

# Ground Water Data Analysis using Data Mining: A Literature Review

P. Tamilarasi, D. Akila

*Abstract--- Data mining is the process of discovering patterns from hidden datasets. Data mining tools are most widely used in ground water quality prediction. Most of the agriculture field relies on ground water. The accurate prediction of ground water quality may help in growth of agriculture sector. There are number of techniques for predicting Ground water quality and Ground water levels such as regression analysis, clustering algorithms. There are many classification algorithms used in data mining. The appropriate use of classification algorithm will enhance the prediction of water quality easy and accurate. This paper conducts literature survey on recent researches in this field up to date. The study reviews on the techniques of analysing ground water data to develop proper models for improving the quality and prediction of ground water.*

*Keywords---Ground Water Quality, prediction, Data Mining, Classification Techniques, clustering algorithms.*

## I. INTRODUCTION

Presently a-days ground water assumes an imperative job in the vast majority of the populace's water utilization. It is important to foresee ground water quality for the successful utilization of water assets. Exact forecast of the ground water quality will improve the utilization of water assets. The quality alludes to the substance, physical, natural and radiological attributes of water. The nature of water can be investigated by utilizing Water Quality Index (WQI). The nature of ground water changes as indicated by physio-substance qualities, including all out alkalinity (Al), Temperature, pH, Electrical Conductivity (EC), Total broke down solids (TDS), Turbidity. The ground water tests are gathered from various areas and the gathered ground water information dissected utilizing information mining strategies. The consequences of investigation are contrasted and some national and global principles or rules for anticipating water quality. The information mining apparatuses and methods are generally utilized for ground water quality appraisal. Some of the classification algorithms of Data mining are,

### A. Decision Trees

A choice tree is flowchart-like tree structure, in this each inner hub signifies a test on partner property, each branch speaks to relate result of the test, and each leaf hub holds a class name. The highest hub in a tree is that the root node. The improvement of choice tree classifiers does not require any area information or parameter setting. Choice trees will deal with both clear cut and information.

### B. Naïve Bayes Classifier (NB)

Naïve Bayes classifier is a probabilistic classifier based on Bayes' Theorem... It can predict class membership based on assumptions between the features. It is easy to build and helps to classify the given input over a set of classes using probability distribution.

### C. K-Nearest Neighbor Method (KNN)

Nearest neighbour classifiers rely upon feature similarity, that is, by differentiating a given test tuple and getting ready tuples that are moderately similar to it. The arrangement tuples are addressed by n characteristics. Each tuple addresses a point in a n-dimensional space. Exactly when given a cloud tuple, a k-nearest neighbor classifier looks the model space for the k-planning tuples that are closest to the darken tuple. These k-getting ready tuples are the k-"nearest neighbors" of the darken tuple. This technique is used for gathering or backslide.

### D. Rule based classification

Guidelines are a decent technique for speaking to data or bits of information. A standard based classifier utilize a lot of IF-THEN guidelines for order.

### E. Support Vector Machine (SVM)

Support Vector Machine algorithm, is a technique for the grouping of both straight and nonlinear information. It utilizes a nonlinear mapping to change the first preparing information into a higher measurement. The fundamental goal of the help vector machine calculation is to find a hyper plane in a N-dimensional space(N—the number of highlights) that especially arranges the information focuses.

## II. LITERATURE SURVEY

In the year 2018 S.V.S Ganga Devi[1] has proposed ground water quality data analysis using classification techniques . Bayesian methods have been investigated. Artificial Neural Networks (ANN)s were used to develop model for prediction and monthly values of water quality parameters. The study was carried out in Kadapa located in A.P, There were 58 water samples collected and analysed in laboratory for physico-chemical properties, conditional inference tree technique was used for analysis. R language was used to perform the water data analysis for accuracy assessment.

**Revised Version Manuscript Received on 22 February, 2019**

**P. Tamilarasi**, Ph.D Research Scholar, Department of Computer Science, VELS Institute of Science, Technology & Advanced Studies, Chennai, Tamilnadu, India. (e-mail:tamilmalu@yahoo.com)

**D. Akila**, Associate Professor, Department of Information Technology, School of Computing Sciences, VELS Institute of Science, Technology & Advanced Studies, Chennai, Tamilnadu, India. (e-mail:akiindia@yahoo.com)

In the year 2018 JyotiBansaland A.K. Dwivedi [2] proposed the study of water quality using the Water Quality Index (WQI) assessment and physico-chemical parameters for the ground water of Indore. Some of the parameters used for water quality index were: electrical conductivity (EC), pH, total hardness, COD, total alkalinity, TDS, turbidity, sulphate, chloride. After calculating WQI, the results were compared with IS: 10500-2012. The results of water quality assessment show that the parameters were high in wet season than summer season.

In the year 2017 Javadi.S and Hashemy.M. [3] proposed ground water vulnerability assessment using vulnerability maps. They have used DRASTIC index. K-means clustering technique was applied in four features of depth to ground water. They have discussed about advantages and disadvantages of DRASTIC method. The study area of the project was Hashtgerd plain located in central part of Alborz Province. There were 4936 samples points data were used in this study.

In the year 2017 Ajinkya P. Chatur, Prof. R. V. Mante, Amol R. Dhakne[4] has proposed the review of various procedures for forecast of ground water level and ground water quality. The methods discussed for ground water level prediction were SVMs(Support Vector Machines), KNN (K-Nearest Neighbor), CBA (Classification Based on Association rule), NB (Naïve Bayes), The techniques discussed for ground water quality and ground water levels were ANN(Artificial Neural Networks), NB, Decision Tree, FNN(Fuzzy Neural Networks) and Back Propagation Neural Network. They concluded that CBA out performed when compared to other methods in terms of their performances.

In 2017 Mr.Sudhir M. Gorade, Prof. AnkitDeo, Prof. PreeteshPurohit [5] proposed investigation of a few information mining arrangement strategies. There were two stages in Classification. The initial step was a model dependent on informational index, and in second step the model was utilized to order an obscure tuple into a class mark. They examined about different order models, Decision Trees, K-Nearest Neighbor, Support Vector Machines, Naive Bayesian Classifiers, Neural Networks. They examined points of interest and burdens of these methods. They reasoned that the vast majority of the students incline toward Decision tree, Bayesian classifiers, back spread, bolster vector machines, in light of the fact that these methods use preparing tuples to build a speculation display. Some languid student like closest neighbor classifiers and case-based thinking. These procedures store preparing tuples in example space and hold up until gave a test tuple

In the year 2016 Priya, Dr. R. Mallika[6] has used linear regression to predict water quality to forecast the crop yield. The study area of this project was Noyyal river basin. Water samples were collected from 12 different places. The data model for TDS using Spatio-Temporal data of Electrical Conductivity (EC) was developed. The source of the data was collected from TWAD(Tamilnadu Water supply And Drainage Development) board for the years 2013 to 2016. Linear Regression data model for TDS using Mg, Na, Cl, EC also developed in this project.

In the year 2016MohdSaleem, AtharHussain and GauharMahmood [7] has proposed the investigation of examining the rground water nature of Greater Noida area by utilizing Water Quality Index (WQI). Some of Nine physico-synthetic parameters, for example, Nitrate, Calcium, Alkalinit, Magnesium, Chloride, Sulfate, Total Hardness, Fluoride and TDS have been decided for examination. The water tests were gathered from ten unique zones from a time of 2015. Water Quality Index was utilized to give data on the nature of any water body. Numerical condition of WQI changes vast number of water quality information into a solitary number. The benefits and negative marks of water quality file were talked about. In this examination, the scientists found that 10% water tests were discovered low quality, 90% water tests were in great quality.

In the year 2016 According to Faisal Aburub, WaelHadi [8] Jordan was one of the water-poorest nations on the planet. Their exploration went for examination of four information mining strategies CBA, SVMs, NB and KNN with respect to order exactness, accuracy and F1 to foresee ground water regions in Jordan. Datasets of 900 ground water zones had been chosen and a test has a place with two classes ("Yes" or "No) containing seven highlights to separate ground water regions. In their examination they inferred that SVMs calculation beat every other calculation as for the order.

In the year 2015 A. Shanmugasundharam, G. Kalpana, S. R. Mahapatra, E. R. Sudharson, M. Jayaprakash [9] have assessed the ground water for drinking and irrigation purpose in Krishnagiriand Vellore Districts. They have collected 31 ground water samples and assessment carried out using physico-chemical components such as EC, TDS,  $\text{HCO}_3^-$ ,  $\text{Mg}^{2+}$ ,  $\text{Cl}^-$ ,  $\text{SO}_4^{2-}$ , pH,  $\text{Ca}^{2+}$ ,  $\text{Na}^+$  and  $\text{K}^+$ . International and National standards are compared with analytical results for all the components. They carried out correlation coefficient analysis for groundwater quality parameters of Krishnagiri district and Vellore district. The results of Wilcox diagram showed that, 6 percent water samples were in the good to permissible condition, 52 percent of the water samples were in the very good to good condition, 32 percent of the water samples were in the permissible to doubtful condition and only one sample was in the doubtful to unsuitable condition.

In the year 2014Shoba G, Dr.Shobha G [10] In this study they focused on reviews of different research papers which are related to data mining algorithms used in groundwater quality prediction. After study, they have concluded that, Back Propagation was used frequently.

In the year 2013 Kamakshaiah. Kolli, R. Seshadri[11] proposed assessment of ground water quality using data mining techniques. The study area was Tadepalli in Guntur district. About 40 Water samples were collected from the bore wells and open wells in the study area. The physico-chemical parameters like TH, TDS,  $\text{NO}_3$  and  $\text{Cl}^-$  were analysed. The results were compared with standard techniques. Based on several water quality parameters,



Water Quality Index was also calculated to get a single value that denotes the overall water quality at a certain area and time. This research on groundwater quality using F-concentration shows that the ground waters were in medium to very hard category and mostly blackish. The value of F-content was almost higher than the maximum permissible limit in 45percent of total water samples.

In the year 2012 Sarala C, Ravi Babu P [12] proposed the investigation of groundwater quality evaluation in the drag wells of Jawaharnagar, in upper Musi catchment territory of Ranga Reddy locale in Andhra Pradesh. The examples from bore wells were gathered from the region for two season's pre rainstorm and post storm in December 2007 and June 2008. The groundwater investigation was done by Arc GIS programming. The quality examination has been improved the situation the parameters like aggregate alkalinity, electrical conductivity, pH, TDS, calcium hardness, magnesium hardness, nitrites, sulfates, chlorides and fluorides. They have inferred that in the whole territory groundwater was dirtied. The investigation saw that the utilization of surface and groundwater for horticultural purposes, drinking and modern use has been expanded yet therefore it was seen that the water was contaminated and influenced soil supplements, the human wellbeing, biomass

and condition in specific territories. In the majority of the areas the water was unsatisfactory for drinking reason.

### III. CONCLUSION

This paper gives brief survey on the various methods and techniques used in predicting the ground water quality. Researchers can propose new more effective methods to predict groundwater quality accurately by considering all these methods. Published papers from year 2012 to 2018 from online research libraries are considered for this review. This review will be useful for the researchers to analyze the quality of ground water with focus on data mining techniques.

### IV. FUTURE ENHANCEMENTS

In future, the authors propose to use these data mining techniques for ground water data analysis can compare the data mining algorithms to find the effective and accurate algorithm for ground water data analysis. The finding of the research study may be helpful to water management board in better water management for achieving the goal of sustainable use of the water resource.

**Table 1: Summary of Survey**

Author	Coverage Area of Study	Methodology	Objectives
S.V.S Ganga Devi[1]	Kadapa, Andhra Pradesh	Conditional inference tree and random forest classification technique	Traditional decision trees are larger in size and they can be pruned whereas conditional inference trees are not pruned and accuracy is improved. Random forest model has 10.87% error which means we can predict with 89.13% accuracy.
Jyoti Bansal and A.K. Dwivedi[2]	Indore	Water Quality Index (WQI)	Some water quality parameters somewhat higher in wet season than in summer season
Javadi, S and Hashemy, M [3]	Hashtgerd plain	Vulnerability Maps, DRASTIC Model	The map shows that central parts of the field are in high risk of groundwater vulnerability
Ajinkya P. Chatur, Prof. R. V. Mante, Amol R. Dhakne[4]	Jordan.	CBA, SVMs, NB and KNN	In all of these methods, CBA outperforms other methods in terms of their performances.
R.Priya and Dr.R.Mallika[6]	NoyyalRiver Basin	Linear Regression	The model analyses the ground water quality in terms of TDS and attempts to approximate future values of data as a linear combination of parameters. It is decisively concluded that this may help the decision maker to predict the crop yield with respect to water quality before harvesting the crop.
MohdSaleem, Athar Hussain and GauharMahmood[7]	Greater Noida, Uttar Pradesh	Water Quality Index	In the examination zone, 90 percent water tests were in great quality condition and just 10 percent water tests were in reasonably poor condition. The water quality record ranges from 16.49 to 64.65.
Faisal Aburub, WaelHadi [8]	Jordan	CBA, SVMs, NB and KNN	The test results show the SVMs algorithm outperformed the other three algorithms in relation to all used measures.
Kamakshaiiah Kolli, R. Seshadri,[11]	Tadepalli	Water Quality Index	The ground water samples were alkaline and also in medium to very hard category and mostly blackish
Sarala C, Ravi Babu P.[12]	Andhra Pradesh	Physico-Chemical Parameters of Groundwater	The groundwater was acidic and also observed that total hardness parameter was in hard to very hard category. The total dissolved solids (TDS) parameter values were in fresh water to saline categories. The fluoride concentration parameter value exceeded the permissible limit.

## REFERENCES

1. F. Aburub and W. Hadi, "Predicting Groundwater Areas Using Data Mining Techniques : Groundwater in Jordan as Case Study," vol. 10, no. 9, pp. 1621–1624, 2016.
2. J. Bansal and A. K. Dwivedi, "Assessment Of Ground Water Quality By Using Water Quality Index And Physico Chemical Parameters : Review Paper," vol. 7, no. 2, pp. 170–174, 2018.
3. A. P. Chatur, P. R. V Mante, and A. R. Dhakne, "Brief Survey of Different Techniques for Prediction of Groundwater Level and Groundwater Quality," vol. 4, no. 11, pp. 168–172, 2017.
4. S. V. S. G. Devi, "Ground Water Quality Data Analysis Using Classification Techniques."
5. S. M. Gorade and P. A. Deo, "A Study of Some Data Mining Classification Techniques," pp. 3112–3115, 2017.
6. I. Journal et al., "Water Quality Prediction Using Data Mining techniques : A Survey," vol. 3, no. 6, pp. 6299–6306, 2014.
7. K. Kolli, "Ground Water Quality Assessment using Data Mining Techniques," vol. 76, no. 15, pp. 39–45, 2013.
8. S. Physics, "Evaluation of groundwater vulnerability using data mining technique in Hashtgerd plain," vol. 42, no. 4, pp. 35–41, 2017.
9. U. P. Pradesh, "Analysis of groundwater quality using water quality index : A case study of greater Noida ( Region ), Uttar," Cogent Eng., vol. 44, pp. 1–11, 2016.
10. R. Priya and R. Mallika, "Ground Water Quality Modelling for Irrigation Using Data Mining Technique and Spatio-Temporal Dates," vol. 12, no. 16, pp. 6097–6101, 2017.
11. C. Sarala and R. B. P., "Assessment of Groundwater Quality Parameters in and around Jawaharnagar , Hyderabad," vol. 2, no. 10, pp. 1–6, 2012.
12. A. Shanmugasundharam, G. Kalpana, S. R. Mahapatra, E. R. Sudharsan, and M. Jayaprakash, "Assessment of Groundwater quality in Krishnagiri and Vellore Districts in Tamil Nadu , India," Appl. Water Sci., vol. 7, no. 4, pp. 1869–1879, 2017.