Rainfall Prediction Using Intelligent Retrieval and Data Analytics

S. Dhamodaran, J.Refonaa ,R.Ranjith Kumar, G.Pavan Kumar

Abstract : Many new approaches are being researched on to predict the rain before hand in order to minimize the amount of damage to a particular area or make the people of the area aware of the rain hit so that they could take some previous safety measures. Machine learning is gaining a ital growth in almost all the technologies and prediction of weather is not an exception. In this paper we have proposed a system that could predict the rain before hand using machine learning techniques. We have used machine learning model that uses a unique algorithm for constructing the rain patterns. This patterns are then used to get the meteoric information on the web and then the datasets are retrieved for the Pinglin lookout of Central Meteoric Administration of Taiwan's Ministry of Communications. Initially the algorithm works on gathering some of the information like rain, humidness and temperature of a particular areas and then making use of random forest for predicting the rain. Apache Spark is used for evaluating the results of the designed model[7]][10]. When compared to other approaches classification of data seems to be an efficient way of predicting the rain as hence it is used in the current work. The evaluation results are performed based on evaluating various parameters and the proposed model seems to provide a better efficiency when compared to the rest of the previous traditional rainfall prediction systems.

Keywords: Big data, Apache Spark, Random forest, Decision tree, Rainfall, Prediction, Machine Learning techniques

I. INTRODUCTION

The flood said to be one among the dangerous hazards that cause severe damages to many parts especially in the parts of India, China, Bangladesh and Taiwan. These floods majorly contribute to almost seventy five percentage of risk of death around the world and it also plays a major role in about two million folks which are full of impact including the mountain landslide. Apart from this rainfall is also responsible for about seventy eight million folks of expose in an annum that lies inside the boundaries of dangerous tropical cyclone. Extraordinarily most of the very dangerous hazards in the country occur due to the frequent rainfall in our country and hence it suffers from numerous disasters. Some of the major reasons for the downfall of the tropical cyclone is due to some of the vital problems such as in Taiwan, which in the end results directly to the severe and voluminous amount of downpour in the country. As there are numerous technologies that are been vitally used for many of the research purposes, detection of weather is not an exception. Many research works are pertaining to predicting the climatic conditions well before to minimize the amount of damage of lives and properties. A rainfall is one of the most important climatic change that needs to be predicted well before as our country is losing numerous lives due to this and also due the disaster that are caused due to heavy rainfall. In this paper, we have proposed a system for predicting the rainfall well before by using various machine learning techniques. The model uses the dataset that is o btained from the Pinglin lookout of Central Meteoric Administration of Taiwan's Ministry of Communications and is used for the prediction of the rainfall. The dataset is used in the machine learning approach to retrieve the data and also other parameters such as temperature, humidity and rainfall. All the parameters are evaluated to make the perfect prediction of the rainfall. The current research work makes use of random forests to predict the rain by making it use in the Apache Spark platform. Though there are numerous methods available classification of data tends to be one of the most used and accurate method for predicting the climatic changes. The choice of which classifier or the tree is used for the classification of the data is one of the most important aspect of any prediction analysis. Tree is majorly used in the techniques as it is simple to precise, straightforward to browse, and doesn't need users to produce advanced parameters. Therefore, call tree is the tree that is used predominantly used by many researchers. Apache Spark is used for the computing platform as it is versatile for computing and the framework is appropriate for building applications like Spark streaming information flow analysis, Spark SQL [2] interaction analysis, MLlib machine learning, etc. 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

In [3], author has performed a typical survey on all the available Neural Network architectures that are intensly used for predicting the rainfall in last twenty five years. The authors highlighted that almost all of the researchers got vital leads to downfall prediction by exploitation Propagation Network, furthermore the statement techniques that used SVM, MLP, BPN, RBFN, and Kyrgyzstani

Revised Manuscript Received on May 10, 2019

S.Dhamodaran, Assistant Professor, Computer Science and Engineering, Sathyabama Institute of Science and Technology,Chennai,India.

J.Refonaa, Assistant Professor Computer Science and Engineering, Sathyabama Institute of Science and Technology,Chennai,India.

R.Ranjith Kumar, Student, Computer Science and Engineering, Sathyabama Institute of Science and Technology,Chennai,India.

G.Pavan Kumar Student, Computer Science and Engineering, Sathyabama Institute of Science and Technology,Chennai,India.

monetary unit are additional appropriate than alternative applied math and numerical techniques. Some limitations have additionally been highlighted. Researchers in [4] used prediction in Asian country. . For future direction it absolutely was urged that few extra options would be enclosed in computer file for downfall prediction like ocean Surface Temperature for the areas around Andhra Pradesh and Southern a part of India. Researchers in [5] foretold monthly downfall by exploitation Back Propagation, Radial Basis perform and Neural Network. For prediction, the dataset was collected from district Tamil Nadu. Researchers in [6] bestowed a Hybrid System . In ANN, MLP works because the data processing engine to perform predictions whereas the Genetic formula was utilised for inputs, the association structure between the inputs, the output layers and to create the coaching of Neural Network simpler. Researchers in [8] mentioned downfall pace in previous years with relation to varied crops seasons like rabi, Kharif, zaid then foretold (rainfall) for future seasons. In [9], one month and 2 month statement models were developed for downfall prediction. The input dataset was selected from multiple stations in North India, spanned on past 141 years. Neural Network exploitation Back Propagation and Levenberg-Marquardt coaching perform were employed in these models. Researchers in [10] bestowed associate degree formula by integration data processing and applied math Techniques. The projected technique foretold the downfall in 5 totally different classes such as: Flood, Excess, Normal, Deficit and Drought. The predictors were selected with highest confidence level, supported association rules and derived from native and world atmosphere. From native environment: wind speed, water level pressure, most temperature, and minimum temperature were taken. From world environment: Indian Ocean dipole conditions and southern oscillation were taken In [11], the researchers have predicted the rate of rainfall by using the Wave Neural Network (WNN), that is associated wwith the degree of intergration od some of the comonly used wave techniques that are inturn used for investigating the overall performance. In [12], the researchers have proposed a generic survey and have also performed an overall analysis of assorted neural networks on predicting the rainfall of the particular area.

III PROPOSED PREDICTION MODEL

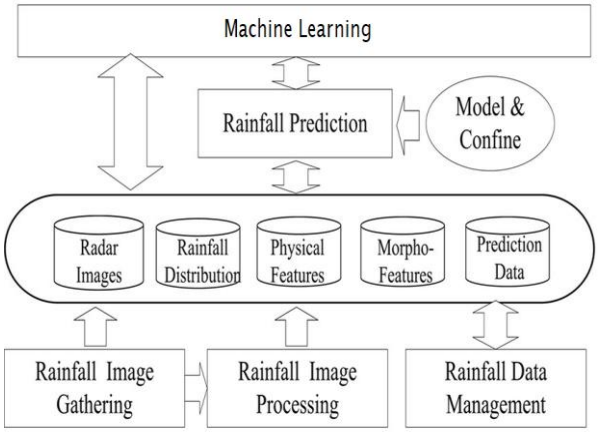

Fig. 1 Block Diagram of Proposed Prediction Model

The huge dataset is hold on and processed in Apache Spark system. The volumes knowledge were pre-processed as shown in fig. 1 and used Bayesian classifiers to get the model. The model is employed to get the prediction of future precipitation supported the dataset. Apache Spark may be a versatile computing framework appropriate for applications like Spark streaming knowledge [13], Spark SQL interaction analysis, etc. thus Apache Spark was used as computing platform. Apache Spark permits the user to load knowledge into a cluster, to beat the disengagement within the existing system we tend to propose a big data based mostly system to extend the potency and accuracy. To handle voluminous knowledge we tend to are victimization Hadoop to store and retrieve data from the distributed hadoop filing system (hdfs). Random forest formula is to be enforced for prediction precipitation.

A. Data assortment and pre-processing

Weather dataset is collected during a comma separated values (CSV) file for the last 3 or four years of precipitation info. The dataset contains the month wise aggregation. The dataset would possibly contain empty values, negative values or error [14]. Dataset is cleansed within the pre-processing. The preprocessing strategies involve of removing records that isn't complete. Once the clean dataset is obtainable we've to arrange it to feed to the machine learning formula.

B. Forest model generation

Forests or random call forests are associate ensemble learning technique for regression and different tasks, by constructing a mess of call trees on coaching time and output the category that's mode of categories (classification) or regression of the individual trees. forest builds multiple call trees and joins them along to induce a lot of correct [15] and stable prediction. One of the most massive advantages of using random forest tree is that, it may be used separately for each classification of the data and also for many of the regression related issues, which is achieved by the use of various machine learning techniques. We tend to train our system with dataset and make the model for future prediction.

C.Prediction, result presentation

it may be used for each regression tasks which it's straightforward to look at the relative it assigns to input areas. Random Forest is additionally thought-about as a awfully handy and simple to use formula, as a result of its default hyper parameters usually turn out a decent prediction result. the amount of hyper parameters is additionally not that top and that they are simple to grasp. The Random forest trained model is employed to predict the precipitation for a selected month. The prediction period is for a pair of month. Python matplotlib library may be accustomed turn out the graphical illustration of the information during a visual format. the expected knowledge is additionally aforethought within the graph together with the prevailing history data.

D. Hadoop implementation

The dataset is often hold on and processed within the same directory of the

machine wherever we tend to run the machine learning formula. however once the dimensions of the information is increased in voluminous size, loading and process the information becomes troublesome. to extend the potency of the system, we tend to use massive knowledge technology to store and retrieve data. Hadoop distributed filing system (hdfs) is employed to store the information within the distributed surroundings. the information.

IV EXPERIMENTAL RESULTS

Compiled a listing of rainfall that have been resulted in landslides on Hadoop implementation as shown in Fig.2. The catalogue was represent the only largest assortment of knowledge on rainfall landslides. Objective and reproducible ID and erectile dysfunction empirical rain thresholds were calculated for the dataset and for a few body regions (regional thresholds). In thresholds were conjointly outlined. thresholds were enforced in a very image landslide warning system supported rain measurements, quantitative rain forecasts and rainfall thresholds.

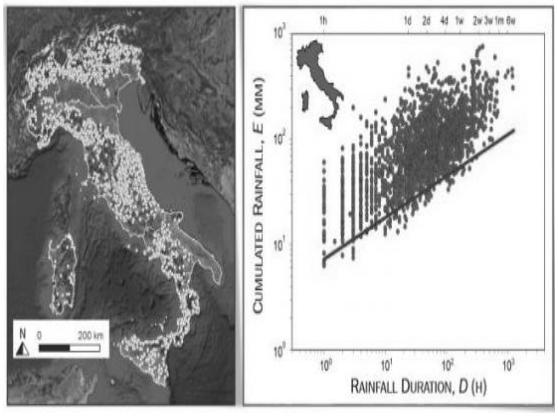

Fig. 2 Result Evaluation

V CONCLUSION

Prediction of weather has intensively got a rapid growth in the imminent technological era. Machine learning techniques and Deep learning Techniques play a vital role in predicting the weather of a specific area or a country for a particular period of time[7]. In this paper, we have proposed an efficient model of using Random Forest Tree for predicting the rainfall in a particular area well before to avoid the loss of lives and property damage. The random forest tree is built using the Apache Spark platform and the model is evaluated on the basis of various parameters. The proposed model is observed to perform better than any of the previously existing models. Future work could include the use of any security protocols for securing the entire system.

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. 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.
4. S. Dhamodaran, Dr. M. Lakshmi "Design And Analysis Of Spatial–Temporal Model Using Hydrological Techniques" IEEE International Conference on Computing of Power ,Energy & Communication" on March 22nd and 23rd, ISBN 978-1-5090-4324-8/17, 2017.
5. 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
6. N. Solanki and G. P. B, "A Novel Machine Learning Based Approach for Rainfall Prediction," Inf. Commun. Technol. Intell. Syst. (ICTIS 2017) - Vol. 1, vol. 83, no. Ictis 2017, 2018.
7. S.Dhamodaran, A.Shrthi ,Adline Thomas, " High-Resolution Flood Hazard Mapping Using Remote Sensing Data", International Conference on Computation of Power, Energy Information and Communication (ICCPEIC),**ISBN:** 978-1-5090-0901-5,IEEE.
8. P. Ramasubramanian, Arputharaj Kannan ,A genetic-algorithm based neural network short-term forecasting framework for database intrusion prediction system. Soft Comput.10(8): 699-714 (2006)
9. N. Mishra, H. K. Soni, S. Sharma, and A. K. Upadhyay, "Development and Analysis of Artificial Neural Network Models for Rainfall Prediction by Using Time-Series Data," Int. J. Intell. Syst. Appl., vol. 10, no. 1, pp. 16–23, 2018.
10. RS Tamilvizhi T, Parvatha Varthini.B., Manoj. K, An Extended Form of MATLAB To-Map Reduce Frameworks in HADOOP Based Cloud Computing EnvironmentsResearch Journal of Applied Sciences Engineering and Technology 12 (9), 900-906.
11. S. Dhamodaran, K. R. Sachin and Rahul Kumar, "Big Data Implementation of Natural Disaster Monitoring and Alerting System in Real Time Social Network using Hadoop Technology", Indian Journal of Science and Technology, Vol 8(22), IPL0278, September 2015
12. M. P. Darji, V. K. Dabhi, and H. B. Prajapati, "Rainfall forecasting using neural network: A survey,"2015 Int. Conf. Adv. Comput. Eng. Appl., no. March, pp. 706–713, 2015.
13. P. Brereton, B. A. Kitchenham, D. Budgen, M. Turner, and M. Khalil, "Lessons from applying the systematic literature review process within the software engineering domain," J. Syst. Softw., vol. 80, no. 4, pp. 571–583, 2007.
14. B. a. Kitchenham et al., "Preliminary guidelines for empirical research in software engineering," IEEE Trans. Softw. Eng., vol. 28, no. 8, pp. 721–734, 2002.
15. B. Kitchenham and S. Charters, "Guidelines for performing Systematic Literature reviews in Software Engineering Version 2.3," Engineering, vol. 45, no. 4ve, p. 1051, 2007.