Water Scarcity Predictor and Supply Scheduling

Swapnil Verlekar[1], Alisha Shah[2], Snehal Kulkarni[3]

[1,2]*Dept. of Computer Engineering, St. Francis Institute of Technology, Maharashtra, India*
[3]*Assistant Professor, Dept. of Computer Engineering, St. Francis Institute of Technology, Maharashtra, India*

Abstract - *The profligate water usage as a consequence of population growth, rapid urbanization and climate change has become a predominant issue. In numerous cases lack of rainfall is followed by acute water scarcity diminishing the ability of water bodies to meet daily water needs. This paper is based on developing a machine learning model for an efficient and accurate water level prediction for a reservoir. The model studies the previous data trends and enhances the ability to diagnose scarcity, assists in providing supply schedule and improve operational efficiency. In addition it plays a pivotal role in knowing about areas with abundance of water resource. It majorly focuses on automating data analysis process and providing an accurate visualization graph describing the water level for each reservoir. The model is hinged on a comparative study of different regression algorithms like Simple Linear Regression and Random forest regression. The results demonstrate that Random forest regression provides higher prediction accuracy due to the use of multiple Decision tree and its Bootstrapping technique.*

Key Words*: *Water scarcity, Prediction, Machine Learning, Random Forest Regression, Linear Regression, Data Visualization, Rainfall, Water Consumption, Supervised Learning, Water Demand, Forecasting.

1. INTRODUCTION

Traditional methods for water shortage treatment include manually working after it has occurred. This method of treatment is not very useful and it has many disadvantages. It is very difficult to treat scarcity after it has occurred. In recent years, Machine Learning has emerged to be useful in prediction. In our system, we will mainly focus on predicting water shortage and scheduling a supply between the reservoirs so that each reservoir has ample amount of water.

Water shortage can mean scarcity in availability due to physical shortage, or scarcity in access due to the failure of institutions to ensure a regular supply or due to a lack of adequate infrastructure. Our primary focus is to encounter the issues caused due to improper and ineffective use of resource. The water shortage already affects every continent. Water use has been growing globally at more than twice the rate of population increases in the last century, and an increasing number of regions are reaching the limit at which water services can be sustainably delivered, especially in arid regions. Water shortage will be exacerbated as rapidly growing urban areas place heavy pressure on neighbouring water resources. Climate change and bio-energy demands are also expected to amplify the already complex relationship between world development and water demand. There is not a global water shortage as such, but individual countries and regions need to urgently tackle the critical problems presented by water stress. Water has to be treated as a scarce resource, with a far stronger focus on managing demand. Integrated water resources management provides a broad framework for governments to align water use patterns with the needs and demands of different users, including the environment. Hence introduction of a system to cater the need for automating the manual analysis and schedule is of essence. Water Scarcity predictor studies the trends and predicts the water level for each reservoir and sets a threshold above which usage of resource may prone to shortage of water.

1.1 Motivation

In 2019, Chennai faced a huge scarcity of water causing a severe drought to occur, forcing the city officials to declare a “Day Zero” a day when the whole City had run dry. Reason being two years of deficient rainfall and lack of water resource management capability based on analytics. There are four reservoirs in the city, namely, Red Hills, Cholavaram, Poondi and Chembarambakkam which acts as a primary source to meet the water needs. Various factor caused a stress on these reservoirs to suffice the water needs of the City. Similar such event occurred in 2004. This urges the need for a system to assist the decision making and prescient of the occurrence of any such event. A machine learning model with the ability to predict the water level in the four major reservoirs which intimates the possibility of a scarcity can help preclude the situation before it actually occurring by proper planning.

1.2 Proposed Solution

The data plays a vital role in the efficiency and accuracy of prediction for any Machine learning model. The proposed solution is, based on the comparative study of various supervised learning algorithm implementing a model to predict the water level for each reservoir and setting an appropriate threshold after calculating the daily requirement and its stress on each reservoir. Accuracy of model is highly depended on impact of parameters selected for prediction. The input parameters would be the factors affecting water level, i.e. Rainfall, Population, Water consumption and available water resources. Based on previous trends and variations in the reservoir water level following the above mentioned parameters the model is trained. As and when water shortage is predicted, it is displayed using graphical visualization. A relation between population and water consumption in a particular reservoir is created from which an ideal amount of water required by that reservoir is derived.

All regions in Chennai are mainly dependent on four major reservoirs. Accordingly, a minimum threshold of water required for each individual water resource is deduced. If a particular resource exceeds above the threshold, that additional portion of water can be supplied to a scarcity affected reservoir. Similarly, if one resource has contributed its share of water to one reservoir and if that reservoir is still suffering from scarcity, other resources with abundance of water will contribute for the same. Finally, a diagnosis is conducted and a report is displayed with an embedded map to calculate the transportation cost and time requirement.

2. REVIEW LITERATURE

In paper [1], the efficiency of two computational intelligent techniques Artificial Neural Network (ANN) and Support Vector Machine (SVM) in water demand forecasting is compared. It was observed that ANN performs significantly better than SVM. In paper [2], based on consumption characteristics, urban water demand forecasting is done. It describes the process of Feed-Forward Neural Network and feedback process of NN model. Results show that the process of NN model for forecasting urban water demand has certain theoretical and practical value which can be used in further studies. In paper [3], a machine learning approach for effective day-ahead forecasting using retrospective metering data and open source weather information provided by meteorological services is proposed. Machine learning models thus can prove effective in prediction of a reservoir water level. In paper [4], factors that have a greater impact on urban water consumption was identified. It was found that the daily max temperature and the historical water consumption data have highest correlation with urban water consumption. Different models were implemented for increasing the accuracy of the system.

3. DATA DESCRIPTION AND ANALYSIS

The water level in each of the four major reservoirs in Chennai is directly proportional to the amount of rainfall received. Other factors which have a major impact on reservoir level are population and daily water consumption. The model needs to be trained on real time historic data for each reservoir and rainfall in each of the particular reservoirs. Real time Data is acquired from data.gov and the water consumption is based on calculation for water requirement per person within the vicinity. Based on the acquired data and considering all the impacting factors data is pre-processed.

Data for different resources is explored and analyzed to support the system. Every year there is a detrimental phase and a replenishment phase (mainly during October to December). In 2004 the region faced acute scarcity of water, since the water level in all four reservoirs was almost diminished. It was germane to select the months which had major repercussion on reservoir water level and map it to the remaining month's data. Hence, a separate dataset for each reservoir mapping all the rain abundance months with reservoir water level for the left months is required. Definitive parameters include Population, Rainfall, Water consumption, Reservoir water level.

3.1 Inferences on rainfall pattern

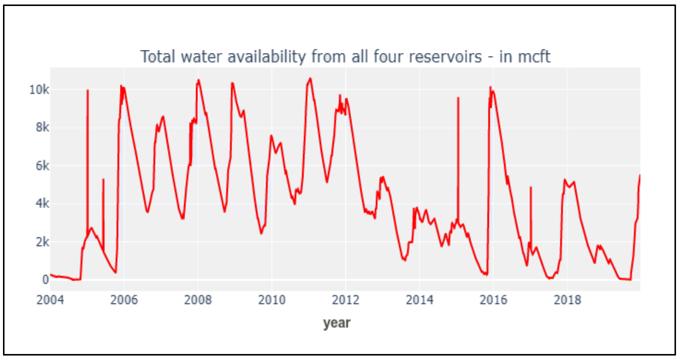

Chart -1: Plot of Water Levels in Four Major Reservoirs

- Summing up all the water availability from four reservoirs, we can see that the water levels reached almost nil thrice (2004, 2017 and 2019)
- Generally after the rainfall, the reservoirs used to get replenished to about 10K mcft. until 2012 which is not the case afterwards due to the lack of rainfall.
- Only during the infamous Chennai floods of 2015, it has reached the 10K level after 2012.
- 2017 is similar to 2019 in terms of depletion of water availability but the levels reached close to 0 during end of august unlike now where the levels reached in the beginning of June itself.

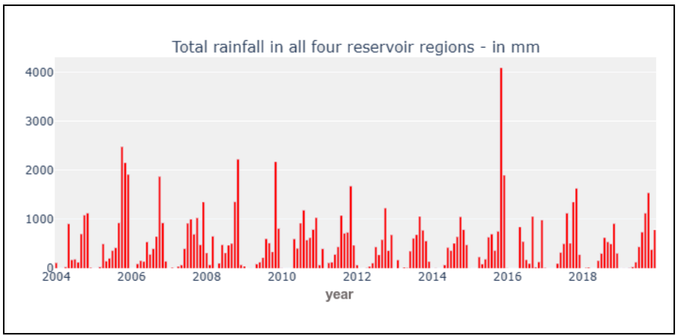

Chart -2: Plot of Rainfall in Four Major Reservoirs

- Looks like the city gets some rains in the month of June, July, August and September due to south west monsoon.
- Major rainfall happens during October of every year which is due to North-east monsoon.
- During the initial years rain from north-east monsoon is much higher than south-west monsoon. But seems like last few years, they both are similar (reduction in rains from north-east monsoon).
- We have got some good rains in August and September 2019, but the water reservoir levels are yet to go up.

4. METHOD OF PREDICTION

The system is based on the supervised learning technique, since we have enough historical data on water level for each reservoir. Different regression algorithms are compared for their prediction accuracy.

4.1 Linear Regression

Linear Regression is a machine learning algorithm based on supervised learning. It performs a regression task. Regression model predicts the value based on independent variables. It is mostly used for finding out the relationship between variables and forecasting. Different regression models differ based on the kind of relationship between dependent and independent variables, they are considering and the number of independent variables being used.
Equation for Linear regression:

$$Y = b0 + b1*X1 + b2*X2 + b3*X3 + \dots$$

Where Y is dependent variable and X notes dependent variables with coefficient b0, b1, b3, …
For the water level predictor, considering reservoir level as dependent variable with rainfall, water consumption and population as independent variable one-to-many mapping is establish for the model to train.

4.2 Random Forest Regression

A Random Forest is an ensemble technique which can be used for performing both Regression and Classification tasks with the use of multiple decision trees and a technique called Bootstrap. Forest is made up of trees and more trees means more robust forest. Similarly, random forest algorithm creates decision trees on data samples and then gets the prediction from each of them and finally selects the best solution by means of voting.
More formally we can write this class of models as:

$$g(x) = f0(x) + f1(x) + f2(x) + \dots$$

Where the final model g is the sum of simple base models fit. For our model we have utilized the Random forest Regressor, it fits a number of decision trees on various sub-samples of the dataset and uses averaging causing low variance and improves the predictive accuracy and controls over-fitting.

5. RESULT AND DISCUSSION

Practical experiment for water level prediction for each reservoir using different Regression algorithm are performed and embedded map for optimal resource supply is hinted.

5.1 Performance of Algorithm

The Random Forest regression model produced the best results compared to other implemented model. The performance was evaluated using the R squared, since the model is regression problem. Random forest regression gave accuracy of 96% which is comparatively large as of linear regression model which is 90%. The MLP regressor gave the least accuracy due to the under fitting of curve. Implementing the Random forest regressor gave ~6% hike in accuracy over linear regression. Calculating the population from year using population growth rate and thereby water consumption per person for the region enhanced the model's prediction and improved the models accuracy.

5.2 Display water level bar graph

The below graph displays water level that is predicted by our system for next 7 months. Those bar show the water level each month which can be used to decide if there is scarcity.

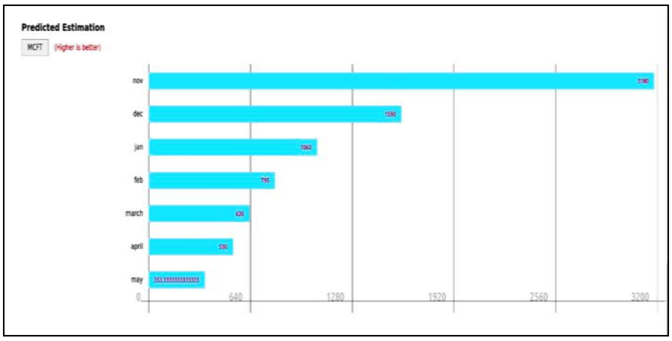

Fig -1: Prediction of reservoir level

5.3 Display threshold suggestion bar graph

The below graph displays water level that is suggested by our system for next 7 months. Those bar show the water level each month which can be used to provide water for all 7 months without scarcity.

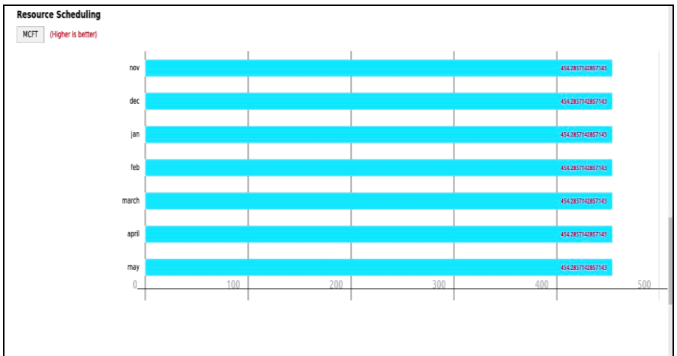

Fig -2: Resource usage schedule

6. CONCLUSIONS

In this paper, based on the comparative study a Random Forest Regressor model for predicting reservoir level is proposed. The water shortage prediction is done based on the rainfall information for 6 months during the rainy season in the 4 major water resources in Chennai. Based on prediction, the system creates a schedule to supply water from the resource with abundant water to the resource that is scarce. Overall, this project focuses on solving one of the biggest problems faced by countries all around the world. Water scarcity will be exacerbated as rapidly growing urban areas place heavy pressure on neighboring water resources. Climate change and bio-energy demands are also expected to amplify the already complex relationship between world development and water demand. And because of this very reason, we believe that this project will have a great impact on the world especially the rural areas.

REFERENCES

[1] Msiza, Ishmael S., Fulufhelo Vincent Nelwamondo, and Tshilidzi Marwala. "Water demand prediction using artificial neural networks and support vector regression." (2008).

[2] Ren, Yongchang, Tao Xing, Xiaoji Chen, E. Xu, and Ying Zhao. "Based on process neural network learning algorithm for prediction of urban water consumption." In 2010 International Conference on Computing, Control and Industrial Engineering, vol. 2, pp. 34-37. IEEE, 2010.

[3] Khalyasmaa, Alexandra, Stanislav A. Eroshenko, Teja Piepur Chakravarthy, Venu Gopal Gasi, Sandeep Kumar Yadav Bollu, Raphael Caire, Sai Kumar Reddy Atluri, and Suresh Karrolla. "Prediction of Solar Power Generation Based on Random Forest Regressor Model." In 2019 International Multi-Conference on Engineering, Computer and Information Sciences (SIBIRCON), pp. 0780-0785. IEEE, 2019.

[4] Hu, Piao, Jun Tong, Jingcheng Wang, Yue Yang, and Luca de Oliveira Turci. "A hybrid model based on CNN and Bi-LSTM for urban water demand prediction." In 2019 IEEE Congress on Evolutionary Computation (CEC), pp. 1088-1094. IEEE, 2019.