European Journal of Molecular & Clinical Medicine
ISSN 2515-8260 Volume 07, Issue 08, 2020

Forecasting & Detection Of Flood Using Random Forest Learning Method.

Shruthi J[1],Sumathi M S[2] ,Srivatsa Raju S[3], Vidya R Pai[4]

[1,3,4]Department of Computer Science and Engineering,B.M.S Institute of Technology Bengaluru, India

[2]Electronics and Telecommunication and Engineering,B.M.S Institute of Technology Bengaluru, India

Abstract

Flood is one of the most destructive natural phenomena which occurs in the nature. The ability to predict this occurrence well play a huge impact on the well-being of the human kind and other natural beings, from the brief history of the study of weather the ancient Mayan's where able to predict floods by the planetary motions which were not that accurate. With the development of technology and human dependency on computers, humans are able to collect large volume of various kinds of data such as planetary positions using mathematical models, Weather data using rain gauge and wind turbines. It is very complex to analyse these data and produce an outcome but with the help of machine learning algorithm once can obtain higher accuracy to forecast flood and alert the region beforehand and avoid losses of the precious.

Keywords: Artificial Neural Network (ANN), K- Nearest Neighbours, Radial Basis Function (RBF), Bayesian Network, Wireless Sensor Network (WSN), Support Vector machine.

1. INTRODUCTION

This paper sheds light about how Random Forest Learning Algorithm helps to predict flood before it happens and also compares the accuracy with other machine learning algorithm used these days traditionally. This paper also shares insights about various machine learning techniques involved in prediction this study can be utilized for better performance characteristics for flood prediction.

2. Motivation

Floods are incredibly destructive and can happen due to many different factors including snow, wind, and even as a result of other natural disasters like hurricanes. Since floods cause a lot of damage to human life, animals and property, we want to find the way to help with this natural disaster. Having a more accurate prediction will allow the damage to be mitigated so counties can even save money for disaster relief in those areas. Theoretically, anyone could use this application but in particular, county officials as well as disaster and prevention relief groups would use the application to predict flooding around their regions of interest and can warn people in a position to take measures against floods. Interested parties of this application include politicians, city planners, disaster relief groups, and government officials dealing with weather-related issues. The application is ideal to generate more insight alongside with current flood detection and prevention methods to protect areas of the country against unavoidable flooding.

3. Machine Learning Techniques Used For Prediction

In the adaptive environment of the present future the machine learning models helps to make better predictions than the traditional mathematical models in support with these model and the modern data science in the 21st century humans are able to predict with much better accuracy in this section presents a brief about the various techniques and their prediction accuracy in case of flood.

Artificial neural network:
With the backpropagation model the results of prediction of flood using this model is has an average accuracy of 92.03% and 88.14%, also the data error involved during the training process is 0.047% and 4.97%.[1]

K-Nearest Neighbors:
With the KNN model the data set consisting of parameters such as temperature, humidity and pressure used to determine the behavior of the model to predict the disaster with weighted moving average technique in a chronological order. The prediction accuracy of this model was proved to be of about 90% but this model does not consider various other parameters such as direction of wind, water level gauges, inland rainfall and so on which can play a key variable in forecasting the flood with higher accuracy. [2]

Radical Basis Function based on IOT:
The RBF is widely used technique to predict weather forecasting, load forecasting and many more, the RBF has three main layers input layer, hidden layer and output layer with each layer the data is classified for certainty as flood or no flood, this model can be used to predict only for a fixed time period such as thirty days. to predict further more data is needed which is a limitation in this case. The smallest error rate in prediction process with RBF is to use 700 times iterations and use the learning rate equal to 0.0007 units. [3]

Bayesian Network Model:
The Bayesian network model is a data driven model, with the increase of error rate in the data set the prediction model fails to perform accurately there are 2 stages for this model Local Bayesian network stage and Global Bayesian network stage where the first divides dataset into small fragments and implements the appropriate function to predict the outcome whereas the later constructs a model for the whole data set and predicts the outcome. [4]

Random Forest :
The random forest model can consider larger dataset for classification, regression and other tasks that operates by constructing a multitude of decision tress at training time and outputting the classification or mean prediction where with a relative small number of sample data set can get a better accuracy, with the more samples might not improve accuracy but can maintain accuracy when large portion of the sample data are missing relatively the computational cost of training a relatively low compared to other models. [5]

4. Flood Predicting Method

The large volume of metadata consisting of entities such as temperature pressure, elevation are recorded in systems where one can access this data and manage using pandas library for extraction of useful data, the daily reports, storm events and precipitation data is also available from the repository, considering these data here for training the random forest model, the daily report data is obtained in CSV format for pandas to analyse and passed to the algorithm as input. The random forest learning algorithm model metadata are heterogeneous which the attributes associated with them makes it complicated for individuals to understand, these data are obtained from various sources and distinct formats which needs to be integrated uniformly to avoid redundancy across the dataset here in this paper the SK learn open source library is used to implement the Random Forest learning model the data set is first cleaned to remove the

redundancy and the null values in the data set for this operation the use of clean data library is utilized the architecture as show in figure 1.1

The main functionality of Random Forest learning algorithm is to choose the sample data randomly and obtain a decision tree where the majority vote is used to make the predictions of the flood the rainfall data consists of only the average rainfall across the country for the month from certain time period, with the SK library visualizer we can obtain the threshold to decide the point where the rainfall above the certain threshold will result in floods in that month the figure 1.2 shows the threshold decision value from the dataset.

5. Data Split

The data set is split using the SK learn test train split library where the data is divided into testing data and training data in 70% of training and 30% testing split the split data is used accordingly for the mentioned cases, from the training split the random forest model is trained to make predictions where it builds n decision trees for the dataset with the built in library of SK learn we can use the predict function to make the predictions of the test data in which the data set in retrieved randomly from the test split model and decision tree is generated for the test split too the prediction is made by the percentage of true values obtain by the random forest model.

6. Result Analysis

The flood prediction using random forest model, the results obtained are 98% for the training model and 95% for the test data, the limitation of the model is when the data set is above the capacity of the hardware the HPC infrastructure needs to be implemented which attracts ways to improve the performance but the collocation of HPC and Big-Data is not easy because of the differences in concepts. The HPC job rigidity create holes in batch scheduler we can use these idle resources as dynamic adaptability for Big-Data workload with the help of Resource and Job Management System's (RJMS) configured to communicate with both Big-Data Systems and HPC's using prolog techniques.

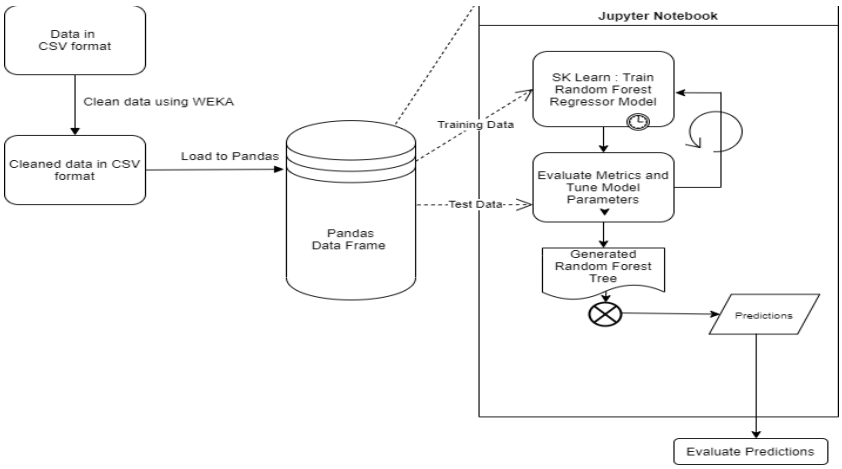

Fig 1: Flood forecasting architecture

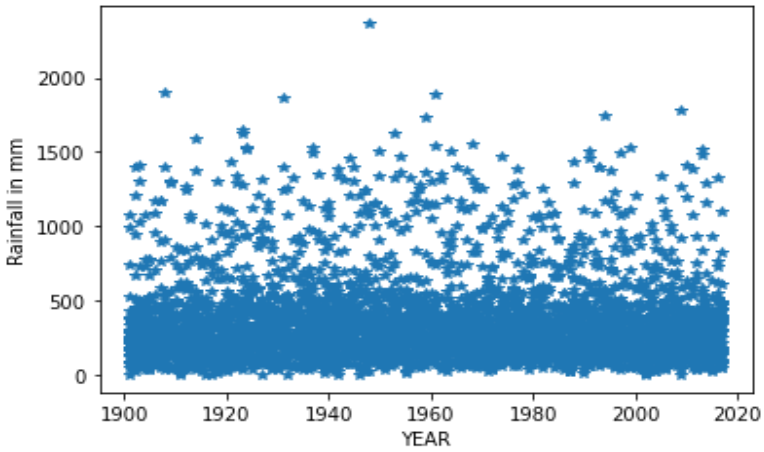

Fig 2: Plot for threshold identification

7. CONCLUSIONS AND FUTURE WORK

This paper discusses about the flood forecasting model using random forest implemented using SK learn library where this can be implemented along with existing architecture for scalability in account of Moore's law where the system must be dynamically scalable and the future work of this paper is where the prediction to be made using much more added parameters such as wind direction, temperature, location specific attributes and so on.

REFERENCES

[1] O. Amale and R. Patil, ―IOT Based Rainfall Monitoring System Using WSN Enabled Architecture‖ 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 789-791.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892,pp.68-73.

[2] C. Nikhil Binoy, N. Arjun, C. Keerthi, S. Sreerag and A. H. Nair, "Flood Prediction Using Flow and Depth Measurement with Artificial Neural Network in Canals," 2019 3rd International Conference on Computing Methodologies and Communication (ICCMC), Erode, India, 2019, pp. 798-801.

[3] J. M. A. Opella and A. A. Hernandez, "Developing a Flood Risk Assessment Using Support Vector Machine and Convolutional Neural Network: A Conceptual Framework," 2019 IEEE 15th International Colloquium on Signal Processing & Its Applications (CSPA), Penang, Malaysia, 2019, pp. 260-265.

[4] K. Ega Kartika, M. A. Murti and C. Setianingsih, "Floods Prediction Using Radial Basis Function (RBF) Based on Internet of Things (IoT)," 2019 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), BALI, Indonesia, 2019, pp. 125-128.

[5] Zainudin, Suhaila & Jasim, Dalia & Abu Bakar, Azuraliza. (2016). Comparative Analysis of Data Mining Techniques for Malaysian Rainfall Prediction. International Journal on Advanced Science, Engineering and Information Technology. 6. 1148. 10.18517/ijaseit.6.6.1487.

[6] Mannan, Md & Abdul, Md & Chowdhury, Mannan & Karmakar, Samarendra & Ahmed, Shamsuddin & Mason, Simon. (2015). Rainfall prediction over northeastern part of Bangladesh during monsoon season. DEW-DROP, A Scientific Journal of Meteorology and Geo-Physics, Bangladesh Meteorological Department. 01.14-25.

[7] Castro, Joel & Salistre, Gabriel & Byun, Y.-C & Gerardo, Bobby. (2013). Flash Flood Prediction Model based on Multiple Regression Analysis for Decision Support System. Lecture Notes in Engineering and Computer Science. 2. 802-807.

[8] C. Thirumalai, K. S. Harsha, M. L. Deepak and K. C. Krishna, "Heuristic prediction of rainfall using machine learning techniques," 2017 International Conference on Trends in Electronics and Informatics (ICEI), Tirunelveli, 2017, pp.1114-1117.

[9] R. V. H. Ginardi, A. Munif, K. W. Nurul, A. Rosyidi, A. Khoiruddin and R. Perdana, "System design to use weather forecast & current condition information on travel route with KNN and graphical data on Android application," 2017 11th International Conference on Information & Communication Technology and System (ICTS), Surabaya, 2017, pp. 287-292.

[10] Bangera, Clyde & Kotian, Priya & Dias, Chrislene & Divya, T & Aithal, Ganesh. (2018). Flood and Heat Wave Prediction using Weighted Moving Average, Anomaly Detection and K-Nearest Neighbours for the City Of Mangalore. 93-97. 10.1109/DISCOVER.2018.8674113.

[11] Y. Wu, W. Xu, J. Fengt, S. Palaiahnakote and T. Lu, "Local and Global Bayesian Network based Model for Flood Prediction," 2018 24th International Conference on Pattern Recognition (ICPR), Beijing, 2018, pp. 225-230.

[12] A. R. Sanubari, P. D. Kusuma and C. Setianingsih, "Flood Modelling and Prediction Using Artificial Neural Network," 2018 IEEE International Conference on Internet of Things and Intelligence System (IOTAIS), Bali, 2018, pp. 227-233.

[13] T. Khan et al., "Flash Floods Prediction using Real Time data: An Implementation of ANN-PSO with less False Alarm," 2019 IEEE International Instrumentation and Measurement Technology Conference (I2MTC), Auckland, New Zealand, 2019, pp. 1- 6.

[14] J. Kwon, N. L. Kim, M. Kang and J. WonKim, "Design and Prototyping of Container-Enabled Cluster for High Performance Data Analytics," 2019 International Conference on Information Networking (ICOIN), Kuala Lumpur, Malaysia, 2019.

[15] D. Dai, R. B. Ross, P. Carns, D. Kimpe and Y. Chen, "Using Property Graphs for Rich Metadata Management in HPC Systems," 2014 9th Parallel Data Storage Workshop, New Orleans, LA, 2014, pp. 7-12.

[16] J. Emeras, S. Varrette and P. Bouvry, "Amazon Elastic Compute Cloud (EC2) vs. In-House HPC Platform: A Cost Analysis," 2016 IEEE 9th International Conference on Cloud Computing (CLOUD), San Francisco, CA, 2016, pp. 284-293.