Scientific Journal of Impact Factor (SJIF): 5.71

e-ISSN (O): 2348-4470
p-ISSN (P): 2348-6406

International Journal of Advance Engineering and Research Development

Volume 5, Issue 10, October -2018

WATER MANAGEMENT USING BIG DATA ANALYTICS

[1]S. Nandhini, [2]Anirudh Bansal, [3]Param Saluja, [4]Sayon Bhattacharyya, [5]Himanshu Joshi

[1,2,3,4,5] SRM Institute of Science and Technology, Chennai, Tamil Nadu

Abstract—*Water is the most precious natural resource for well sustained life. As we know that population and urbanization grow at very high rate which in result cause serious disturbances in whole value-chain in water ecosystem. Due to climate change and human mishandling the resources, recently NITI AAYOG presents report which clearly indicates that by 2030 India will face a worst water crisis especially our demand of potable water will outstrip supply. So a proper and robust system is in the need of the hour. After 3 decades of various policy and programs we have not achieved the sustained way and also our policymakers also failed to consolidate all stakeholder and data related to every component of policy. As mentioned one of the reasons for failure in implementation is that there is no proper consolidation and analysis of the data of various programs, river, streams, aquifers etc. So we proposed a robust system which runs on the principles of Big Data, and we implemented these data sets with the help of various new analytical algorithms such as Random forest techniques and some machine learning techniques. Not only we have gathered the information from historical and governmental sources, we provide visualization of analysis. Ergo our outcome not only show various trends in water quality but it also helps in understanding the parameters and their needs, which in turn helps to find root cause of failure and helps in future planning.*

Index Terms—*Water Management, Big Data, CPCB*

I. INTRODUCTION

$6^{T\ H}$ goal in SDG is Clean water and Sanitation, which directly indicates important parameter which is quality of water. Contaminated water causes around 500 million deaths due to various water-borne diseases. Annually around $400 billion are spent on various water bills including maintaining and regulating quality of water. India is rich in water resources as we have a very good network of rivers and Himalayans glaciers that are capable of meet various daily needs. But eventually due to over usage and more dependence our quality of water decrease very rapidly as pollution are at very high pace. So if we want to achieve SDG goal by time we have to see this problem holistically so that we can create a system of proper monitoring and analytically efficient.

Quality of water is most important aspect in whole eco-logical chain of water ecosystem. It plays a vital role in deciding the usage as we need clean and potable water for drinking, water contains all micro nutrient for irrigation, for well sustained life etc. Now for monitoring of quality our government had given the responsibility to Central Pollution Control Board (CPCB) which have networked of monitoring stations on rivers across the country. CPCB is an apex body in the field of water quality management in India. It has around 1032 monitoring stations in the country spread over all important water bodies. So there is need for a proper and robust system which helps us not only in predicting the outcome but also able present the reality of the resources used in process. For a rational planning of any water quality management program, we need to know the nature and extent of water quality degradation. Therefore, a scientific water quality monitoring system and proper analytical tools are prerequisite.

Rapid growth at population, pollution and urbanization our water resources are going at very bad stage where degradation rate is very high. This trend is very alarming because quality of water must remain wholesome so that our existence does not disturb so badly. It is therefore necessary to keep vigilant watch on quality of available fresh waters whose major sources in our country are rivers and also other sources viz. grey water, rainwater etc.

So making a robust and efficient system in data analysis in water quality management, or we can say that for fully water management system. We also want to improve efficiency in current CPCB monitoring system at step 7 of data manage-ment. Some specific aims are as follows:-

For planning of pollution control strategies.

To assess nature and extent of pollution control needed in different water bodies or their tributaries.

To evaluate effectiveness of currently existence system. To evaluate water quality trend over a period of time.

To understand the environmental nature of different pol-lutants.

To assess the fitness of water for different uses.

As we discussed the importance of water quality and its analysis for future planning, now some light on monitoring as well how CPCB handles the information about quality parameters.

In the field of water quality management, there has been a steady evolution in procedures for designing system to obtain information on the changes of water quality. The 'monitoring' comprises all activities to obtain 'information' with respect to the water and its sources.

Water quality monitoring is a complex subject, and the scope of it is both deep and very important for us. Water quality monitoring has a direct relation with chemistry, bi-ology, statistics, economics and now data analytics, machine learning etc.Its scope is also related to the types of water uses and functions which are manifold and the nature of the sources of water such as surface water (rivers and lakes), sea water groundwater etc.

We are not handling this system for only analysis purposes, but we have obligations by legal matter to protect and do something in a sustainable manner. Now some legal sources which gave us a sense of responsibility to perform these tasks.

Article 48A enjoins the state to protect and improve the environment and safeguard the forests and wildlife in the country.

Article 51A(g) states that the "fundamental duty of every citizen is to protect and improve the natural environment including forests, lakes, rivers and wildlife and to have com-passion for living creatures
Water policy 2002; especially section 25 which is about Science and Technology role and various domain.
The Water (Prevention & Control of Pollution) Act, 1974 Environment (Protection) Act, 1986

II. APPROCH TO MONITORING

As we till now discuss the importance of management and data analysis for water quality now sees a how current system works It occurs in two stages, which are as follows:

1. Water Quality Monitoring Network

Due to economic considerations, monitoring network de-sign, sampling frequencies, choice of variables and frequency of laboratory analysis should be determined on the basis of the information requirements, the hydraulic and hydrologic constraints, and variability in water body characteristics, the end-use of water that drains to and from the water bodies. It is also important to optimize the amount of efforts required and information generated and its importance to fulfill the set objectives. The scoping and designing step are the foundation of the entire water quality monitoring system. The main objective of the design should be to minimize the cost of monitoring without sacrificing the desired information to the level of precision and accuracy. Scoping and designing of water quality monitoring system is based on clear scientific understanding of :

1. Issues
2. Relevant background information
3. Monitoring objectives
4. Desired outcomes
5. Appropriate methods
6. The dynamics and characteristics of water systems

2. Analytical Quality Control

Water samples are being analyzed for 28 parameters con-

sisting of physico-chemical and bacteriological parameters for water samples apart from the field observations. Besides this, 9 trace metals Mercury (Hg), Arsenic (As), Chromium (Cr), Cadmium (Cd), Lead (Pb), Copper (Cu), Nickel (Ni), Zinc (Zn)

and Iron (Fe) and 15 pesticides (Alpha BHC, Beta BHC, Gama BHC (Lindane), OP DDT,PP DDT, Alpha Endosulphan, Beta Endosulphan, Dieldrin, Aldrin, Carbonyl (Carbonate), 2-4 D, Malathion, Methyl parathion, Anilophos and Choropyriphos are analyzed in selected samples collected by various devices.. In view of limited resources, limited numbers of organic pollution related parameters are chosen for frequent monitoring i.e. monthly or quarterly and major cations, anions, other inorganic ions and micro pollutants (Toxic Metals & POP's) are analyzed once in a year to keep a track of water quality over large period of time. The water quality data are reported in Water Quality Statistics yearbooks.

SYSTEM OF CPCB FOR WATER QUALITY MONITOR-ING:
Central Pollution Control Board started national water qual-ity monitoring in 1978 under Global Environmental Monitor-ing System (GEMS), Water Program. Monitoring Program was started with 24 surface water and 11 groundwater stations.

Parallel to GEMS, a National Program of Monitoring of Indian National Aquatic Resources (MINARS), was started in 1984, with a total of 113 stations spread over 10 river basins. The present network comprises 1032 stations on rivers, water bodies and subsurface waters. Some components are as follows:

1) Spectrometer probe can measure SS, COD, BOD, TOC, Nitrate, Color and Turbidity. It uses UV spectrometer principle.

2) Optical spectra in liquid media.

3) Various light techniques for various wavelengths.

4) Optical sensors operate using the principle of optical or fluorescence can monitor dissolved oxygen & tempera-ture.

5) pH, Conductivity, Salinity and Temperature can be mea-sured/monitored using the principal of multi parameter electrode.

6) Amperometric membrane can monitor free chlorine (Cl2 + HOCl + OCl-) or total chlorine (free chlorine + combined chlorine).

All the water quality monitoring stations will be fitted with GSM, GPRS for communication with the central receiving sta-tion. The central receiving station will have software for data acquisition, data analysis, data display and report generation. Zonal offices of CPCB and State Pollution Control Boards will have direct access to the data from central receiving station.

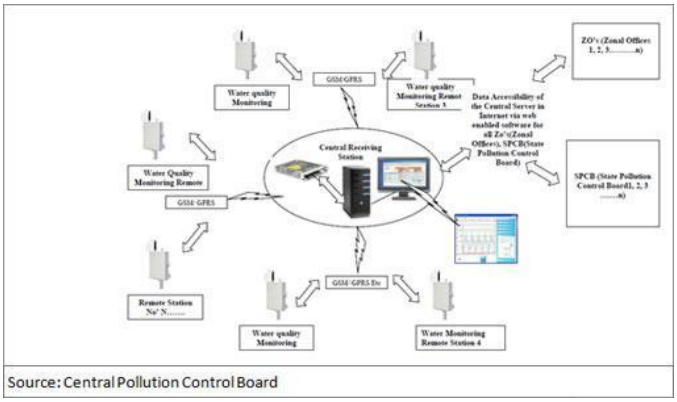

Fig. 1. CPCB monitoring system pictorial diagram

BIG DATA ANALYTICS:

Before we see the main part of big data first we look into the problem with some traditional problem. Some tradi-tional statistical methods are computational intelligence, linear regression, data collection, clustering method, discriminant analysis, Multidimensional Scaling, Single linkage, complete linkage, Average linkage etc.

Some disadvantages of these techniques are:

1) Chances for delay are more which in turn cause to loss of information.
2) Fluctuation is very high.
3) Not able to handle very large variety and volume of data.
4) Not very efficient in predicting future outcomes.

Random Forest is a supervised learning algorithm. It creates a forest and makes it somehow random. The forest it builds

Fig. 2. Data Processingflowchart

is an ensemble of Decision Trees, most of the time trained with the "bagging" method. The general idea of the bagging method is that a combination of learning models increases the overall result. Random forest builds multiple decision trees and merges them together to get a more accurate and stable prediction.

It needs SK learn library packages in python and R language and also it needs Hadoop Environment. It is Machine Learning tool which can implemented in any latest configured hardware system.

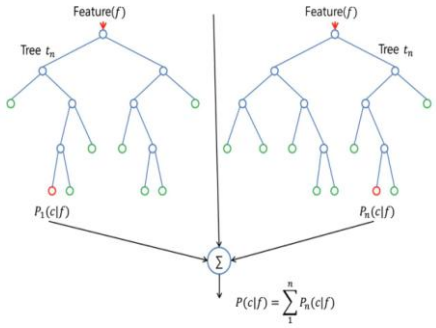

Fig. 3. Data Structure for random forest algorithm Advantages of Random Forest Algorithm:

1) Increasing the Predictive Power
2) Increasing the Models Speed
3) Alreadyinclude regression so time-consuming
4) Cost effective
5) It is a very handy and easy to use algorithm.

III. RESULT/OUTCOME:

We use a data set about various quality parameters in different states. We use this algorithm for analysis the data set and predict the trends in relations to quality of water.
Trend of quality of water in Tamil Nadu:

It shows us that in Tamil Nadu content of iron in water is

66.94%. So drinking water quality is not so up to the standard.

After iron content, salinity content is about 26%.

Trend of quality of water of Gujarat:

In Gujarat, we observe that unlike TN iron content is very minimal here but salinity and nitrate amount is somewhat is high. The salinity comprises about 30%, nitrate amount is around 35% and fluoride amount is around 38%. The iron content is less than 1%.

Trend of quality of water in India:
Contents in percentage are as follows:

1) Iron : 47.95%

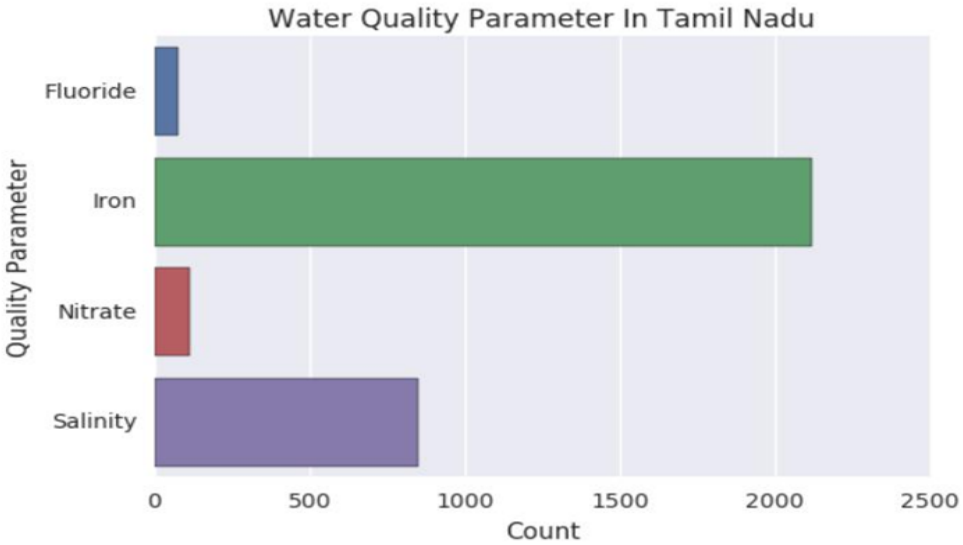

Fig. 4. Visualization of analysis of parameters(Tamil Nadu)

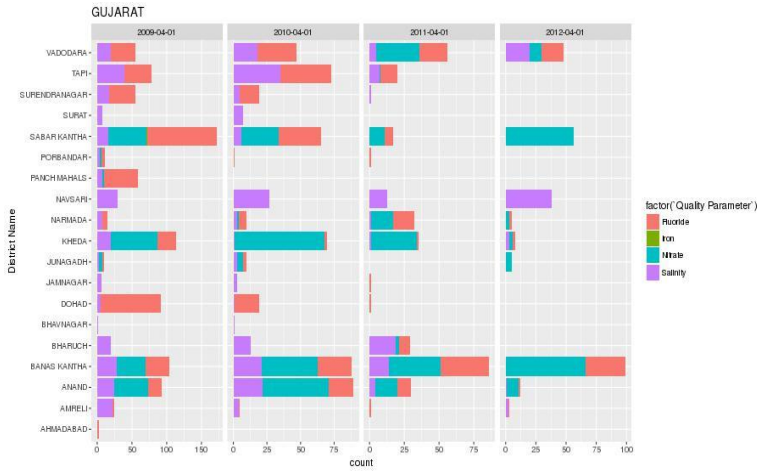

Fig. 5. Visualization of analysis of parameters(district wise Gujrat)

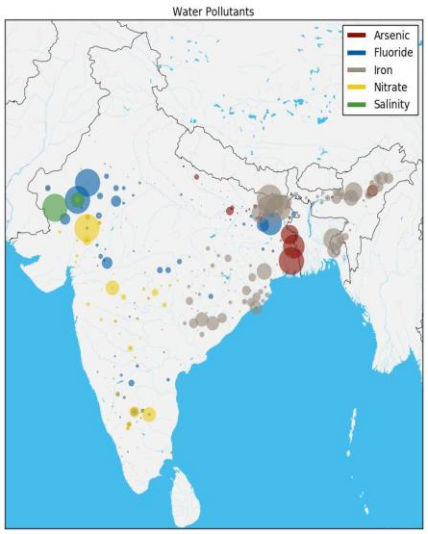

Fig. 6. Visualization of analysis of parameter(map wise)

2) Arsenic : 7.9%

3) Nitrate : 4.01%

4) Salinity : 22.92%

5) Fluoride : 17.7%

This analysis clearly show that availability of drinking water is not satisfactory condition as quality is not up to standard, Hence our water bill is very high as cost for filtration is high.

Some specific outcome from analysis:

1) With the help of analytical data we can make inventory for basins for different parameter such as pollution etc.
2) Water quality data is used for Identification of Polluted Water Bodies based on violation of desired water quality criteria.
3) Allocation of resources for planning for various use of water can easily be managed by quality data.
4) Water quality data is used for Query Response i.e. to reply Parliament Questions, Public Queries, Public Interest Litigation filed in Supreme Court and Various High Courts and to fulfill the requirement of Non-Governmental Organization, Students, and Researchers.

FUTURE EXPANSION:

Improvement in network design of monitoring design.

Incorporating internet connectivity in communication through IOT enabled devices.

Analysis on groundwater and its usage by more geospatial visualization.

Analysis on Grey water and its sources and develop the more predicting system for distribution and usage pattern. Analysis on monsoon and water storage reservoirs.

More on quality parameters such as pesticides, chromium, lead, copper etc..

IV. CONCLUSION

As we know water is very precious natural resource and its conservation importance also have same importance. Many children are trapped in vicious cycle of malnutrition which not only harm their lives but as whole affect the human resource of the country. So proper planning and analysis is need of the hour. For achieving this aspect our analysis system is trying to reduce cost as well as time along with increasing the efficiency, efficacy, accuracy, robustness of the system. With the advent of machine learning and IOT, we can increase the predictability of system to manifold times.The real-time, automated, collection of water quality data will greatly assist on obtaining a secure, safe and sustainable quality of water to those who rely on it. The automated collection of water quality data collection is just a first step in developing an efficient repository of information that can be used by stakeholders, such as regulatory agencies, industry, education, research, communities, as well as the public in general. Hence, this system is act as bridge between policymakers and stakeholders so that proper planning as well usage can be done. Our system will help in achieving real time trends in different parameters such as quality, usage, sources etc. With the help of Hadoop environment and machine learning techniques we implemented one algorithm but according to complexity of data sets we can switch to another algorithms so that our predictability and accuracy factor is always remain at strong side in a spectrum of quality analysis.

V. APPENDICES

CRITERIA OF QUALITY OF WATER:

The Central Pollution Control Board has classified water resources of the country according to their uses for setting water quality objectives for different water bodiesTable I

WATER QUALITY MONITORING STEPS

It consists of 8 steps which cover all aspects of monitoring system.All steps are based on scientific techniques.Table II

REFERENCES

Qasem A. Al-Radaideh, Mahmud M. Al-Zoubi;2018 9th International Conference on Information and Communi-cation Systems (ICICS)

cwc.gov.in /Hydrological/2016-17/WYB_2016-17/10B_K&GBO_Hyderabad_Godavari_WYB_2016-17.pdf

www.wwfenvis.nic.in/Database/RealTime_4515.aspx

cwc.gov.in /Hydrological/2016-17/WYB_2016-17/Ganga_WYB_2016-17.pdf

Zhenhai Wang,Bo Xu;2017 13th International Conference on Natural Computation, Fuzzy Systems and Knowledge Discovery (ICNC-FSKD 2017)

cpcb.nic.in/water-quality-monitoring-of-drai

Guidelines for Water Quality Monitoring/CPCB/ MINARS/27/2007-08

Water Quality Monitoring System for River Ganga'/NMCG
R.M.Bhardwaj, IWG-Env, International Work Session on Water Statistics, Vienna, June 20-22 2005 cseindia.org/cpcbs-real-time-water-quality-monitoring—458

wris.nrsc.gov.inmowr.gov.in

Central water commission and Central Ground water Board annual reports

		TABLE I
		WATER QUALITY CRITERIA BY CPCB
Designated best use	Class	Criteria
Drinking water source without conventional	A	
treatment but after disinfections		1. Total coliform organisms MPN/100mL shall be 50 or less.
		2. pH between 6.5 and 8.5. Dissolved oxygen 6 mg/l or more.
		3. Biochemical oxygen demand 2 mg/l or less.
Outdoor bathing (organized)	B	
		1. Total coliform organisms MPN/100ml shall be 500 or less.
		2. pH between 6.5 and 8.5.
		3. Dissolved oxygen 5 mg/l or more.
Drinking water source with conventional	C	
treatment followed by disinfection		1. Total coliform organisms MPN/ 100ml shall be 5000 or less.
		2. pH between 6 and 9
		3. Dissolved oxygen 4 mg/l or more. Biochemical oxygen demand 3 mg/l or Less
Propagation of wild life, fisheries	D	
		1. pH between 6.5 and 8.5.
		2. Dissolved oxygen 4 mg/l or more.
		3. Free ammonia (as N) 1.2 mg/l or less
Irrigation, industrial cooling, controlled	E	
waste disposal		1. pH between 6.0 and 8.5 .
		2. Electrical conductivity less than 2250 micro mhos/cm.
		3. Sodium absorption ratio less than 26.
		4. Boron less than 2mg/l.

TABLE II

TABULAR FORM OF CPCB MONITORING SYSTEM

Steps taken in procedure
Step 1 :Setting Water Quality Monitoring Objectives
Step 2 :Assessment of Resources Availability
Step 3 :Reconnaissance Survey
Step 4 :Network Design
Step 5 :Sampling
Step 6 :Laboratory Work
Step 7 :Data Management
Storage
Statistical analysis and big data anaysis
Presentation
Interpretation
Reporting
Step 8 :Quality Assurance

