Library Progress International *Print version ISSN 0970 1052*
Vol.44 No. 3, Jul-Dec 2024: P.18159-18167 *Online version ISSN 2320 317X*

Original Article *Available online at www.bpasjournals.com*

Water Quality Diagnosis using IOT and KNN Classifier Integrated with Think Speak Cloud

Dr.S.Selvi[1], Dr G Revathy[2], Dr.N.S.Kavitha[3] & Dr.S.Russia[4]

1- Associate Professor,Department of Artificial Intelligence and Data Science,
Builders Engineering College,Nathakadaiyur, Tiruppur Dt. Email:dselvi.tpr@gmail.com
2- Assistant Professor, Department of CSE, Srinivasa Ramanujan Centre, SASTRA Deemed University, Kumbakonam. revathyjayabaskar@gmail.com. 0000-0002-0691-1687
3-Assistant Professor(Sr.Gr.), Department of Information Technology, Dr. N. G. P. Institute of Technology, Coimbatore, :nskavi17@gmail.com
4-Professor, Department of CSE, Velalar College of Engineering and Technology,
Erode, russiavcet@gmail.com, 0000-0002-0271-4109.

How to cite this article: S.Selvi, G Revathy, N.S.Kavitha, S.Russia (2024) Water Quality Diagnosis using IOT and KNN Classifier Integrated with Think Speak Cloud. *Library Progress International,* 44(3), 18159-18167.

ABSTRACT

The textile and dyeing industries pave a foremost providing responsibility in our Indian Economy and creates a non discrete monetary progress in nation. However, Water pollution causes a major downfall simultaneously. The discharge of chemical wastages from the industries into the river has a direct impact on the biological organisms. To face this challenge and to prevent the water resources from getting defected, this project is being proposed with a novel solution. This paper provides a multi technological solution, by integrating the Machine Learning techniques with a Cloud-Based Wireless Sensor Network (WSN). This would lead to the dismissal of Waste Water Chemical effluent discharge into the river in an efficient manner. The waste water is monitored using Internet of Things(IoT) sensors for various attributes that include pH, conductivity, dissolved Oxygen-Biological Oxygen Demand(BOD),Chemical Oxygen Demand(COD),Total Dissolved Solid (TDS) Content,etc . In order to monitor pH, conductivity, and dissolved oxygen characteristics from wastewater released into water sources, this suggested paper uses an integrated cloud-based environment. The system uses AT commands in conjunction with the HTTP GET technique to gather and upload sensor data to the ThingSpeak cloud via GPRS internet connectivity in order to deliver an Internet of Things (IOT) based real-time online monitoring. Additionally, the system is supported with an SMS gateway service provided by the Telerivet mobile messaging platform to send message alerts to the responsible organisation.The experimentation has been simulated using Jupyter Notebook for different values of K and the data acquired through IOT sensors are being visualized in live using ThinkSpeak Cloud. The comparison of various classification models were made and its found that KNN provides a betters classification when compared to the other relevant models.
Keywords:*Water Quality, KNN, ThinkSpeak, IOT*

1. INTRODUCTION:

An improper management of waste water discharge by the industries into the river is a very serious threat universally to the health of biological organisms and in extreme cases, it may also lead to the higher rate of mortality. The presence of colored chemical materials and dye stuff ventures into the waste water bodies are the major components identified as the cause of deterioration in our everyday life. These venomous eminence and their deceivability issues of water bodies has been found to be the unsolvable issues in the most recent years. Various conventional techniques have been made in the existing projects and they are found to provide a minimal effect in the dismissal of complete poisonous substances in to the water bodies. The BOD signifies the consumption of dissolved oxygen when gets decomposed in the water. The COD indicates the consumption of oxygen when the water sample is oxidised. The normal pure water has a conductivity of 0.5 to 3 µs/cm, whereas the inorganic dissolved solids like chloride, nitrate, etc. in the waste water cause a intolerable higher conductivity.

The accumulation of increased alkaline levels and toxic chemicals significantly affect the pH level. The occurrence of inorganic salts, organic matter and other dissolved materials promotes a higher TDS in the waste water.

The various IOT sensors are utilized to monitor and measure the above said components and prohibits from the discharge of waste sewage water from the industries into the water bodies. The data gathered from the sensors are agreed collectively in a non discrete time series to the ThinkSpeak Cloud through a Wireless Sensor Network. A Wireless Module naturally a Zigbee network or a Wimax Network is indulged in transferring a fastest communication of the sensor data to the Cloud. The ThinkSpeak Cloud compiles the data and uphold KNN Supervised Machine Learning Algorithm to predict the percentage of poisonous effluents in the sewage water and concludes whether the water is being permissible to be discharged from the industry or not. The KNN algorithm thoroughly inhales and analyzes the compiled data. It results out an accurate prediction of the amount of poisonous organic and inorganic substances in the water and it how far affects the health of biological organisms. Based on the outcome of KNN, a SMS alert notification report would be sent to the concerned Government Authorities and to the industries. Hence an assuring the well being of water is a great challenge which could be solved by this proposed work.

2. RELATED WORKS:

IoT serves as a communication tool and is used by the internet to sense and control devices, including other smart appliances. [4] [5] [8]. IoT encourages the development of intelligence in linked objects. The intervention of Humans is reduced considerably and provides a faster approach using IoT. ThinkSpeak is an open cloud platform that makes it possible to collect, analyse, and act on real-time data via an open API[20]. ThinkSpeak[15] stores the sensors data in channels and provides either a private or a public channel. ThinkSpeak uses read key and write key for accessing the sensors data. The write key is meant for uploading the sensor data into Clod and read key is used in private channel for data access. Carrio [21] and Xively [22]could also be alternate to ThinkSpeak, but The usage of an open source API by ThinkSpeak, ease of data exports, utilisation of live graphs for data visualisation, and accessibility of public channels for data storage all contribute to this. In contrast to ThingSpeak, Xively only offers a free trial edition and has few notification features. Carriot [22] lacks a ready-made dashboard, making it more expensive and less user-friendly. ThinkSPeak demonstrates why it is a superior IoT Cloud Platform.

Yono Zakaria et.al[1] built a prototype to monitor the various characteristics of waste water like pH, Conductivity and dissolved Oxygen content. The data are being collected by the concerned sensors, observed via a Telerivet Message Communication plation and ThinkSpeak Cloud. Various improvement methodologies are summarized by Changxin et. al[2] for improvising the quality of river water.Lakshmi Devi et. al[3] has provided an unique solution to discover the contamination level of water and unveiled an notification mechanism for the controlling the effluent discharges into the river water. A Cloud based solution has been provided by Anzar Ahmed et. al[6] for continuous monitoring and identification of the water contamination level of reserved water. An IOT[10] relied Smart Water Quality Diagnosis system with lower cost [12] [13][14]and a less complex framework using a controller and Wi-Fi Model was designed by Geetha et. al[7]and Prasad et.al[8].

3. PROPOSED METHODOLOGY:

3.1 The System Architecture:

The structural design of the proposed KNN integrated with ThinkSpeak Cloud for prediction and monitoring industrial sewage water discharge into water bodies is illustrated in Figure 1. The proposed system consists of the following major components: WSN sensor nodes, Raspberry Pi, Wireless Module, Cloud platform for IoT, KNN based data prediction analysis and the SMS-notification module.

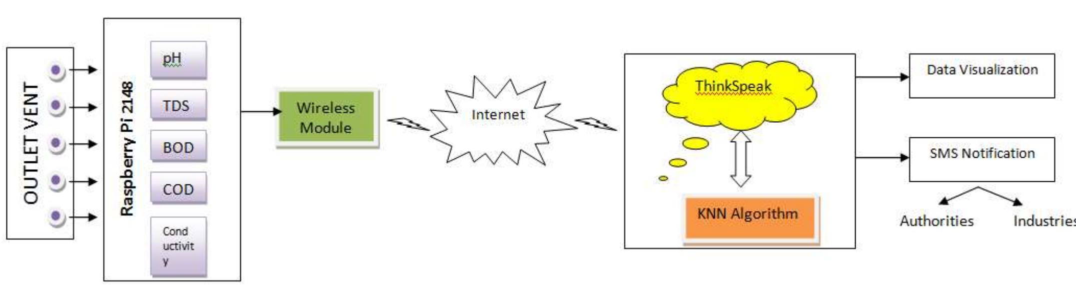

Figure 1:KNN integrated with Cloud and WSN for prediction and monitoring industrial sewage water discharge into water bodies

3.2 Data Visualization and SMS Alert Notification:

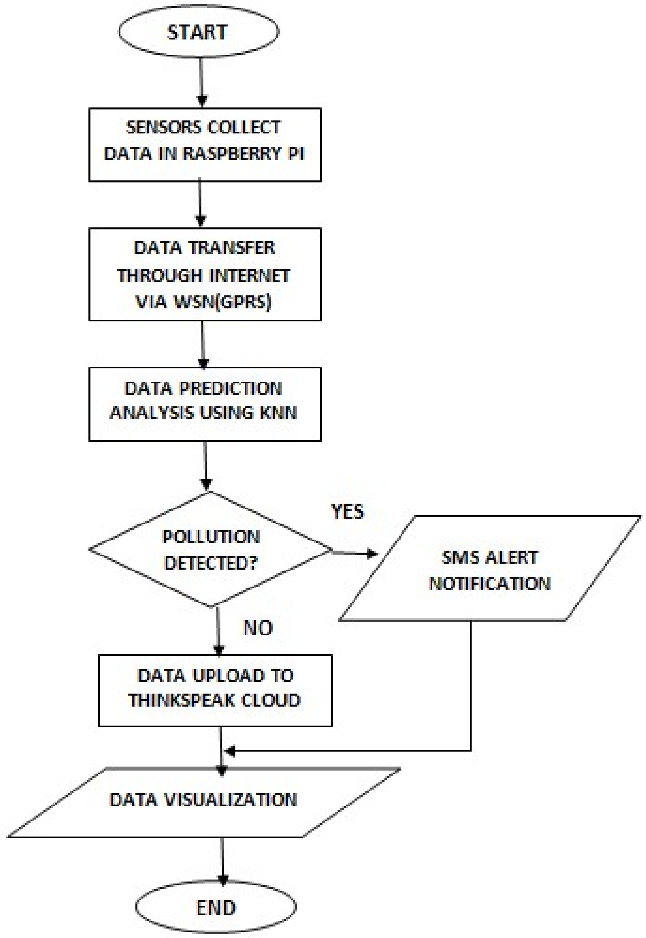

Figure 2: Flowchart of Data Visualization and SMS Alert Notification Process

Figure 2 portrays the flowchart summarizing about the data collection from the waste water till the SMS notification and Data Visualization. The Raspberry Pi incorporate the following sensors: pH, TDS, BOD, COD and conductivity sensors. These sensors collect the data from the sewage water, compile the data and sent through GPRS via internet to the Data Prediction and Analysis. KNN algorithm is employed to segregate the data and predict whether the organic and inorganic substances present in the sewage waste is below the threshold value. If it is below, then the water is allowable to be dissolved in tot the water bodies. Else, pollution is detected. Telerivet mobile messaging platform is used to notify an SMS alert messaging service to the respective Government authorities and to the industries. The compiled data is then lively visualized in the form of Plotted/Line graph to the users.

The proposed paper comprises of SMS Notification process, which acts as a means for communicating the detection of Pollution as an alert system. The KNN algorithm accompanied by the WSN sensor nodes using Raspberry Pi kit, if predicting the presence of Pollution in the sewage outlet sources, initiates an alert system to the concerned Government Authorities and the respective industries. Telerivet [17] mobile messaging platform is used for implementing this service. Installing the Android app for Telerivet on the Android devices is a pre-requisite component. Tanzania used a Telerivet mobile messaging platform for the SMS alert notification and also the platform used SWAHILI as the back end language. There are also other similar SMS platforms like TERA [18] and Twilio [19].

There are two destinations where SMS trigger notifications to be transmitted in this project. Once the Water Pollution is detected, "POLLUTION DETECTION ALERT" is being reported to the Government Authorities and POLLUTION DETECTION ALERT.INITATE RECYCLE” to the concerned industries. This is being achieved through an SMS gateway service. Figure 3 depicts the SMS triggering methodology, in case of any pollution detection by the sensor nodes.

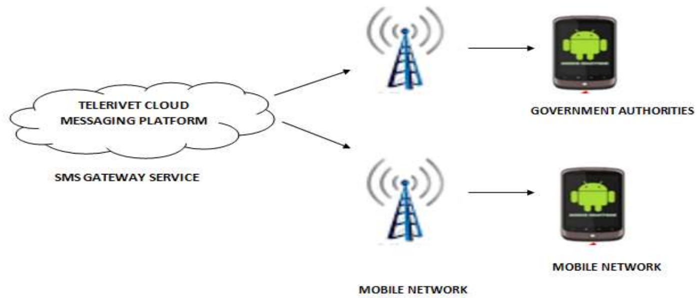

Figure 3: Proposed SMS gateway service

Following are the keywords used as the Pollution indicator for SMS triggering alert:

1) Deaths — A keyword indicating of the passing of an animal, plant, bird, or aquatic organism (such as a frog) after ingesting toxic water.
2) Illness — A key sign that people get ill after drinking contaminated water.
3) Smell — Bad odour inducing from the wastewater discharge.
4) Color — Colored wastewater discharged

An SMS alert is generated whenever one of the aforementioned pollution keyword indicators is sent to the SMS gateway service. This authenticates the possibility of pollution existence by the release of industrial wastewaters.

3.3 Data Conversion from IoT to Cloud Platform

ThinkSpeak Cloud is being deployed for live relay and it could be possible by uploading the data using the HTTP GET method with the avail of GPRS internet connectivity. Also a set of AT commands has been used for further implementation. Initially, api.thinkspeak.com is established using a TCP connection by the AT command AT+CIPSTART. Sensor data is converted into a String format using the command AT+CIPSEND and the HTTP GET method with the API Key. Finally, AT+CIPCLOSE command is used to close the TCP connection once the data is sent.

3.4 ThinkSpeak Cloud

The following procedures have been followed by ThingSpeak which is facilitated with an Open Source IoT analytics platform services:

- Cumulative Aggregation of data
- Visualization of Live Graph
- Analyze live data streams in the cloud.

The Raspberry Pi senses the data, traverses through the communication channel and are either stored in a private or a public channel. Then it is further generates an instant, dynamic visualization of live data, and send alerts. By default, private channel is opted for storage. ThinkSpeak also allows interacting with social media, web services and other devices. In our project, the Cloud interacts with Android devices as SMS alert notifications. One of the main advantages of using ThinkSpeak Cloud is any number of channels could be created for a single application. Using a REST API or a MQTT API, data is transmitted from devices or local gateways to ThingSpeak..

The data stored in a central location can be easily accessed either offline or online. The data are collectively from IOT devices and equipment. Further KNN algorithm is being used to develop the prediction analysis of the existence of Water Pollution. One of the reasons to chose ThinkSpeak[16] is due to its fastest transmission of data as once every second and faster response times. The live charts could also be customized as per our requirement. The live charts are visualized for each of the expected parameters.

3.5 KNN Prediction Analysis of Water Pollution

KNN Algorithm is responsible for the prediction of the received data from the sensor nodes. The Machine is trained with test samples of various parameters (pH, BOD, COD, TDS, and Conductivity) separately for future prediction. Based on the compiling of all the prediction outcomes, the Machine will provide whether the sewage water is being polluted and how far it would affect the life of Biological Organisms.

The Flowchart for the KNN Prediction Process is depicted in the **Figure 4:**

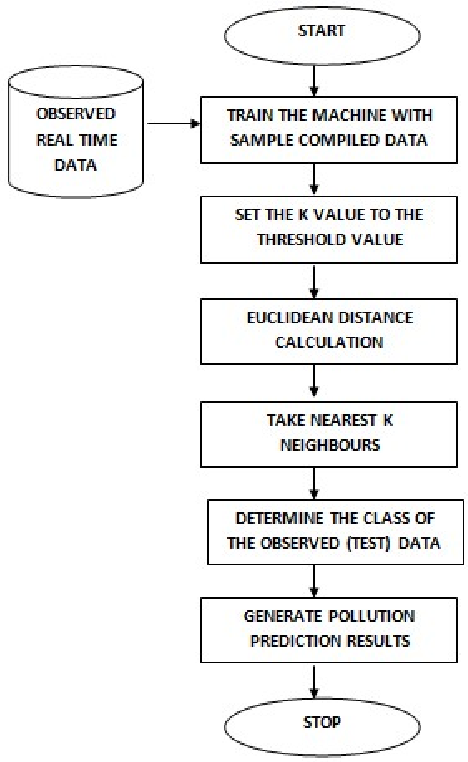

Figure 4: KNN Prediction Process

KNN algorithm receives the sensor (normalized) data as an input. The algorithm considers 70% of the input data for training, the machine gets trained for pollution prediction. The remaining 30% of the data is considered to be the testing data. The KNN algorithm set the value of K with the threhold value for each of the categorical parameters (pH,BOD,COD,TDS and Conductivity). The machine sets that if the data falls down the threshold value(K), then there is no pollution, else if the data is above threshold value(K), then there is an occurence of pollution. Euclidan distance is calculated among the data points and take the K- nearest neighbours. Among these K neighbours, the data points are counted in each category. The majority of points in a category gives the result.

KNN algorithm is run over all the sensors data seperately and all the outputs are combined finally. Analysing the combined prediction and based on the majority, the Pollution result is drawn out by the algorithm.

3.6 RESULTS AND DISCUSSIONS

The data read by the IOT sensors from the wastage water with essential parameters like pH,BOD,COD, Conductivity,etc is experimented. The proposed paper is being implemented using Anaconda Navigator and Jupyter Notebook for prediction.

3.6.1 Violin Plot Representation

The dataset is of non linear seperable type and it is represented in the form of Violin Plot (Figure 5). The geometric distribution of the data could be visualized effectively using Violin plot.

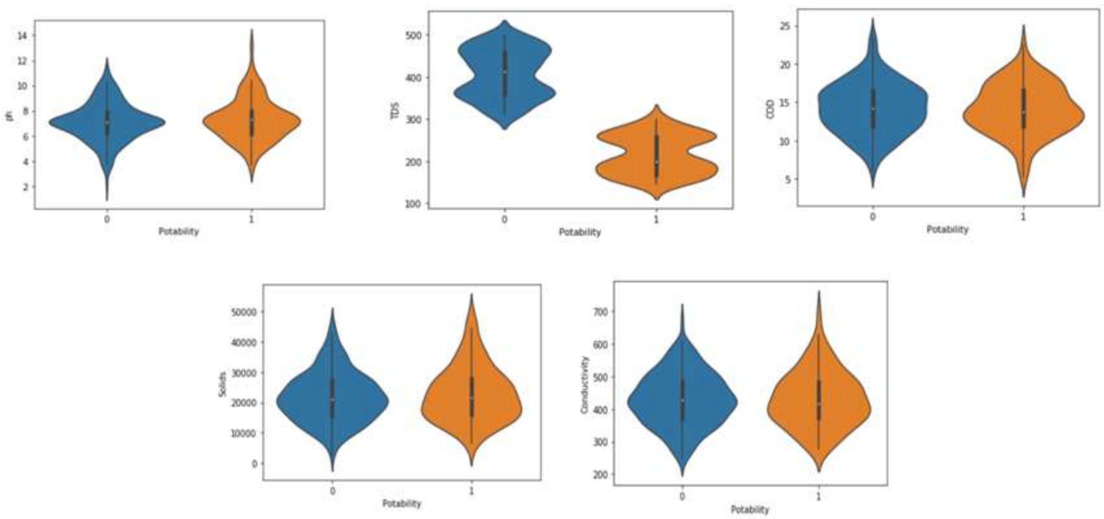

Figure 5: The dataset representation as a Violin Plot

The data of water potability suitable for human use or not has been splitted and symbolized in Figure 5.Also the dependent variable's distribution agaist all the other independent attributes has been provided in the Figure 5.

3.6.2 Potability Analysis with Independent Attributes

From the obsevations, Table 1 provides the attribute wise analysis of the data set.

Table 1: Potability Analysis

Independent Attributes	Potability (0-Not Suitable for Human Use)	Potability (1 - Suitable for Human Use)
pH	49.4%	51.6%
COD	79.42%	20.58%
TDS	78.30%	21.70%
BOD	97.1%	2.9%
Conductivity	95.3%	4.7%

The potability measure against the independent attributes has been tabulated in Table 1. It is found that COD and Conductivity content is far behind the acceptable range to support for the human usage.

3.6.3 KNN with Normalization Measures

One the data set has been imported into the machine, the varying values measured on a different scale has been normalized into an uniform scale. Without scaling the data, it would be complicate for experimentation and for further analysis with other parameters. The proposed work applied KNN with three normalization techniques:Z-Score, Min-Max and Feature Scaling.K values are taken to be 25,55,75,95. Accuracy of Prediction is being recorded accordingly.

Table 2 and Figure 6 provides the performance of KNN algorithm with the different Normalization measures.

Table 2: Prediction Accuracy of KNN with Z-Score, Min-Max and Feature Scaling

K	Prediction Accuracy		
	Z-Score	Min-Max	Feature Scaling
25	83.96%	95.44%	88.54%
55	85.25%	87.61%	86.65%
75	99.37%	81.85%	92.31%
95	94.58%	79.96%	94.56%

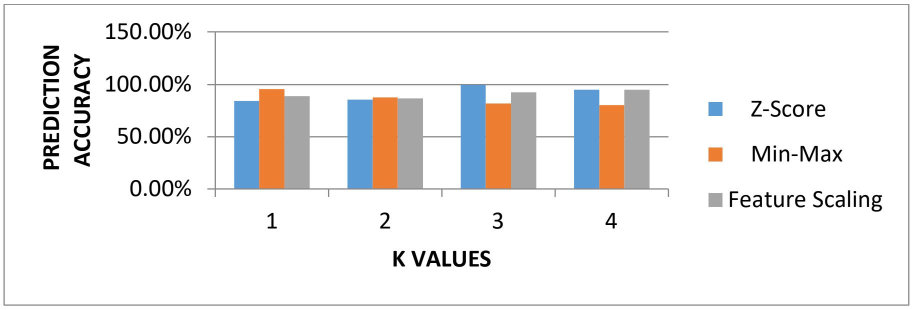

Figure 6: KNN Prediction Accuracy with Z-Score, Min-Max and Feature Scaling

3.6.4 Comparision of Models

A Comparision of various classification models such as Gaussian Naive Bayes(NB),Logistic Regression and Support Vector Machines(SVM) has been made with KNN classifier. The results has been displayed in the Table 3 and diagrammatically represented in Figure 7 & 8. The metrics such as F1 Score, Kappa and Mean Square Error(MSE) have been taken into account for both training and testing data set. 70% of the dataset was considered for training the machine and 30% was taken as the testing data set. F1 Score is the one of the prominent evaluation metric since it balances eventually both the precision and recall.Kappa is chosen for comparision since it provides a conclusion about how far better the performance of a classifier based on the frequency of classes. Also Kappa replicates the reliability of a classification algorithm. Kappa lies between the range of 0 and 1. 0 indicates a dis-agreement of a model and 1 represents a total perfect agreement of a model. Error is the one representing Mean Squared Error which penalizes even the negligible errors and helps in optimizing the model.

Table 3: Comparision of ML Classification Models

Model	Training_ Accuracy(%)	Testing _ Accuracy(%)	F1 Score(%)	Kappa	MSE
Gaussian NB	54.85	49.95	89.53	0.8697	4
Logistic Regression	63.01	64.26	89.61	0.7945	4
SVM	78.82	79.56	91.12	0.9256	2
KNN	89.62	91.54	98.24	0.9688	2

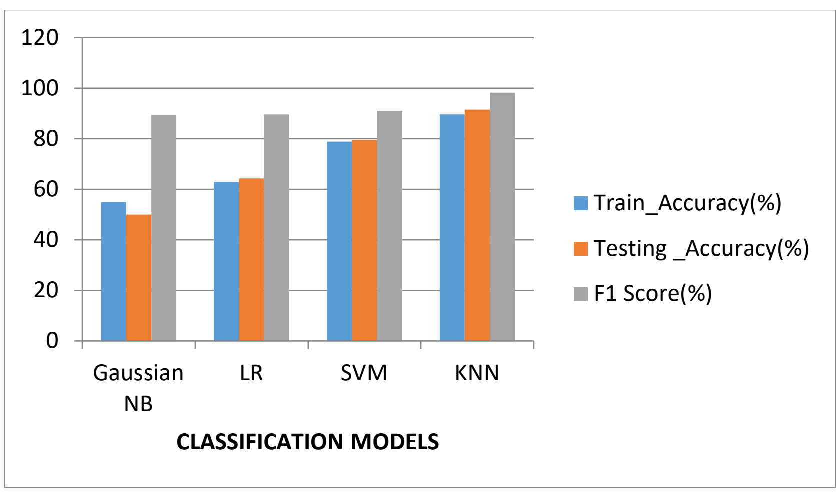

Figure 7: Comparision of ML Classification Models:Training_Accuracy,Testing Accuracy,F1 Score

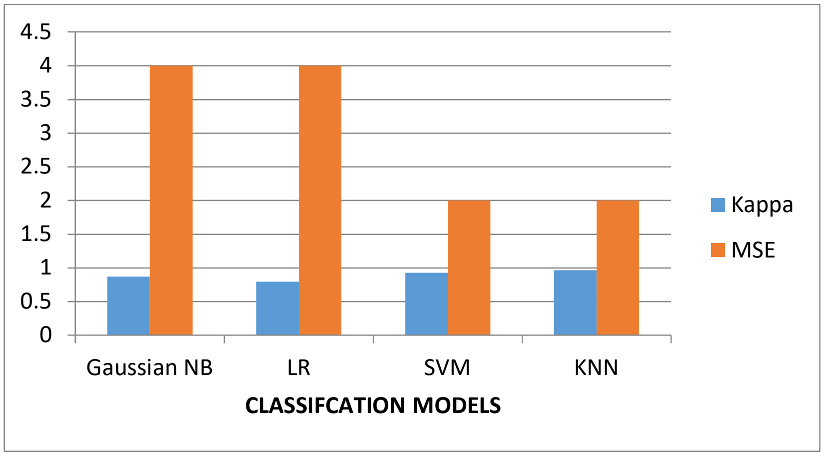

Figure 8: Comparision of Models:Kappa & MSE

Figure 9 explores a Visualization Live Graph of Varying parameters with time for water in ThinkSpeak Cloud. The data sensed through IOT could be made public to become aware about the sewage particles content in the outlet.

Figure 9: Visualization Display of Varying Parameters with Time in ThinkSpeak Cloud

3.7 CONCLUSION AND FUTURE WORK

Classifying the potablility of waste water during the dismissal of Chemical Effluent into the river water is one of the major concern in the society and it directly affects the life of living organisms.The proposed paper provides a better solution by the introducing the KNN algorithm for analysing and classifying the potability of discharge. The attribute values from the sensors are being continuously monitored using ThinkSpeak Cloud platform. The platform gives a clear and detailed segregation of the waste water chemical characteristics. Simultaneously an alert notification would be delivered tot the respected Government authorities in case if the discharge is highly polluted.The experiment is simulated using Jupiter Notebook with all necessary libraries. From the results, it is found that the classification using KNN gives 98.24% accuracy when compared to other classification models. The proposed model could be further extended by applying hybrid techniques with KNN and implementing in real time.

REFERENCES

[1]Yona Zakaria, Kisangiri Michael," An Integrated Cloud-Based Wireless Sensor Network for Monitoring Industrial Wastewater Discharged into Water Sources", Wireless Sensor Network,Scientific Research Publishing,ISSN:1945-3078,pp.290-301,2017.

[2] Changxin Cai," Research Progress in Water Quality Improvement", IOP Conference Series: Materials Science and Engineering, 2019. doi:10.1088/1757-899X/484/1/012050.

[3] Lakshmi Devi Ramakrishnan, Saravanan Govindarajan ,"River Pollution Control System through Efficient Monitoring of Industrial Effluent Discharge", Journal of Green Engineering, Vol. 10, No. 10,pp.9943-9952,2020.

[4] Varsha Lakshmikantha, Anjitha Hiriyannagowda, Akshay Manjunath, Aruna Patted, Jagadeesh Basavaiah , Audre Arlene Anthony ," IoT Based Smart Water Quality Monitoring System", Global Transitions Proceedings,Vol. 2,pp.181-186,2021.

[5] Ajith Jerom B , Manimegalai R , Ilayaraja V "An IoT Based Smart Water Quality Monitoring System using Cloud", International Conference on Emerging Trends in Information Technology and Engineering,2020.

[6] Anzar Ahmad Shashi Shekhar, Abhijeet Roy," Cloud Based Water Reservoir Quality Monitoring System",International Journal of Recent Technology and Engineering, ISSN: 2277-3878, Vol.8 ,No.6, 2020.

[7]Geetha S, Gouthami.S, "Internet of Things Enabled Real Time Water Quality Monitoring System", DOI 10.1186/s40713-017-0005-y, 2017.

[8] Prasad. M. Pujar, Harish. H. Kenchannavar, Umakant. P. Kulkarni ," Water Quality Assessment and Monitoring for River Malaprabha using the Internet of Things(IoT) System", International Journal of Recent Technology and Engineering, ISSN: 2277-3878, Vol. 8, No. 2, pp.3839-3844, 2019.

[9] M.J.V. Kumar , K. Samalla , Design and Development of Water Quality Monitoring System in IoT, Int. J. Recent Technology of Engineering, Vol.7, No.5S3, ISSN:2277–3878,2019 .

[10] A.N. Prasad, K.A. Mamun, F.R. Islam, H. Haqva, "Smart water quality monitoring system", in: Proceedings of the 2nd Asia-Pacific World Congress on Computer Science and Engineering (APWC on CSE), 2015, pp. 1–6, doi: 10.1109/AP- WCCSE.2015.7476234 .

[11] A.T. Demetillo, M.V. Japitana, E.B. Taboada, "A System For Monitoring Water Quality in A Large Aquatic Area Using Wireless Sensor Network Technology", Sustain. Environ. Res. 29 ,No.12, doi: 10.1186/s42834-019-0009-4 .

[12] B. Anuradha , R. Chaitra , D. Pooja , "IoT Based Low Cost System for Monitoring of Water Quality in Real Time", Int. Res. J. Eng. Technol. (IRJET) Volume: 05 (Issue: 05) (2018) May-2018 .

[13] T. Sugapriyaa , S. Rakshaya , K. Ramyadevi , M. Ramya , P.G. Rashmi , "Smart water Quality Monitoring System for Real Time Applications", Int. J. Pure Appl. Math. 118 (2018) 1363–1369 .

[14] S. Pasika, S.T. Gandla, "Smart Water Quality Monitoring System with Cost-Effective using IoT", Heliyon 6 (7) (2020), doi: 10.1016/j.heliyon.2020.e04096.

[15] Simitha.K, Subodh Raj M.S.,"IOT and WSN Based Water Quality Management System", 3rd International Conference on Electronics, Communication and Aerospace Technology, doi:10.1109/ICECA.2019.8821859, 2019.

[16] Abbas Hussien Miry, Gregor Alexander Aramice," Water Monitoring and Analytic Based ThingSpeak", International Journal of Electrical and Computer Engineering,Vol.10, No. 4, pp.3588-3595.doi: 10.11591/ijece.v10i4.

[17] Telerivet SMS Platform. https://telerivet.com/about

[18] Trilogy Emergency Relief Application (TERA) IFRC. http://www.ifrc.org/en/what-we-do/beneficiary-communications/tera/

[19] Twilio Global Short Message Service API. https://www.twilio.com/sms

[20] ThingSpeak Documentation. https://www.mathworks.com/help/thingspeak/

[21] Carriots—Internet of Things Platform. https://www.carriots.com/

[22] Xively-IoT Platform for Connected Devices. https://www.xively.com/