Analysis of Water Quality using Supervised Machine Learning Classifiers

Danish Paliwal, Aditya Maheshwari, Yogendra Singh Solanki, Prasun Chakrabarti

Abstract*: The paper points out analysis of water with the use of supervised machine learning classifiers. The parameters include turbidity, ph level, water level, temperature and dissolved oxygen. Curve fitting has also been applied in order to analyze the water quality.

***Keywords:-* PH, Dissolved Oxygen, Temperature, Turbidity, Water quality, Analytics.**

I. INTRODUCTION

Water is a limited natural resource. So preserving water is important for protection of our environment [1]. Lake Monitoring System (LMS) provides an understanding of Lake Ecosystem and dynamics. A study was conducted on whether the lake water was safe for living beings or not. The accumulated data proved that water quality was affecting aquatic life, animals drinking that water and also affected the lives of swimmers. This degrading water quality affected health of people and around 1000 children per year died suffering from diarrhea and almost 12 million people died suffering from Typhoid.

Increase in global warming, dumping industrial waste in water bodies, acid rain, sewage drainage in lakes causes lake pollution. Herein lies the significance of testing quality of water. The parameters includephlevel [2],turbidity [3], waterlevel, temperature and dissolved oxygen [4].

II. RESEARCH METHODOLOGY

A study was conducted in 2 lakes of Udaipur: Govardhan Sagar (GS) [5] and Fateh Sagar (FS) [6] which noted data through sensors placed in the lakes. To get the real time pollution data, a system has been developed in which we decided to measure 5 parameters like pH, turbidity, dissolved oxygen, water level and temperature. The dataset used is https://bit.ly/2Rv7tJs.

Weka 3.8.3 tool has been used to analyze the water quality. The curve fitting has been generated in order to predict the quantified value of each parameter towards analysis of quality of water. In addition to this, IBM SPSS has been used for generating the graphs.

Revised Manuscript Received on March 15, 2020.

Danish Paliwal*, Techno India NJR Institute of Technology, Udaipur313003 (Raj.), India. **E-mail: danish24paliwal@gmail.com**
Aditya Maheshwari, Techno India NJR Institute of Technology, Udaipur313003 (Raj.), India
Yogendra Singh Solanki, Techno India NJR Institute of Technology, Udaipur313003 (Raj.), India
Prasun Chakrabarti, Techno India NJR Institute of Technology, Udaipur313003 (Raj.), India

III. EXPERIMENTAL RESULTS AND ANALYSIS

We have used various classifiers to classify data which includes LogisticRegression, NaiveBayes, RandomForest, FilteredClassifier, SMO, VotedPercepton, laxy.IBK, lazy.LWL, lazy.Kstar, DecisionTable, RandomTree and J48. The results of these classifications are as follows:

Table1: Simulation results using Weka 3.8.3.

Classifiers	Total no. of instances	Pres.	Rec.	Acc.	Kp.
LogisticRegression	3060	0.998	0.99	99.83	0.99
lazy.IBK	3060	0.98	0.98	98.7	0.97
lazy.LWL	3060	1	1	99.96	0.99
lazy.Kstar	3060	0.993	0.993	99.3	0.98
RandomForest	3060	1	1	99.96	0.99
RandomTree	3060	0.999	0.999	99.9	0.99
NaiveBayes	3060	0.997	0.997	99.67	0.993
DecisionTable	3060	1	1	99.9	0.99
J48	3060	1	1	99.96	0.99
VotedPercepton	3060	1	1	99.96	0.99
SMO	3060	1	1	99.96	0.99
FilteredClassifier	3060	1	1	99.96	0.99

Pres. = Precision, Rec. = Recall, Acc. = Accurate, Kp. = Kappa Statistics

Value of Kappa Statistics varies from 0-1. If value is greater than 0.8 then it is accepted, if it's ranging between 0.6-0.8 it's substantial, if it is ranging between 0.41- 0.6 it is moderate, 0.21-0.4 is fair and 0.01-0.2 is none to slight [7].

The graphical representation of ph level, turbidity and dissolved oxygen with respect to time is as follows:

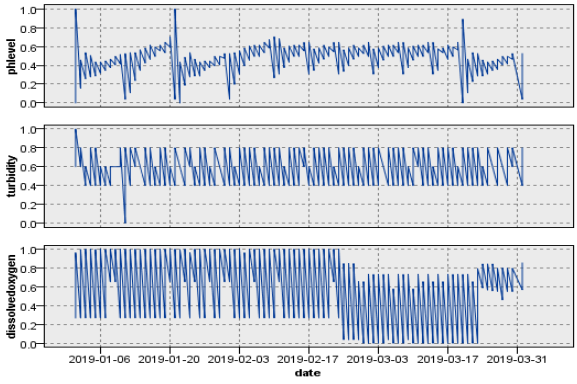

Figure 1: Time Plot for phlevel, turbidity and dissolved oxygen based on data survey of GovardhanSagar(GS)

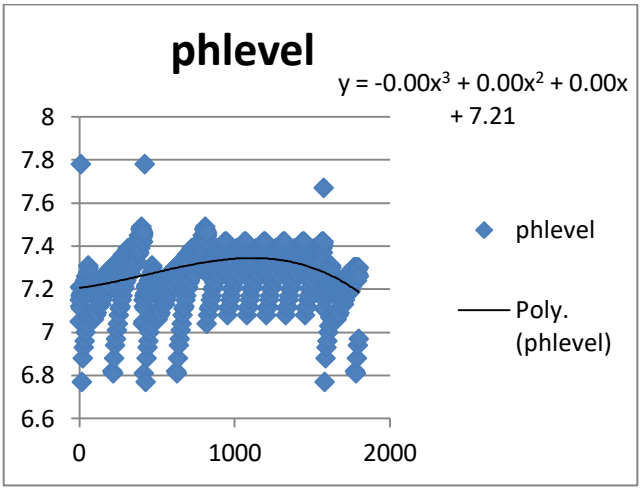

Figure 2: Curve fitting of phlevel based on data survey of Govardhan Sagar(GS)

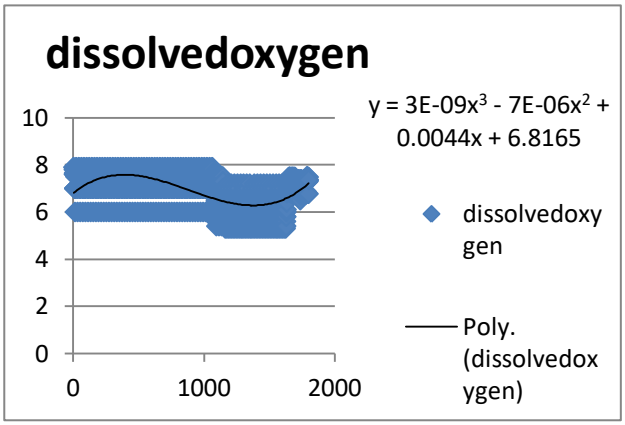

Figure 3: Curve fitting of dissolved oxygen based on data survey of Govardhan Sagar(GS)

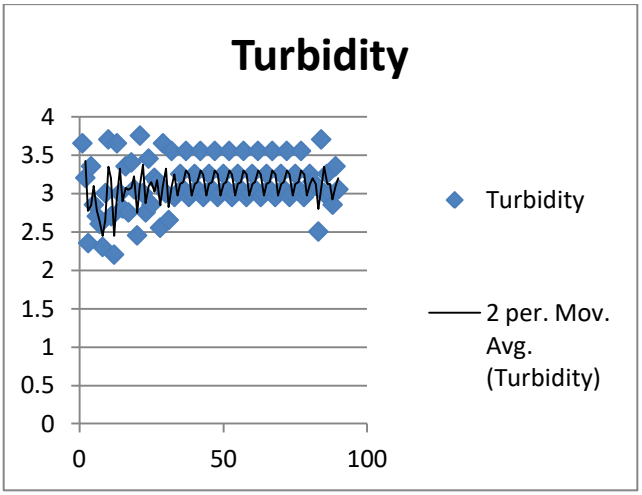

Figure 4: Curve fitting(moving average period= 2) of turbidity based on data survey of Govardhan Sagar(GS)

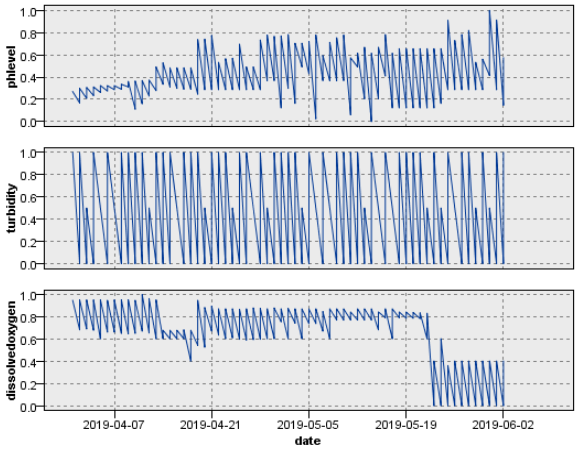

Figure 5: Time Plot for phlevel, turbidity and dissolved oxygen based on data survey of FatehSagar(FS).

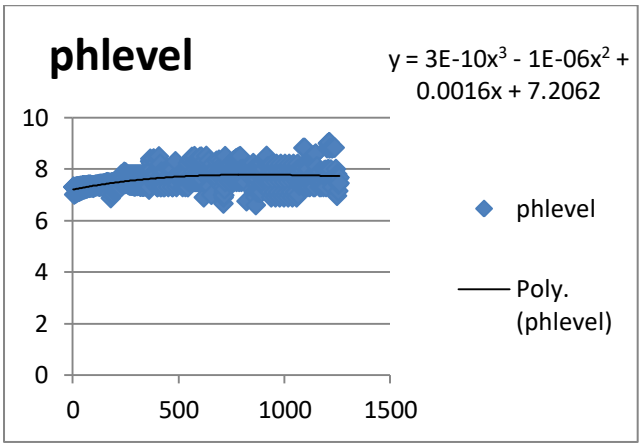

Figure 6:Curve fitting of ph level based on data survey of FatehSagar(FS)

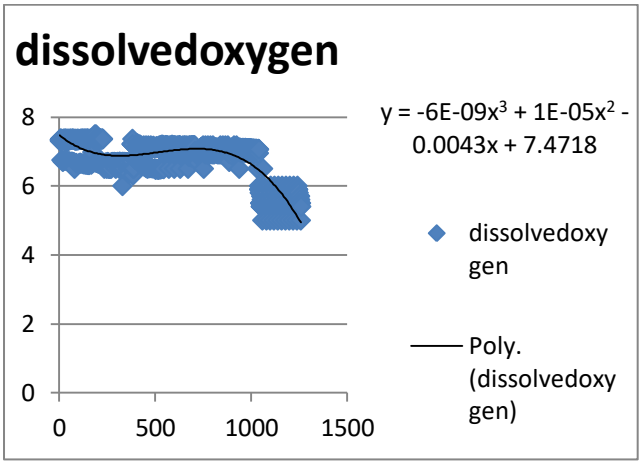

Figure 7: Curve fitting of dissolved oxygen based on data survey of Fateh Sagar(FS)

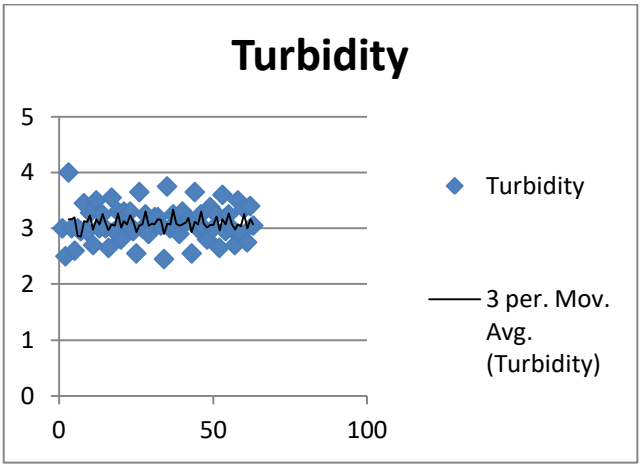

Figure 8: Curve fitting of turbidity (moving average period=3) based on data survey of FatehSagar(FS)

The observations are as follows:
In case of quality analysis of GS-

- The quantity of dissolved oxygen decreased in end of February month and increased in end of March month.
- From February end to mid of March month, the measure of phlevel is almost uniform.
- Turbidity remained almost uniform from mid January to March month.

In case of quality analysis of FS-

- Phlevel increased gradually with respect to time.
- Turbidity maintains a uniform pattern.
- From mid may to June first week, the measure of dissolved oxygen gradually decreased.

IV. CONCLUSION

The best accuracy estimate towards quantification of water quality is achieved by lazy. LWL, RandomForest, J48, Voted Percepton, Filtered Classifier and SMO classification. The experimental results entail that as phlevel decreases, dissolved oxygen also decreases hence increase in turbidity. Dissolved oxygenis alsoaffected by temperature.As temperature increases, amount ofdissolved oxygen decreases in water.

REFERENCES

1. M N Barabde, S R Danve, "Continuous Water Quality Monitoring System for Water Resources at Remote Places", International Journal of Engineering Research and General Science, Vol.2, no.3, Part 2 , May-June 2015, ISSN 2091-2730.
2. Ph: Definition and measurements https://www.usgs.gov/special-topic/water-science-school/science/ph-and-water?qt-science_center_objects=0#qt-science_center_objects
3. Himanshu Patel, R.T. Vashi, in " Characterization and Treatment of Textile Wastewater, 2015
4. Dissolved oxygen and aquatic life https://www.fondriest.com/environmental-measurements/parameters/water-quality/dissolved-oxygen/
5. https://myudaipurcity.com/goverdhan-sagar-lake/
6. "Lakes"Archived from the original on 27 March 2013
7. Mary L. McHugh "Interrater reliability: the kappa statistic" Oct 2012
8. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, "Electron spectroscopy studies on magneto-optical media and plastic substrate interfaces (Translation Journals style)," *IEEE Transl. J. Magn.Jpn.*, vol. 2, Aug. 1987, pp. 740–741 [*Dig. 9th Annu. Conf. Magnetics* Japan, 1982, p. 301].

AUTHORS PROFILE

Danish Paliwal, is a final year student pursuing B. Tech in Information Technology, Techno India NJR Institute of Technology, Udaipur. She is also pursuing specialization in Business Analytics certified by IBM.

Aditya Maheshwari, is working in the capacity of Project Lead- New Initiative & Research at Techno India NJR Institute of Technology Udaipur. He is an Open Source Advocate, leading technical speaker and Expert of User Research (UI/UX). Leading students by creating and developing projects applying cloud technologies at an engineering institute. Achieved specialization and certification in IBM Watson and Block Chain and applied in developing projects.

Yogendra Singh Solanki, is working as Asst. Professor, Department of Electronics and Communications, Techno India NJR Institute of Technology, Udaipur. He is practicing the IoT Devices, AI/ML and Data Science in Medical Analytics.

Prof Dr. Prasun Chakrabarti, has received his PhD (Engg) from Jadavpur University in 2009. He is working as Executive Dean (Research and International Linkage) and Institute Distinguished Senior Chair Professor, Techno India NJR Institute of Technology. He has 182 publications, 9 books and 34 filed Indian patents in his credit. He has supervised ten PhD candidates successfully. On various research assignments, he has visited Waseda University Japan (2012 availing prestigious INSA-CICS travel grant), University of Mauritius (2015), Nanyang Technological University Singapore (2015,2016,2019), Lincoln University College Malaysia (2018), National University of Singapore (2019), Asian Institute of Technology Bangkok Thailand (2019) and ISI Delhi (2019). He is a Fellow of IET (UK), IETE, ISRD (UK), IAER (London), AE (I), CET (I) and senior member of the IEEE (USA).