ISSN: 0974-2115
www.jchps.com Journal of Chemical and Pharmaceutical Sciences

Mining analysis of iron contaminant in river water quality

S. Brintha Rajakumari*[1], C. Nalini[1], Rajkumar Samuel[2]
[1]Dept. of CSE, Bharath University, Chennai.
[2]Director - Technical, Hubert Enviro Care Systems Private Limited, Chennai.
***Corresponding author:E-mail:brintha.ramesh@gmail.com,**

ABSTRACT

Water pollution is an important issue in India. Growth and rapid development of economy, water pollution is becoming more critical. In order to prepare efficient water pollution trace and metal dataset and predict the pollution level of each state, a case study was carried out. This paper presents the study report about river water quality and analyzed the Iron metal in river and reduced the row size using PIVOT to predict the high contaminant of Iron and take necessary steps based on the report.

KEY WORDS: Data Mining, SQL Aggregation, River Water Quality, Water Pollution.

1. INTRODUCTION

Data Mining is one of the important research areas motivate to find the meaningful information from large number of datasets. At present it is popular in water quality analysis because of detecting useful information in water quality and provides lots of benefits such as finding BOD Level, trace of Metals, mining water quality and also predict the diseases based on the quality of water.

Normally industrial waste and city sewage are discharged in to the rivers. The Central Pollution Control Board (CPCB) and Tamilnadu Pollution Control Board (TNPCB) provides lot of data about river water quality which are very difficult to understand. So it is necessary to build a new dataset model to predict the exact information with data mining technique which is used to analyze trace and toxic metal in river water and predict various diseases based on the metal. The papers organized as follows: section 2 discusses the related review papers based on the river water quality analysis. Section 3 and 4 describes the study about the system and discussed the result. Finally conclude the paper.

Related Works: The paper suggests that watershed instrument to control the wastewater pollution. A watershed is defined as catchment area that ultimately drains into a particular body of water. There are four main features in this approach such as identifying and prioritizing water quality problems in the watershed, participatory planning to involve all users in problem identification and organization of suitable water quality targets, identification of most effective strategies to reach the water quality standards and coordinating activities with other agencies, and measuring achievement during increased and more efficient monitoring and other data gathering. Finding missing, duplicate and erroneous issues in data are time-consuming activity. The author developed a Profiler visual analytical tool for accessing data quality issues in tabular data.

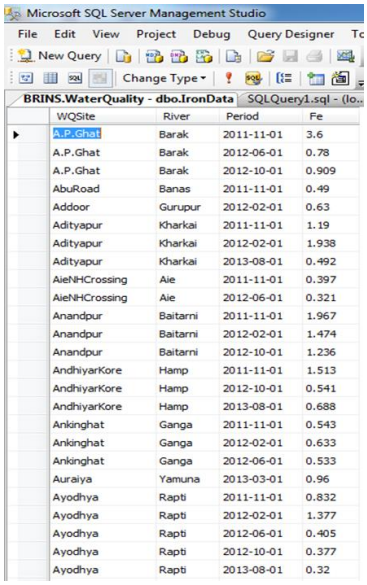
The paper presents classification and cluster analysis of total coliform in water quality. The literature review revealed that research has been conducted on the river. Therefore, there is an urgent need for a study on water quality of river which can be used for developing management policy and for improving the water quality of the river. Quality of water is determined based on its chemical, biological and physical parameters. The paper used k mean clustering algorithm with various weight values and used high dimensional and complex data matrices.

The paper present the horizontal representation of data used for data analysis of data mining which also gives better view of huge amount of data in summarized form.

2. MATERIALS AND METHODS

System Study: The data set used for this analysis is based on trace and toxic metal in river water quality which is taken from CPCB. The metal data is converted into SQL server database format before applied a necessary preprocessing technique. The dataset contain three non-numeric attribute and one numeric attribute such as Water Quality monitoring Site, name of the River, period of the water analysis and contaminant of the Iron metal. BIS (Bureau of Indian Standard) (10500, 2012) recommend upto 0.3 mg/L is an acceptable level of Iron in drinking water.

The dataset is in table 1 contains 492 observations of Iron metal which are all not fit level for river water quality in the period of Sep2011, Feb2012, June2012, Oct2012, Mar2013 and Aug2013.



WQSite	River	Period	Fe
A.P.Ghat	Barak	2011-11-01	3.6
A.P.Ghat	Barak	2012-06-01	0.78
A.P.Ghat	Barak	2012-10-01	0.909
AbuRoad	Banas	2011-11-01	0.49
Addoor	Gurupur	2012-02-01	0.63
Adityapur	Kharkai	2011-11-01	1.19
Adityapur	Kharkai	2012-02-01	1.938
Adityapur	Kharkai	2013-08-01	0.492
AieNHCrossing	Aie	2011-11-01	0.397
AieNHCrossing	Aie	2012-06-01	0.321
Anandpur	Baitarni	2011-11-01	1.967
Anandpur	Baitarni	2012-02-01	1.474
Anandpur	Baitarni	2012-10-01	1.236
AndhiyarKore	Hamp	2011-11-01	1.513
AndhiyarKore	Hamp	2012-10-01	0.541
AndhiyarKore	Hamp	2013-08-01	0.688
Ankinghat	Ganga	2011-11-01	0.543
Ankinghat	Ganga	2012-02-01	0.633
Ankinghat	Ganga	2012-06-01	0.533
Auraiya	Yamuna	2013-03-01	0.96
Ayodhya	Rapti	2011-11-01	0.832
Ayodhya	Rapti	2012-02-01	1.377
Ayodhya	Rapti	2012-06-01	0.405
Ayodhya	Rapti	2012-10-01	0.377
Ayodhya	Rapti	2013-08-01	0.32

Figure.1.Iron DataSet

3. RESULT AND DISCUSSION

Analyzed trace and toxic metal in the water quality of river dataset and applied traditional vertical aggregation function such as min, max and count to find the river which contains minimum and maximum unfit level of Iron in river water is in table.3 and 4. The table.2, contains the number of records of unfit iron content based on the period.

```
SELECT  period , COUNT(*)as TotalRecord, MIN(fe)as Minimum,MAX (fe)as Maximum
  FROM [WaterQuality].[dbo].[IronData] group by period
with temp as
(
  select *, rank() over(partition by period order by fe) as rowno
  from WaterQuality .dbo.IronData
)
select WQSite,River,Period,Fe
from temp
where rowno = 1

with temp as
(
  select *, rank() over(partition by period order by fe desc) as rowno
  from WaterQuality .dbo.IronData
)
select WQSite,River,Period,Fe
from temp
where rowno = 1
```

The rank() function is used to predict the high value of iron and the resultant table 4 contains the maximum contaminant of Iron of 6.6 in Nagavali river water in the period Nov-2011.

```
SELECT  [Wqsite], [River],  [2011-11-01], [2012-02-01],  [2012-06-01],[2012-10-01], [2013-03-01],  [2013-08-01]
FROM  [Waterquality].[Dbo].[Irondata]
PIVOT
(
     Fe
     FOR [Period] IN ([2011-11-01], [2012-02-01],  [2012-06-01],[2012-10-01], [2013-03-01],  [2013-08-01])
) AS P
```

Results Messages

	period	TotalRecord	Minimum	Maximum
1	2011-11-01	158	0.301	6.645
2	2012-02-01	113	0.304	3.675
3	2012-06-01	43	0.306	2.367
4	2012-10-01	66	0.31	3.299
5	2013-03-01	20	0.313	1.106
6	2013-08-01	92	0.31	3.184

Figure.2. No. of Minimum and Maximum of Iron unfit data

	WQSite	River	Period	Fe
1	DhengBridge	Bagmati	2011-11-01	0.301
2	TeestaBazar	Teesta	2012-02-01	0.304
3	Bareilly	Ramganga	2012-06-01	0.306
4	Dholai	Rukni	2012-10-01	0.31
5	Motinaroli	Kim	2013-03-01	0.313
6	Balrampur	Rapti	2013-08-01	0.31
7	JaiNagar	Kamla-Balan	2013-08-01	0.31
8	Sonapur	Digaru	2013-08-01	0.31

Figure.3. Minimum of Iron unfit data in river

Results Messages

	WQSite	River	Period	Fe
1	Srikakulam	Nagavali	2011-11-01	6.645
2	Srikakulam	Nagavali	2012-02-01	3.675
3	Kesinga	Tel	2012-06-01	2.367
4	Panposh	Brahmani	2012-10-01	3.299
5	Ekmighat	Adhwara	2013-03-01	1.106
6	Sarangkheda	Tapi	2013-08-01	3.184

Figure.4. Maximum of Iron unfit data in river

	WQSite	River	2011-11-01	2012-02-01	2012-06-01	2012-10-01	2013-03-01	2013-08-01
1	Ekmighat	Adhwara	0.393999993801117	NULL	0.463999998664856	NULL	1.10599994659424	0.5
2	Santeguli	Aghanashini	0.47299998998642	1.18499994277954	NULL	NULL	NULL	NULL
3	AieNHCrossing	Aie	0.397000014781952	NULL	0.321000009775162	NULL	NULL	NULL
4	Jamtara	Ajay	0.693000018596649	0.544000029563904	NULL	0.43299999833107	NULL	1.08200001716614
5	Nutanhat	Ajay	1.12100005149841	1.12100005149841	0.554000020027161	0.953000009059906	NULL	1.31599998474121
6	Rudraprayag	Alaknanda	2.61500000953674	0.75	NULL	NULL	NULL	0.970000028610229
7	Gadat	Ambika	1.36899995803833	NULL	NULL	NULL	NULL	1.25699996948242
8	DhengBridge	Bagmati	0.300999999046326	NULL	NULL	NULL	NULL	0.504999995231628
9	Hayaghat	Bagmati	0.428999990224838	NULL	NULL	NULL	0.453999996185303	NULL
10	Anandpur	Baitarni	1.96700000762939	1.47399997711182	NULL	1.23599994182587	NULL	NULL
11	Champua	Baitarni	1.0789999961853	3.32299995422363	NULL	1.23099994659424	NULL	0.312000006437302
12	Chitrasani	Balaram	0.791999995708466	NULL	NULL	NULL	NULL	NULL
13	Matigara	Balson	0.763999998569489	0.939999997615814	NULL	NULL	NULL	NULL
14	AbuRoad	Banas	0.490000009536743	NULL	NULL	NULL	NULL	NULL
15	Kamalpur	Banas	0.476000010967255	NULL	NULL	NULL	NULL	NULL
16	Bamni	Banjar	NULL	NULL	NULL	0.527999997138977	NULL	0.435999989509583
17	A.P.Ghat	Barak	3.59999990463257	NULL	0.779999971389771	0.908999979496002	NULL	NULL
18	B.P.Ghat	Barak	2.49000000953674	NULL	0.944000005722046	NULL	NULL	NULL
19	Fulertal	Barak	2.4300000667572	1.067999958992	1.27400004863739	0.386999994516373	NULL	NULL
20	BekiRoadBri...	Beki	0.550000011920929	NULL	0.370999991893768	NULL	NULL	NULL
21	Mathanguri	Beki	0.564999997615814	NULL	0.53600001335144	NULL	NULL	NULL
22	Ganod	Bhadar	1	NULL	NULL	NULL	NULL	NULL
23	Holehonnur	Bhadra	0.703999996185303	0.430000007152557	NULL	NULL	NULL	NULL
24	Tehri	Bhagirath	0.537000000476837	1.01600003242493	NULL	0.312999993562698	NULL	1.49300003051758
25	Uttarkashi	Bhagirath	1.13999998569489	0.598999977111816	NULL	NULL	NULL	1.66799998283386

Figure.5. Iron Data using PIVOT

In the table 5 contains 184 rows which are reduced from 492 observations and used SQL PIVOT command to transpose the period column of the Iron table to reduce the row size of the table. It has lot of null cell. From the result set one can easily analyze the data record of unfitable level of Iron in the above six periods and take the necessary steps to prevent the river water quality of Iron. From that official easily take decision to avoid pollution of water based on nearby industries, sewage water and any other sources to pollute the water.

4. CONCLUSION

Water pollution of river is a serious economic and public health problem. Standard monitoring of the water quality of river is required to evaluate the condition of water. This monitoring is useful in saving the river water from more degradation. From the literature study it is understandable that the water of river in India is very much polluted but still can be used for agriculture and all other activities. We have used rank and Pivot function to evaluate the dataset and presented the horizontal layout of data which will be useful identify high iron content in the river water. In future we have planned to combine all the metal and prepare the dataset.

5. ACKNOWLEDGEMENTS

I thank Mr.A. Manoharan, Environment Consultant and Ex. Sr.scientist & Zonal officer of Central Pollution Control Board for giving wonderful idea to do the research based on river water quality.

REFERENCES

Achudhan M, Prem Jayakumar M, Mathematical modeling and control of an electrically-heated catalyst, International Journal of Applied Engineering Research, 9 (23), 2014, 23013.

Brintha Rajakumari S, Nalini C, A Comparative Analysis of Horizontal Layout Representation of Data", International Journal of Advance Research In Science and Engineering, 4 (1), 2015, 984-990.

Brintha Rajakumari S, Nalini C, An Efficient Compression Method to Reduce the Data Transfer Cost in the Cloud, International Journal of Innovative Research in Computer and Communication Engineering, 3 (3), 2015, 2176-2179.

Brintha Rajakumari S, Nalini C, An efficient cost Model for data storage with horizontal layout in the cloud" , Indian Journal of Science and Technology, 7 (S3), 2014, 45-46.

Brintha Rajakumari S, Nalini C, An efficient Data Mining data Set preparation using aggregation in relational database, Indian Journal of Science and Technology, 7 (5), 2014, 44-46.

Brintha Rajakumari S, Nalini C, Preparation Of Dataset For Mining Analysis Using Cancer Dataset, Int J Pharm Bio, 6 (3), 2015, 883 – 889.

Gopalakrishnan K, Sundeep Aanand J, Udayakumar R, Electrical properties of doped azopolyester, Middle - East Journal of Scientific Research, 20 (11), 2014, 1402-1412.

Gopinath S, Sundararaj M, Elangovan S, Rathakrishnan E, Mixing characteristics of elliptical and rectangular subsonic jets with swirling co-flow, International Journal of Turbo and Jet Engines, 32 (1), 2015, 73-83.

Han J and Kamber M, Data mining: concepts and techniques, 2nd ed. The Morgan Kaufmann Series, 2006.

Hui Zou, Zhihong Zou, and Xiaojing Wang, An Enhanced K-Means Algorithm for Water Quality Analysis of The Haihe River in China, Int. J. Environ. Res. Public Health, 12, 2015, 14400-14413.

Ilayaraja K, Ambica A, Spatial distribution of groundwater quality between injambakkam-thiruvanmyiur areas, south east coast of India, Nature Environment and Pollution Technology, 14 (4), 2015, 771-776.

Ivana D Radojevic, Dusan M Stefanovic, Ljiljana R Comic, Aleksandar M Ostojic, Marina D, Topuzovic and Nenad D Stefanovic, Total coliforms and data mining as a tool in water quality monitoring,African Journal of Microbiology Research, 6 (10), 2012, 2346-2356.

Kazmi AA, Akansha Bhatia, Azfar Shaida, Meena Sharma, Markus Starkl and R. C. Trivedi,A Short Screening Study On Water Quality Of Indian Rivers, And Lakes,Resour. Soc, 33 (3), 2013

Kerana Hanirex D, Kaliyamurthie KP, Kumaravel A, Analysis of improved tdtr algorithm for mining frequent itemsets using dengue virus type 1 dataset: A combined approach, International Journal of Pharma and Bio Sciences, 6 (2), 2015, 288-295.

Lingeswaran K, Prasad Karamcheti SS, Gopikrishnan M, Ramu G, Preparation and characterization of chemical bath deposited cds thin film for solar cell, Middle - East Journal of Scientific Research, 20 (7), 2014, 812-814.

Premkumar S, Ramu G, Gunasekaran S, Baskar D, Solar industrial process heating associated with thermal energy storage for feed water heating, Middle - East Journal of Scientific Research, 20 (11), 2014, 1686-1688.

Sargaonkar A, Deshpande V, Development of an overall index of pollution for surface water based on a general classification scheme in Indian context. Environ. Monit. Assess, 89, 2003, 43–67.

Sean Kandel, Ravi Parikh, Andreas Paepcke, Joseph M. Hellerstein, Jeffrey Heer, Profiler: Integrated Statistical Analysis and Visualization for Data Quality Assessment, ACM, 2012, 21-25.

Sundar Raj M, Saravanan T, Srinivasan V, Design of silicon-carbide based cascaded multilevel inverter, Middle - East Journal of Scientific Research, 20 (12), 2014, 1785-1791.

Thooyamani KP, Khanaa V, Udayakumar R, Application of pattern recognition for farsi license plate recognition, Middle - East Journal of Scientific Research, 18 (12), 2013, 1768-1774.

Thooyamani KP, Khanaa V, Udayakumar R, Efficiently measuring denial of service attacks using appropriate metrics, Middle - East Journal of Scientific Research, 20 (12), 2014, 2464-2470.

Thooyamani KP, Khanaa V, Udayakumar R, Partial encryption and partial inference control based disclosure in effective cost cloud, Middle - East Journal of Scientific Research, 20 (12), 2014, 2456-2459.

Thooyamani KP, Khanaa V, Udayakumar R, Using integrated circuits with low power multi bit flip-flops in different approch, Middle - East Journal of Scientific Research, 20 (12), 2014, 2586-2593.

Thooyamani KP, Khanaa V, Udayakumar R, Virtual instrumentation based process of agriculture by automation, Middle - East Journal of Scientific Research, 20 (12), 2014, 2604-2612.

Thooyamani KP, Khanaa V, Udayakumar R, Wide area wireless networks-IETF, Middle - East Journal of Scientific Research, 20 (12), 2014, 2042-2046.

Udayakumar R, Kaliyamurthie KP, Khanaa, Thooyamani KP, Data mining a boon: Predictive system for university topper women in academia, World Applied Sciences Journal, 29 (14), 2014, 86-90.

Wang XJ, Zou ZH, Zou H, Water quality evaluation of Haihe River with fuzzy similarity measure methods. J. Environ. Sci, 25, 2013, 2041–2046.

Xu HS, Xu ZX, Wu W, Tang FF, Assessment and spatiotemporal variation analysis of water quality in the Zhangweinan River Basin, China. Procedia Environ. Sci, 13, 2012, 1641–1652.