Developing a Water Quality Model for Upper Lake of Bhopal Using Data Mining Methods

[1] Sanjeev Gour, [2] Dr. Mamta Gour, [3] Purushottam Sharma.

Abstract

Water is the basic need to human life and the health of our environment. Water quality is fundamental for any River or Lake Health. Lakes are very important as water sources and for fishing, recreation, tourism, and water transport. Thus for water quality protection, the water quality analysis can be used to develop management plans. Approach of this research study is to analyze the water quality data of upper lake of Bhopal city using data mining techniques like classification and association and give a model in terms of rules and relationship among various water parameters which affects the quality of lake water like BOD, COD Chloride, Nitrate, Hardness, and Alkalinity etc. It can contribute valuable support in the management programs for upper lake of Bhopal city.

Index terms: association, classification, data mining.

I INTRODUCTION

Our water resources are very important to our environmental life also have great social economic value to human life, and if the quality of water becomes degraded this resource will lose its value. Human life highly depends on water quality that is useful drinking, irrigation, watering stock and fishing and many others needs. One of the major source of drinking water for the people of the Bhopal city, is Upper lake (also called Bada Talab) which is used for serving around 40% of the residents with approximate 30 million gallons (140,000 m^3) of water per day but dumping of waste into the Upper Lake by People alongside the lake is affecting the water quality of upper lake of the city. Because of this affected quality of water of the Upper Lake, there are complaints of water-borne diseases. So the quality of water in the upper lake has gradually deteriorated due to pollution and ecological damage caused by human activities in the surrounding area of lake [1].In this study, we consider certain water quality parameters which affects the quality of water and analyzed them using data mining methods. Numerous efforts to model water quality have been made by research communities nowadays [2]. Accurate water quality classification and prediction would provide us with a better understanding the behavior of lake water and provide information for better decision making in regards to water conservation. In this Study, a data-mining approach like classification, Association for analyzing water quality is introduced and this can be support water management programs conducting for improving the quality of water of upper lake.

II STUDY AREA (UPPER LAKE)

Upper Lake also called Bada Talab lies on the Western side of the Bhopal city, capital of Madhyapradesh (India). It is one of the oldest water resource in the state which was built in the 11th century by Raja Bhoj while the construction of a dam across the Kolans River. Upper lake has a catchment area of 36.1 sq km and is surrounded by the garden, Kamla Park. This beautiful lake is a one of good source of entertainment for the locals as well as the tourists. Peoples enjoy water sport activities like boating, parasailing, and water skiing in this upper lake.

III. DATA MINING AND TECHNIQUES

Data mining uses application of database that look for hidden information and patterns in a group of data that can be used to predict future behavior of particular object. It finds previously unknown relationships among the data. In this work, the objective of data mining is to develop a model to be used in water management plan. Here we used two methods of data miming which are used for developing a model; Classification and Association.

ISSN: 2278 – 1323

Classification: Classification models classify categorical class labels and it consists of predicting a certain output on the basis of given input. In order to classify the outcome, the classification algorithm processes a training set which contains a set of attributes and the respective output values The algorithm then used to find the relationships between the attributes that can be used to predict the outcome[3].

Association rule: Association rules are IF→THEN statements that are used to find hidden relationships between unrelated data in any database. These rules are created by analyzing data for IF→THEN patterns and using the concept of *support* and *confidence* to find the most important relationships between data. *Support* is an indication of how frequently the items appear in the database. *Confidence* indicates the number of times the IF→THEN statements have been found to be true. Here we used this technique to uncover the relationship between various water quality parameters of upper lake [4].

IV. ABOUT DATA MINER

To create a classification model for water quality we used data miner software called Rapid Miner. Rapid Miner Studio 6.5.002 (Basic Edition) is free software available on the Internet. It has great capabilities to process and mined knowledge from large databases using set of Data mining techniques, methods and algorithms. (Figure 1)

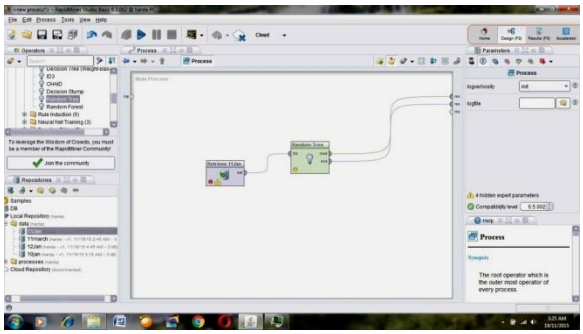

Figure 1: Application Environment of Rapid Miner.

V. DATASET DESCRIPTIONS

For this experimental work we have collected the data (water quality parameters) of surface water of upper lake of Bhopal city from eighteen stations in Duration Year 2010, 2011, 2012 (Winter + summer + rainy Months).We have approximated sixteen Parameters selected (Table 1)with their average value (in unit mg/l) in MS-EXCEL 2007 file format.(Figure 2).

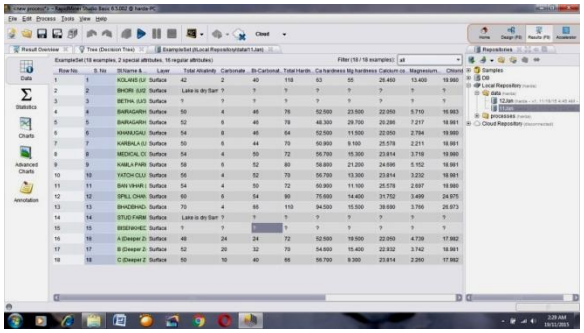

Figure 2: Explorer View for Dataset Used.

SN	Parameter	Unit
1	St.Name & No	Name of station & Number
2	Total Alkalinity	mg/L
3	Carbonate Alkalinity	mg/L
4	Bi-Carbonatealkalinity	mg/L
5	Total Hardness	mg/L
6	Ca hardness	mg/L
7	Mg hardness	mg/L
8	Calcium content	mg/L
9	Magnesium content	mg/L
10	Chloride	mg/L
11	Phosphate	mg/L
12	Total Phosphorus	mg/L
13	Org.Phosphorus	mg/L
14	Nitrate	mg/L
15	BOD	mg/L
16	COD	mg/L

Table 1: Water Quality Parameter Selected

ISSN: 2278 – 1323

VI. EXAMINED CLASS

For classification and association process we used attribute "Station Name and Number "from sixteen parameters as the examined class for classify water quality data from the training dataset (Table 2).These are eighteen stations of upper lake of Bhopal city.

S. No	St.Name & No.
1	KOLANS (U/1)
2	BHORI (U/2)
3	BETHA. (U/3)
4	BAIRAGADH (U/4)
5	BAIRAGADH EAST (U/5)
6	KHANUGAU (U/6)
7	KARBALA (U/7)
8	MEDICAL COLLEGE (U/8)
9	KAMLA PARK (U/9)
10	YATCH CLUB (U/10)
11	BAN VIHAR (U/11)
12	SPILL CHANEL (U/12)
13	BHADBHADA (U/13)
14	STUD FARM (U/14)
15	BISENKHEDI (U/15)
16	A (Deeper Zone)
17	B (Deeper Zone)
18	C (Deeper Zone)

Table 2: Station Name and Number

VII. EXPERIMENT SETUP

The Experiment work carried out in application environment of Rapid Miner. first we imported the Excel file from local repository and select the 15 parameters from dataset as regular attribute, 2 special and 1 parameter used as a Class label (examined attribute),also assign their domains ,types and then select different process of data mining like decision tree, random tree ,association rule etc as a Operator which are then applied to the example dataset in application window of Rapid Miner (figure 1), and connect the input file to the output tool and then execute the selected operator. Every execution of selected operation for given input example dataset, we have found a model with decision tree, Predictions, statistics, charts, and association rules(figure 3 to 5 and Table 3).

Interpretation of model: The Model in table -3 is in the form of IF→THEN patterning such as: IF (attribute-values1; attribute-values2; & attribute-values3 &&…) →THEN (predicted Class Attribute).

Example: From table-3, rule- *"TH=160(Max) & TA>120→U13***" *implies that* IF** attribute "Total Hardness" found 160 mg/l at its maximum value and "Total alkalinity" is greater than 120 mg/l **THEN** the predicted /classified class station is **U13** (i.e. BHADBHADA).

VIII. SOME RESULATANT SCREENSHOTS DURING EXPERIMENT.

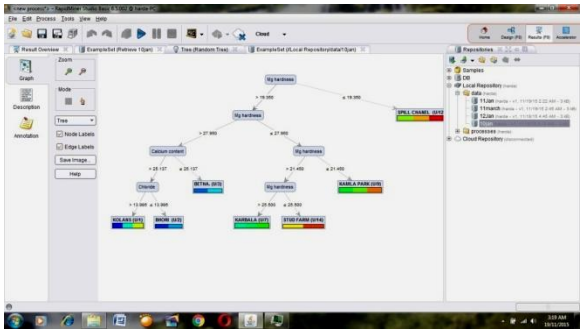

Figure 3: Random Tree generated for UL (January-2010)

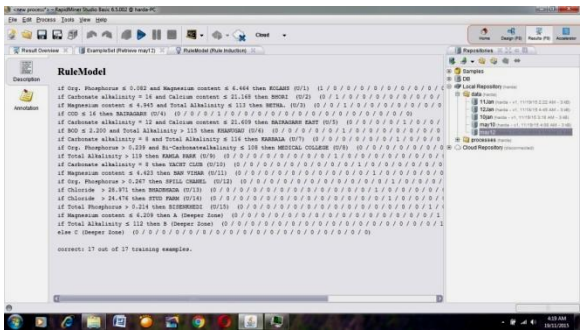

Figure 4 : Rule Model Generated for UL (May-2012)

ISSN: 2278 – 1323

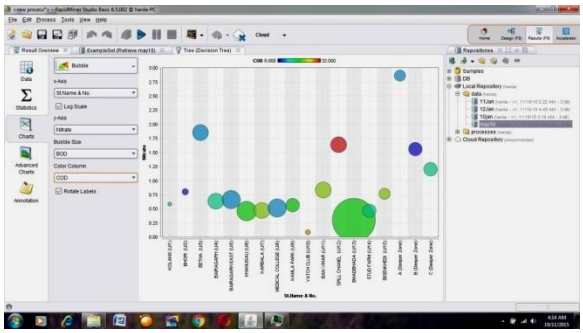

Figure 5: Association between BOD and COD with Nitrate of UL (May-2010).

IX. RESULTS AND DISCUSSION

	Winter Season			**Summer Season**			**Rainy Season**		
Year→	2010	2011	2012	2010	2011	2012	2010	2011	2012
IF → THEN	[1]TH=160(Max)& TA>120→**U13.** [2] BOD + COD (High) & CA<4 &N=0.75→ **U9**. [3] COD (LESS) & N (HIGH)=3.0 & IF TA>99 → **A Zone**	[1] TH=110 Ca(Max) && BOD+COD (high)& N=.025 →**U13.** [2] TH=120 ,Mg & CA<=5→**U1** [3] N=2.75 &BCA<=42→ **A Zone**	[1] BOD+COD(high)' N=0.95 &TH=110 &Ca+Mg(High)& if TA>101 →**U13.** [2] TH <=62 &Ca+Mg less→ **U9**	[1] CaC<18.96→**U7,U8**. [2] N=1.75, OP>1.811 with COD(high)→ **U12.** [3] N< 0.194 with less BOD→ **U10.** [4] TA <=51, TH =80 with less Ca→ **U4**	[1] N=0-0.5, BOD=High→**U13.** [2] TH=120with high Mg +Ca & CA <=5→**U1.** [3] N<= 0.50 with high BOD→ **U-4,6,10,11, 12 13.**	[1]TH=135 Ca +Mg=High & C<28.971 →**U13**. [2] MgC <=4.423 with (Min) TH→**U11.** [3] TA >112 & N=0.3(less)→**C Zone.**	[1]TH=120 with Ca+Mg= (LOw) && TA >95→**U13**.	[1] TH=140(High) &TA>110 &CA=2 →**U14.** [2] High BOD +COD & N=1.2 → **U13.**	NA

Table 3: Classification Model with Class -attribute "Station Number".

Abbreviation: TH=Total Hardness**, TA**=Total alkalinity**, Ca**=Calcium Hardness**, Mg**=Magnesium Hardness**,N**=Nitrate**, C**= chloride, **CA**= Carbonate alkalinity, **CaC=** Calcium content**, MgC=** Magnesium content. **BCA**=Bi-Carbonate alkalinity, **OP**= Org-Phosphorus**,** **BOD**= Biochemical Oxygen Demand, **COD**=Chemical Oxygen Demand (In Unit mg/l).

ISSN: 2278 – 1323

The present study is based on the classification and association of certain parameter like Total Hardness, Chloride BOD; COD, Nitrite, Total alkalinity, carbonate alkalinity; phosphate etc. which are essential to assess the water quality of any water resource .An attempt was made in the study to assess the water quality changes during a decade i.e. 2010-2012. The results obtained are shown in Table 3 and figure 3 to 6.

We have some classification and association rules from our model that In month of Jan during years from 2010 to 2012 ,the water of upper lake at station U13 (Bhadbhada) have found high value of Total hardness between 110 to160 as due to high concentration of calcium and magnesium hardness found at U13 and total alkalinity >=100 to 120.Also nitrate level at U13 found very low (0.25-0.95 mg/l) as due to high concentration of BOD and COD while in all other station Nitrate level vary as in variation in BOD & COD. The water of upper lake at station Bisenkhari (U15) has found minimum value of Total Hardness. In summer season during year from 2010 to2012, Total hardness is nearly about 110 to 135 due to Ca and Mg hardness and chloride<28.971 mg/l. Nitrate level found 2010 from 0.50 mg/l with high BOD and Mg contents<2.807mg/l. In rainy seasons during year 2010 to2012, at station U13 ,Total hardness found high up to 120 to 140 while low value of Ca hardness and Mg hardness as due to alkalinity level high >95mg/l to 110 mg/l. Many more such classification and association facts have found from the model (see table 3). It indicates the river has been gradually deteriorated and needs conservation on plan for both cases of pre monsoon & post monsoon season. The main activities which are mainly responsible for deteriorate the water quality is increase in Mg and Ca hardness and contents .therefore there is a serious need of water management plan for the upper lake of Bhopal city.

X. REFERENCES

[1] Kumar Amit & Chaudhary Om: Article entitled- "Conservation Plan of Upper Lake, Bhopal: Water-Quality Centric Case Study- International Journal of Emerging Technology and Advanced Engineering. (Website: www.ijetae.com) ISSN 2250-2459, Volume 3, Issue 1, January2013.

[2] Phillips Evan: Thesis entitled "Water Quality Modeling and Rainfall Estimation: A Data Driven Approach. The University of Iowa, July 2011.

[3] Fabric voznika and Leonard oviana: Article entitled- "data mining classification", July 2008.

[4]Tanna Paresh & Ghodasara Yogesh :Article entitled -" Improved Frequent Pattern Mining Algorithm with Indexing". IOSR Journal of Computer Engineering, Volume 16, Issue 6, PP 73-78,Dec-2014.

[5] Khan Muzaffar and Ishtiyaq Majeed Ganaie : Article entitled" Assessment of Physico-Chemical Parameters of Upper Lake Bhopal M.P. India" - International Journal of Engineering Research and General Science Volume 2, Issue 4 June -July, 2014 .

[6] Sharma M.K., Choubey V K and Dwivedi: : Article entitled "Phosphorous Modelling of Upper Bhopal Lake Using Best Subset Procedure "-National Institute of Hydrology, Roorkee-246 67, Uttarakhand, India. World Lake Conference: 851-856.

AUTHORS PROFILE:-

[1] **SANJEEV GOUR-** Professor and Head-Department of Computer Science. L.B.S. College of Professional Studies, Harda-461331, M.P. (India) .He has published Six research papers in International journals in the field of Data mining.

[2] **Dr. MAMTA GOUR –** Assistant Professor (G.F.), Department of Chemistry, Govt.P.G. College, Harda-461331, M.P. (India).She obtained her Ph.D. degree in chemistry from Barkatullah University; Bhopal in 2011.She has published four research paper in International journals.

[3] **PURUSHOTTAM SHARMA-** Research Scholars, Barkatullah University, Bhopal, M.P. (India).

ISSN: 2278 – 1323