JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue

JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)

An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Predicting The Quality Of Drinking Water Using Machine Learning

[1]Mrs A.Porselvi, [2]Adelin Kiruba S, [3]Malavika R,[4]Pavithra R

[1]Assistant Professor, [2]Student, [3]Student,[4]Student
[1,2,3,4] Department of Computer Science and Engineering
[1,2,3,4] Panimalar Institute of technology, Chennai, Tamil Nadu, India

Abstract : Water contamination a general rule alludes to the arrival of toxins into water that are unsafe to human well being and, thus, to the earth overall. It is regularly alluded to as one of the most hazardous dangers that mankind has at any point confronted. It hurts creatures, yields, and woodlands, in addition to other things. To stay away from this issue, AI strategies should be utilized to expect water quality from foreign substances in the transportation area. Subsequently, evaluating and gauging water quality has turned into a basic exploration field. The objective is to explore AI based answers for water quality gauging with the most noteworthy precision through expectation.The administered AI strategy (SMLT) is utilized to examine the dataset to catch a few snippets of data, like variable distinguishing proof, uni-variate examination, bi-variate and multi-variate investigation, missing worth medicines, and information approval, information cleaning/getting ready, and information representation. Our examination gives a definite manual for model boundary awareness investigation in association with execution in water quality contamination expectation by precision computation. To foster an AI based procedure for viably foreseeing the Water Quality Index esteem by forecast, looking at directed characterization AI calculations yields the best precision. To analyse and examine the exhibition of different AI calculations utilizing the given vehicle traffic office data set, recognize the disarray framework, and sort information by need; subsequently, the adequacy of the proposed AI calculation method will be assessed by contrasting and the best exactness with accuracy, Recall, and F1 Score.

Keywords- **Water quality, Data set, Machine Learning, Accuracy**

I. Introduction

After air, water is undoubtedly the most valuable natural resource. Regardless of the way that water covers most of the world's surface, simply a little level of it is drinkable.Customary investigations of water sources are needed to ensure that the water is protected to polish off. Poorly maintained water bodies indicate environmental degradation. Subsequently, water quality is basic as far as both the climate and the economy.It is impossible to overestimate the significance of water quality analysis. After years of investigation, certain standard methodologies for measuring water quality have been discovered. There are a few well-established methods for assessing water quality. These protocols were discovered after years of research. There are a slew of rules for determining water quality. Water quality investigation estimates the substance, physical, and natural attributes of water determined to suggest safe use as per globally perceived guidelines. The water quality analysis is carried out using some established methods. As a result, techniques for sampling, maintaining samples, and analyzing samples in line with needed requirements have been developed. AI is a part of man-made brainpower that is arising as a science. Due to the developing measure of information, AI depends on the reason of tracking down the most straightforward model for the new information among the earlier information. The objective of this exploration is to show scholastics AI, which has been progressively famous as of late, just as its applications. AI calculations are used to shrewdly analyze this information and produce the suitable genuine applications. Directed, unaided, semi-regulated, and support learning are probably the most widely recognized types of learning calculations. The character and qualities of data, as well as the performance of the training algorithms, determine the effectiveness and hence efficiency of a machine learning solution.

II. RELATED WORK

There is high importance in drinking water safety and security as it renders high impact on public health and life. Many researchers are seeking different ways and techniques to provide good quality of water to ensure people's health[2]Customary investigations of water sources are needed to ensure that the water is protected to polish off. Poorly maintained water bodies indicate environmental degradation.. There is a slew of rules for determining water quality. Water quality investigation estimates the substance, physical, and natural attributes of water determined to suggest safe use as per globally perceived guidelines. The water quality analysis is carried out using some established methods[1]. In a survey done in the log one valley to assess the physical chemical quality of water sources. It identified chemical quality water concerned with critical and direct intervention. Only by the source of contamination it is possible to implement proper solution to the quality issues[3]. pH measuring scale is used to measure the water pH in a sub-sample of collected drinking water. The samples were collected from rural Bangladesh out of 24 upazilas 12 of the samples were slightly alkaline (pH 7.4 ± 0.4)[4]. Water is a vital resource. Risk-free drinking water is a primary need for Mankind and its human's basic right [5]. A study was done to understand the status of Chandrapur which is the Fourth most populated city in India. Growth of Industrial area increased anthropogenic activities. Domestic wastes are dumped into various water bodies without and proper treatment. This affects the human health ultimately[6]. Since water quality is related to physical, chemical and biological property of water there is high chance of these parameters getting affected by pollution. Depending on the activities pollutants are disposed into water bodies[7]. During a study in Kerala Minerals and Metals area, Chavara, Quilon district Bacterial contamination was found in the well using the Coliform test.Effective maintenance is required to monitor the quality of the water[8]. The centralization of SS, the likelihood of inorganic in SS, and the rates of iron dynamically expanded as the far off from the storage compartment fundamental or water conveyance tank, that is, close to the impasse of a primary, developed. Changes in SS fixations and basic arrangements, rather than lingering chlorine focuses, might be more delicate and exact signs of water quality decay in water mains. The monotonic SS increment towards a vitally's impasse was not viewed as in one significant, where SS focuses expanded in the center, then, at that point, step by step diminished before bit by bit expanding closer the super's impasse [9]. The analysis of well water is most vital for human health risk and irrigation purpose[10].

Around the world, the weight on accessibility of perfect and new water assets is broadly expanding because of populace development, fast industrialization and financial turn of events [11]. The nature of the water we drink is significant since it has extraordinary general well being concern. It is a significant danger cause for significant relative rate of diarrheic sicknesses in Nepal. Out of 84 water samples that were collected, pH of 15.48% samples was found to be on top of the WHO permissible guideline values. Parallelly, arsenic value of 85.71% samples was found to be on top of WHO value [12]. From Ambika river in Gujarat a decision of legitimate arrangement of info factors from all attainable information factors during AI model advancement is significant for procuring great model were found [13]. The wisdom which clarifies whether an environmental factor has an influence on the health of mortal beings started as epidemiology. Statistically that in the region where water force with sand filtration was used where the prevalence of cholera was lower than in other regions[14]. Retrogression analysis done in Moradabad,India suggests the conductivity of drinking water is an important parameter and it's significantly identified with ten parameters out of twelve water quality parameters studied [15]. In areas where population density is high and mortal use of the land is acute, ground water becomes especially vulnerable. Nearly any exertion whereby chemicals or wastes may be released to the terrain, either deliberately or accidentally, has the implicit to contaminate ground water [16]. The EPA gave surrounding water quality standards in 2012 for sporting waters for two list of bacterial waste defilement: Escherichia coli and streptococci [17]. The Guidelines lay out numerical "guideline values" for elements of water or water quality indexes, as well as describe appropriate minimum requirements for safe practise to protect consumers' health. It is prudent to break down rules with regards to homegrown or common, social, money related, and social variables to decide mandatory limits[18]. The most frequent standards for evaluating water quality are ecosystem health, mortal contact safety, and drinking water. Various plots were analysed and compared throughout the design process[19]. To drive the advancement of refining water assets utilizing advances and guidelines redesigned water treatment process are done to further develop the water quality to give better wellspring of water has turned into a typical exercise worldwide. Based on the impurities present different advances and blends are utilized for drinking water production[20].

Water is the main asset for people to get by in this world. In spite of the fact that water covers 71% of its world's surface, just 2.5 percent is new and drinkable. Modern contamination, marine unloading, radioactive waste, underground stockpiling spillage, an unnatural weather change, and different variables are dirtying the water supply. Individuals' health suffers immediately as a result of water contamination[21]. Currently, the quality of the water is determined through expensive laboratory tests. As a result, an alternate technique that is both efficient and cost-effective must be chosen. In the subject of resource management, several research projects are underway to improve river water quality[22]Water covers the majority of the earth, but only a little amount of it can be consumed. Because water is used for so many things, it's critical to keep track of its purity[23]. The Water Quality Index (WQI) and its sub-files are utilized to recognize the primary contamination sources that cause water quality corruption[24]. Human health depends on the use of clean and safe water, hence supply of safe potable water supply is crucial. Among the numerous water sources, groundwater is the safest to consume. The groundwater, on the other hand, becomes increasingly polluted as the population and industrialisation rise[25]. Nearly 3.4 people around the world died due to water related diseases. Waterborne illness became a result of a lack of monitoring. Human health is harmed as a result of changes in pH levels[26].The increment in contamination in the dissolve area came about because of sewage and modern effluents. To organize examining and understanding, the examples are gathered from across the city and afterward isolated into four zones of erode district[27]. Water is the subject of many investigations. To give a practical and viable drinking water, it is important to acquire reliable outcomes[28]. Lakes and reservoirs are some main source of water. But the water quality is influenced by several factors such as anthropogenic activities, disposal of sewage and industrial waste. It is important to monitor reservoirs to stop exploiting aquatic resources. The exploitation and degradation of water resources has risen as a result of increasing population and urbanization.This leads to contamination of water resulting to waterborne diseases as well[29]. Water availability, both in quantity and quality, is critical for humanity. The majority of the population relies on surface water for drinking. Water is frequently exploited for residential, commercial, and industrial purposes. The majority of water sources are hazardous to use due to pollution. The groundwater supply provides a safe and dependable source of drinking water[30].

III. MATHEMATICAL BACKGROUND

Four types of algorithms are compared to know the best accuracy for predicting the water quality.K-fold cross validation is used to compare and evaluate each algorithm.Confusion matrix is used in each algorithm the performance of the each classification model can be described using this matrix.

True Positive Rate(TPR) = TP / (TP + FN); False Positive rate(FPR) = FP / (FP + TN)

Accuracy = (TP + TN) / (TP + TN + FP + FN)

Precision = TP / (TP + FP)

Recall = TP / (TP + FN)

Measure = 2TP / (2TP + FP + FN)
F1 Score = 2*(Recall * Precision) / (Recall + Precision)

IV. ARCHITECTURAL DIAGRAM

Figure 1 At first the data set is collected i.e. the parameter requires to predict the quality of the water is predicted. The predicted values for each parameter present in water is stored in the dataset. The values are processed and cleaned using data mining. The water data set consist of the value that are safe to consume by the people based on the records previously stored. For the Input Information a raw data is given as input. The data is pre-processed using Machine Learning Technique. The values of the parameters acquire and recorded are compare using the ML algorithm. The comparison is done to obtain a best accuracy of the result. Finally, after classification of the model, result is obtained.

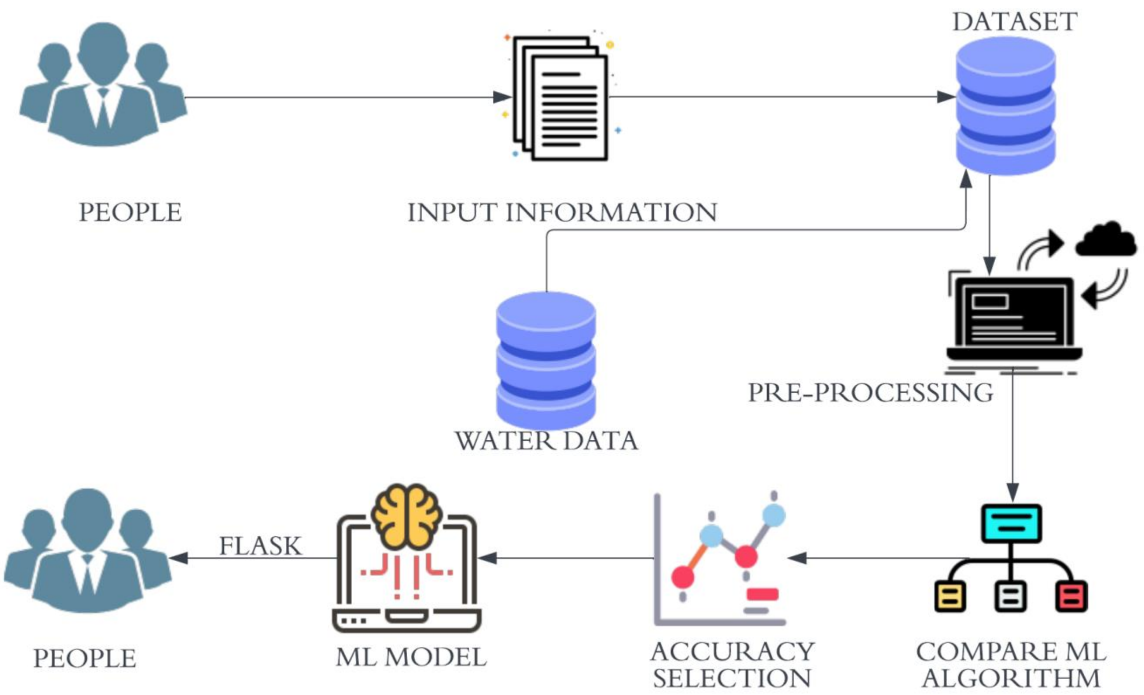

Fig. 1 System flow data

V. MODULES

Data Pre-processing Technique

Machine learning validation techniques are utilized to get the forecast mistake of the Machine Learning (ML) model, which can be expected as near the genuine blunder pace of the dataset as could really be expected. Information assortment, information investigation, and the most common way of managing information content, quality, and design can all amount to an extended plan for the day. Understanding your information and its properties is useful during the information distinguishing proof cycle; this data will help you pick which calculation to use to fabricate your model. The primary goal of information cleaning is to identify and remove errors and anomalies from data so that it may be used for examination and navigation. An assortment of information cleaning errands are performed utilizing Python's Pandas library, with a specific accentuation on the most troublesome information cleaning task, missing information, and the capacity to clean information all the more rapidly. It likes to invest less energy cleaning information and additional time testing and displaying. The means and philosophies for cleaning information will vary contingent upon the dataset. Information is given as contribution, with the normal result being the evacuation of uproarious information.

Data Analysis of Virtualization.

Data visualization is a significant ability in applied insights and machine learning. Measurements truly does for sure zero in on quantitative depictions and assessments of information. Data gives a significant set-up of instruments for acquiring a subjective arrangement. In this module Data is given as input and Visualized data is expected as output. Pre-processing refers to the modifications made to our data before it is used in the calculation. Data Preprocessing is a method of transforming raw data into a flawless informative index. When data is obtained from numerous sources, it is usually arranged in a chaotic fashion that isn't practical for evaluation. The information should be properly coordinated to achieve better results from the applied model in Machine Learning strategy. Some Machine Learning models require data to be organised in a specific way; for example, the Random Forest calculation rejects invalid attributes. As a result, when doing the arbitrary woods computation, invalid characteristics should be ruled out from the start. Another idea is that the data collection should be automated so that various Machine Learning and Deep Learning calculations can be run on the equivalent dataset.

Comparing Algorithm for high accuracy

It is imperative to think about the exhibition of different AI calculations reliably, and this theme will tell you the best way to foster a test tackle in Python utilizing scikit-figure out how to do as such. You might acquire a thought of how exact each model is on concealed information utilizing resampling approaches like cross approval. It should have the option to use these assessments to choose a couple of the best models from the set you've constructed. The way in to a reasonable correlation of AI calculations is to guarantee that every technique is assessed similarly on similar information, which can be cultivated by convincing every calculation to be assessed similarly on a similar data. Each strategy is assessed utilizing the K-overlay cross approval method. Prior to looking at calculations, introduce Scikit-Learn libraries and construct an AI model. Pre-processing, direct model with strategic relapse strategy, cross approval with KFold technique, gathering with arbitrary timberland strategy, and tree with choice tree classifier are completely remembered for this library bundle. Moreover, the train set and test set have been arranged.

Deployment using flask framework

Flask is a micro web framework written in Python. Flask is classified as a micro-framework because it does not require any specific devices or libraries. It lacks an information base reflection layer, structure approval, and other components that rely on third-party libraries to perform routine operations. Extensions, on the other hand, can be used to extend application functionality as if they were built into Flask itself. Extensions are available for object-relational mappers, form validation, upload handling, several open authentication protocols, and other framework-related features. Armin Ronacher of Pocoo created the flask framework in 2004 from a worldwide organisation of Python fetishists. When Ronacher and Georg Brand built a bulletin board system in Python, the Pocoo projects Werkzeug and Jinja were born. In April 2016, the Pocoo team disbanded, and the development of Flask and related libraries was transferred to the newly formed Pallets project.

VI. RESULTS AND DISCUSSION

The analytical process started by mining the data and by processing it. Then the values missing are found and explanatory analysis is done. In the end the model is constructed and evaluated. The advisable quality on public data set is primed higher accuracy score is will be found out. This application result to find the Water Quality status using the machine learning technique. The best precision value is yield using the AI. The processed values are compared and checked for accuracy to classify the model. Using the flask web framework, the result is provided to know the quality of the water.

VI. CONCLUSION

The water quality index determines the quality of one of the most vital resources for survival: water. There are various ways to check the quality of water, and there will be many more in the future. These are some of the approaches for predicting water quality and they are mentioned above. To check the quality of water in the past, one had to go through an expensive and time-consuming lab analysis. This project looked into a different way of utilizing machine learning to forecast water quality using only a few simple water quality measures. To work out the water quality record, a set of sample supervised machine learning algorithms was used.

REFERENCES

[1] Chacko, S., & Tom, T. (2016). Analysis of water quality of samples collected from Thevara Region, Kerala, India.

[2] Tsitsifli, S., & Kanakoudis, V. (2017, June). Drinking water quality and safety assessment–a review. In *Proceedings of the 6th International Conference on Environmental Management, Planning, Engineering (CEMEPE2017), Thessaloniki, Greece* (pp. 25-30).

[3] Sorlini, S., Palazzini, D., Sieliechi, J. M., & Ngassoum, M. B. (2013). Assessment of physical-chemical drinking water quality in the Logone Valley (Chad-Cameroon). *Sustainability*, *5*(7), 3060-3076.

[4] Akter, T., Jhohura, F. T., Akter, F., Chowdhury, T. R., Mistry, S. K., Dey, D., ... & Rahman, M. (2016). Water Quality Index for measuring drinking water quality in rural Bangladesh: a cross-sectional study. Journal of Health, Population and Nutrition, 35(1), 1-12.

[5] Meride, Y., & Ayenew, B. (2016). Drinking water quality assessment and its effects on residents health in Wondo genet campus, Ethiopia. Environmental Systems Research, 5(1), 1-7.

[6] Pratiksha Tambekar, Pravin Morey, R. J. Batra and R. G. Weginwar.Quality assessment of drinking water: A case study of Chandrapur District (M.S.).Journal of Chemical and Pharmaceutical Research, 2012, 4(5):2564-2570 Research Article ISSN : 0975-7384 CODEN(USA) : JCPRC5

[7] Sneha S. Phadatare and Prof. Sagar Gawande (Guide).Review Paper on Development of Water Quality Index.International Journal of Engineering Research & Technology (IJERT) ISSN: 2278-0181.Vol. 5 Issue 05, May-2016.

[8] C. Shaji, H. Nimi and L. Bindu.Water quality assessment of open wells in and around Chavara industrial area, Quilon, Kerala.Journal of Environmental Biology September 2009, 30(5) 701-704 (2009)©Triveni Enterprises, Lucknow (India)

[9] Matsui, Y., Yamagishi, T., Terada, Y., Matsushita, T., & Inoue, T. (2007). Suspended particles and their characteristics in water mains: developments of sampling methods. *Journal of Water Supply: Research and Technology—AQUA*, *56*(1), 13-24.

[10] Jaishree Karmore ., Dr. Rajni kant , Vrushali Karmore., . Assessment of Qualitative Water Analysis of Groundwater from Walgaon Village, District Amravati, Maharashra State by Using Different Parameters 2021 JETIR March 2021, Volume 8, Issue 3.

[11] Ibrahim, M. N. (2019). Assessing groundwater quality for drinking purpose in Jordan: application of water quality index. *Journal of Ecological Engineering*, *20*(3).

[12] Aryal, J., Gautam, B., & Sapkota, N. (2012). Drinking water quality assessment. *Journal of Nepal Health Research Council*, *10*(22), 192-196.

[13] PREDICTION OF WATER QUALITY PARAMETER OF AMBIKA RIVER BY ARTIFICIAL INTELLIGENCE BASED MODELS”", International Journal of Emerging Technologies and Innovative Research (www.jetir.org), ISSN:2349-5162, Vol.5, Issue 4, page no.439-445, April-2018

[14] Magara, Y. (2002). Classification of water quality standards. *Water Quality and Standards*,

[15] Kumar, N., & Sinha, D. K. (2010). Drinking water quality management through correlation studies among various physicochemical parameters: A case study. *International journal of environmental sciences*, *1*(2), 253-259

[16] Bedient, P. B., Rifai, H. S., & Newell, C. J. (1994). *Ground water contamination: transport and remediation*. Prentice-Hall International, Inc..

[17] Stephan, C. E., Mount, D. I., Hansen, D. J., Gentile, J. H., Chapman, G. A., & Brungs, W. A. (1985). *Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses* (p. 98). Washington, DC: US Environmental Protection Agency.

[18] Edition, F. (2011). Guidelines for drinking-water quality. *WHO chronicle*, *38*(4), 104-108.

[19] Shrivastava, S. (2014). *Water quality analysis of water bodies of Kantajhar Basti* (Doctoral dissertation).

[20] Liu, G., Zhang, Y., Knibbe, W. J., Feng, C., Liu, W., Medema, G., & van der Meer, W. (2017). Potential impacts of changing supply-water quality on drinking water distribution: A review. *Water research*, *116*, 135-148.

[21] Taruna Juneja and Alankrita Chaudhary.Assessment of water quality and its effects on the health of residents of Jhunjhunu district, Rajasthan:A cross sectional study.Journal of Public Health and Epidemiology Vol. 5(4), pp. 186-191, April 2013 DOI: 10.5897/JPHE12.096 ISSN 2006-9723 ©2013 Academic Journals

[22] GasimHayder, Isman Kurniawan Hauwa Mohammed Mustafa. Implementation of Machine Learning Methods for Monitoring and Predicting Water Quality Parameters Article Volume 11, Issue 2, 2021, 9285 - 9295

[23] Ritabrata Roy.An Introduction to Water Quality Analysis.International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056,Volume: 06 Issue: 01 | Jan 2019

[24] Andreea-Mihaela Dunca.Water Pollution and Water Quality Assessment of Major Transboundary Rivers from Banat (Romania),Hindawi Journal of Chemistry,Volume 2018, Article ID 9073763, 8 pages

[25] Roohi Rawat and A. R. Siddiqui.Assessment of Physiochemical Characteristics of Drinking Water Quality in Allahabad Metropolitan City, India.The Oriental Anthropologist 19(1) 121–135, 2019© 2019 Oriental Institute of Cultural and Social Research and SAGE

[26] Prajakta Patil , Sukanya More, Atharv Deshpande, Harshal Todkar,Sanjeev Wagh.Water Quality Monitoring for Disease Prediction using Machine Learning.International Journal of Future Generation Communication and Networking Vol. 13, No. 3, (2020), pp. 1240–1248.

[27] Arivoli Appavu, Sathiamoorthi Thangavelu, Satheeshkumar Muthukannan,Joseph Sahayarayan Jesudoss and Boomi Pandi.Study of water quality parameters of cauvery river water in erode region.Journal of Global Biosciences,ISSN 2320-1355.Volume 5, Number 9, 2016, pp. 4556-4567

[28] Peiyue Li1 and Jianhua Wu.Drinking Water Quality and Public Health.Exposure and Health (2019) 11:73–79.Received: 16 January 2019 / Revised: 16 January 2019 / Accepted: 21 January 2019 / Published online: 4 February 2019 © Springer Nature B.V. 2019.

[29] Archana Solanki,Himanshu Agrawal,Kanchan Khare. Predictive Analysis of Water Quality Parameters using Deep Learning International Journal of Computer Applications (0975 – 8887) Volume 125 – No.9, September 2015.

[30] Rajesh Prajapati and Ram Bilas.Determination of water quality index of drinking water in varanasi district, up, India.Journal of Scientific Research Vol. 62, 2018 : 1-13. Banaras Hindu University, Varanasi ISSN : 0447-9483.