**Mini Project Report**

On

**“Karad Tehsil Water Quality Prediction using Machine Learning.”**

Submitted To

**SHREE SANTKRUPA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GHOGAON.**

**Affiliated to DBATU University**

In Partial Fulfillment of

The Requirement for the Degree of

**BACHELOR OF ENGINEERING (COMPUTER SCIENCE &ENGINEERING)**

Submitted By

1. **Mrs. Akanksha Satish Chavan**
2. **Mrs. Sayali Sampat Pawar**
3. **Mrs. Pooja Gorakh Patil**
4. **Mrs. Shradha Vilas Surywanshi**

Under The Guidance of

**Asst. Prof. AshokSadavare**



**Department of Computer Science and Engineering**

**S.I.E.T. GHOGAON, Tal-Karad.**

**YEAR 2023-24**



**SHREE SANTKRUPA INSTITUTE OF ENGINEERING AND TECHNOLOGY, GHOGAON.**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING**

**CERTIFICATE**

This is to certify that following group members

1. **MRS. AKANKSHA SATISH CHAVAN (Roll NO-52)**
2. **MRS. SAYALI SAMPAT PAWAR (Roll NO-51)**
3. **MRS.POOJA GORAKH PATIL (Roll NO-47)**
4. **MRS. SHRADHA VILAS SURYWANSHI (Roll NO-44)**

of Third Year (Computer science & Engineering) has successfully completed their work entitled **“Karad Tehsil Water Quality Prediction using Machine Learning”** towards the partial fulfillment of award of Bachelor of Engineering (Computer science) degree as laid by the DBATU university during the academic year **2023-2024.**

|  |  |  |
| --- | --- | --- |
| **Asst. Prof. . Ashok Sadavare** | **Asst. Prof. Patange.S.S** | **Dr. S. B. Kulkarni** |
| Guide | H.O.D. | Principal, SIET Ghogaon |

**ACKNOWLEDGEMENT**

I would like to express my profound gratitude to **Asst. Prof. Patange.S.S, HOD of Computer science department**, and **DR. S. B. Kulkarni, Principal of SSIET Ghogaon** for their contributions to the completion of Mini project **“Karad Tehsil Water Quality Prediction using Machine Learning.”**

I would like to express my special thanks to our mentor **Asst. Prof. Ashok Sadavare.**

for his time and efforts he provided throughout the Semester. Your useful advice and suggestions were really helpful to us during the project’s completion. In this aspect, I am eternally grateful to you.

I would like to acknowledge that this project was completed entirely by me and not by someone else.

1. **MRS. AKANKSHA SATISH CHAVAN (Roll NO-52)**
2. **MRS. SAYALI SAMPAT PAWAR (Roll NO-51)**
3. **MRS.POOJA GORAKH PATIL (Roll NO-47)**
4. **MRS. SHRADHA VILAS SURYWANSHI (Roll NO-44)**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| **S.N.** | **Title** | **Page No.** |
| I | List of Images | 5 |
| II | Abstract | 6 |
| 1. | Introduction | 7 |
| 2. | Objectives | 8 |
| 3. | Literature review | 9 |
| 4. | Project details | 11 |
| 5. | Advantages | 24 |
| 6. | Conclusions | 25 |
| 7. | References | 26 |

**I. LIST OF IMAGES & FIGURES**

1.1 Implementation process

2.1 Data Preprocessing

3.1 Visualization

4.1 Percentage of Potablity and Not-Potablity

5.1Confusion Matrix for the Support Vector Machine

6.1Accuracy of the support vector machine and Logistic regression

**II. ABSTRACT**

Water is a crucial and indispensable resource for sustaining human life, and maintaining its quality is of utmost importance for the well-being of individuals. When drinking water becomes contaminated, it poses severe health risks, including diseases like diarrhea, cholera, and various other waterborne ailments. As a result, ensuring safe and clean water becomes crucial to promote public health. Recent findings indicate that a significant number of approximately 3,575,000 people lose their lives each year due to water-related illnesses. Notably, machine learning algorithms have emerged as powerful tools for effectively predicting water quality, enabling timely and precise monitoring of water resources. This research focuses on multiple algorithms to forecast water potability based on the physicochemical properties of water samples obtained from the Drinking Water dataset available on (https://jaljeevanmission.gov.in). This dataset comprises 15 distinct parameters, namely Temperature, Colour, Odour, Taste, Turbidity, pH,TDS, Total\_Alkalinity, Chloride, Fluoride, Nitrate, Sulphate, Total\_Hardness, Total\_Hardness, Conductivity

.

|  |
| --- |
|  |

**INTRODUCTION:**

Water quality has a direct impact on public health and the environment. Water is used for various practices, such as drinking, agriculture, and industry. Among various sources of water supply, due to easy access, rivers have been used more frequently for the development of human societies. Using other water resources such as groundwater and seawater sometimes assisted with problems.

In regions where water treatment facilities are inadequate, such as developing nations and rural areas, the ability to predict water potability is of utmost importance. Traditional methods of monitoring water quality, which involve costly and time-consuming laboratory and statistical investigations, have proven to be inefficient. Therefore, there is a pressing need for a more efficient and cost-effective alternative.

In recent years, the advent of machine learning (ML) techniques has offered a promising avenue for addressing these limitations and improving the accuracy and efficiency of water quality prediction. ML algorithms, ranging from classical regression and classification models to more advanced deep learning architectures, have shown remarkable capabilities in modeling complex relationships within water quality datasets. By leveraging large volumes of historical data and incorporating various environmental variables, ML models can predict key water quality parameters with higher precision and provide valuable insights for resource management and decision-making.

**OBJECTIVES**

* The goal is to built the rigid machine learning model that can accurately predict portability of water on some inputs.
* For various reasons, predicting WQI and WQC using machine learning models is critical for assessing water suitability: Just-in-time water quality monitoring: Predictive models allow for real-time or near-real-time estimate of WQI and WQC, which is more efficient and cost-effective than standard laboratory analysis.
* **Accuracy Improvement**: Enhance the accuracy and reliability of water quality predictions compared to traditional methods by leveraging machine learning algorithms.
* **Real-Time Monitoring**: Develop models capable of providing real-time or near real-time predictions of water quality parameters, enabling timely interventions and adaptive management strategies.
* **Spatial and Temporal Resolution**: Improve spatial and temporal resolution of water quality monitoring by utilizing machine learning techniques to interpolate and extrapolate data across space and time.
* **Early Warning Systems**: Establish early warning systems for identifying and mitigating potential water quality issues, such as pollution events, algal blooms, or contaminant spills, using predictive modeling approaches.

By addressing these objectives, water quality prediction using machine learning aims to enhance our ability to monitor, manage, and protect water resources for the benefit of both human populations and ecosystems.

**LITERATURE SURVEY:**

In 2018, Ali Heidar Nasrolahi along with Amir Hamzeh Haghiabi and Abbas Parsaie predicted the Water Quality of a river bed in Iran Tireh River by taking pH, Na, Ca, Mg, etc such components into consideration. Performance was tallied by using several ML and DL algorithms. It was observed that results of SVM was the front runner and gave the best accuracy. ANN gave acceptable accuracy for practical purposes.[1]

In 2019, Umair Ahmed et.al explained ways to efficiently predict water quality using supervised Machine Learning. Harrowing diseases have been in increased proportions due to the depreciation and deterioration of water quality at an alarming rate which was a direct impact of rapid urbanization and industrialisation. Their research monitors and works with supervised Machine Learning algorithms to calculate Water Quality Index (WQI) and Water Quality Class (WQC), the former being a singular index which describes the general quality of water and the latter being the derivative and distinctive class on the basis of WQI.[2]

In 2020, Navideh Noori et.al explained the water quality prediction using SWAT-ANN coupled approach. For solving environmental problems Machine Learning algorithm such as Artificial Neural Networks is being used widely. They illustrated the application of SWAT-ANN for water quality prediction.[3]

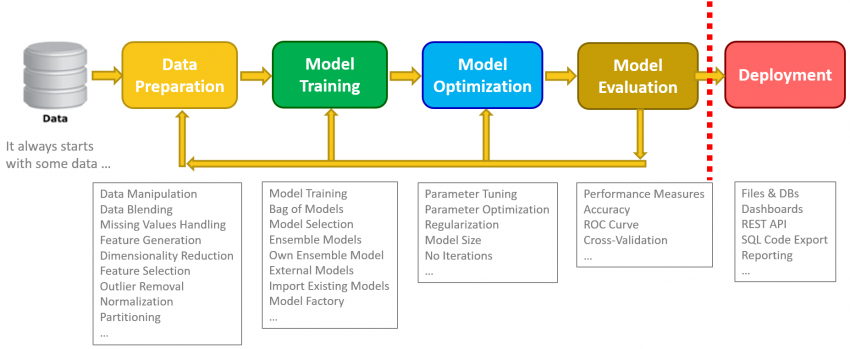
In 2022, Jin-Won Yu et.al explained the use of AI algorithms for the water quality prediction. Combined the power of data decomposition, fuzzy C-means clustering and bidirectional gated recurrent model for the prediction of water quality.[4]

In 2022, Manisha Koranga et.al discussed the use of Machine Learning Algorithms for water quality prediction for Nanital Lake, Uttarakhand. Analysed the use of machine learning algorithms and used eight regression algorithms and nine classification algorithms. Three algorithms Random Forest, SVM and Stochastic Gradient Descent comes out to be the most effective machine learning algorithms.[5]

**Project Detail:**

**Methodology:**

**1.1 Implementation process:**



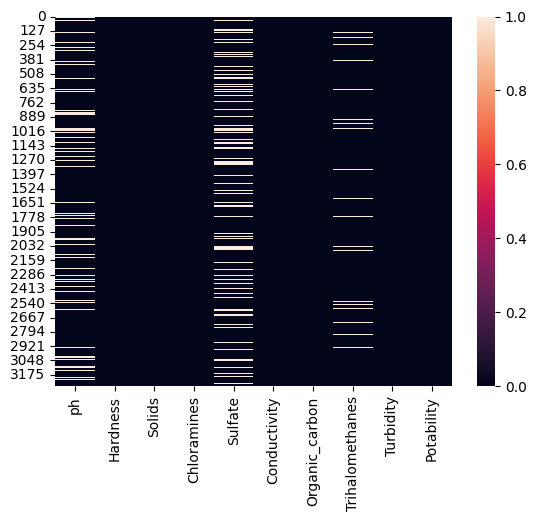
**Collecting Data:**

The present study utilized a publicly available dataset on Jal Jeevan Mission (https://jaljeevanmission.gov.in) as the primary source of data. This dataset comprises 1426 observations of water quality collected from different locations OF Karad Taluka and includes 15 distinct chemical parameters, namely Temperature, Colour, Odour, Taste, Turbidity, pH,TDS, Total Alkalinity, Chloride, Fluoride, Nitrate, Sulphate , Total Hardness, Conductivity which is used to make a prediction using various machine learning algorithms.

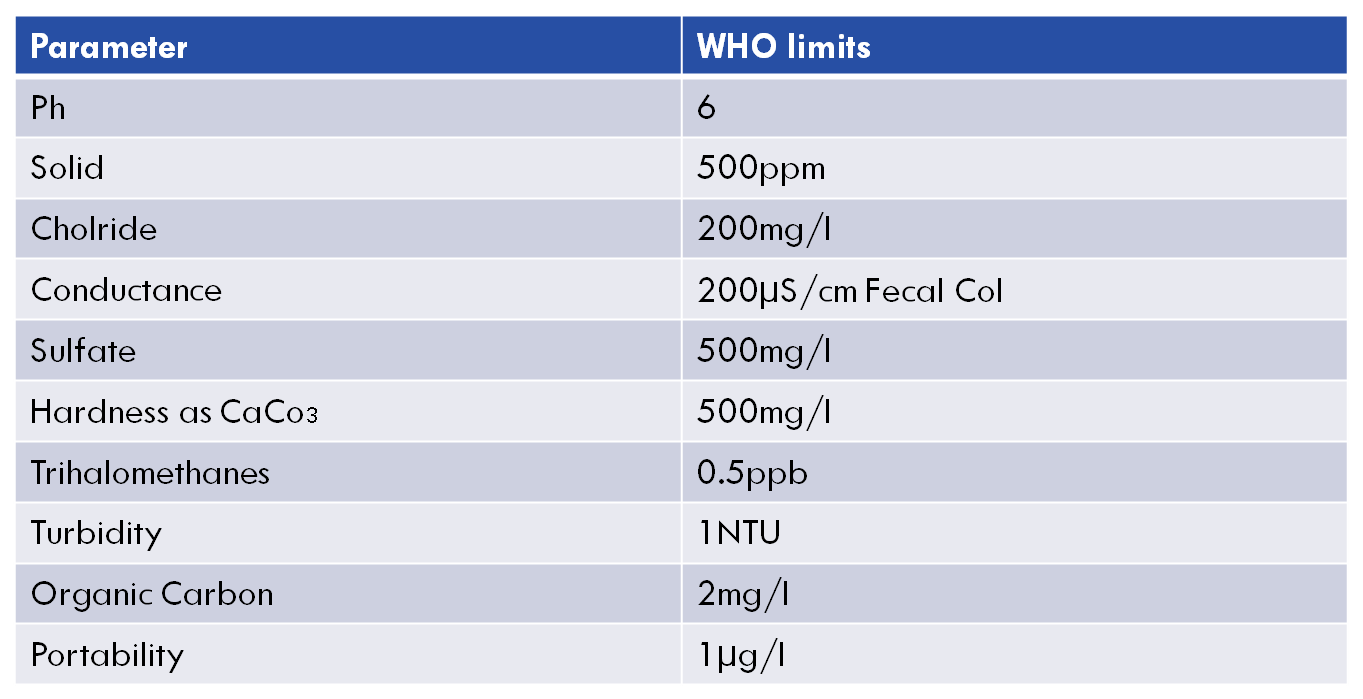
**2.1 Data Preprocessing:**

As all the null values present in dataset data cleaning carried out and the missing values are filled null value in each parameter is fill using the mean value of respective parameter

**Heatmap for Null Values**



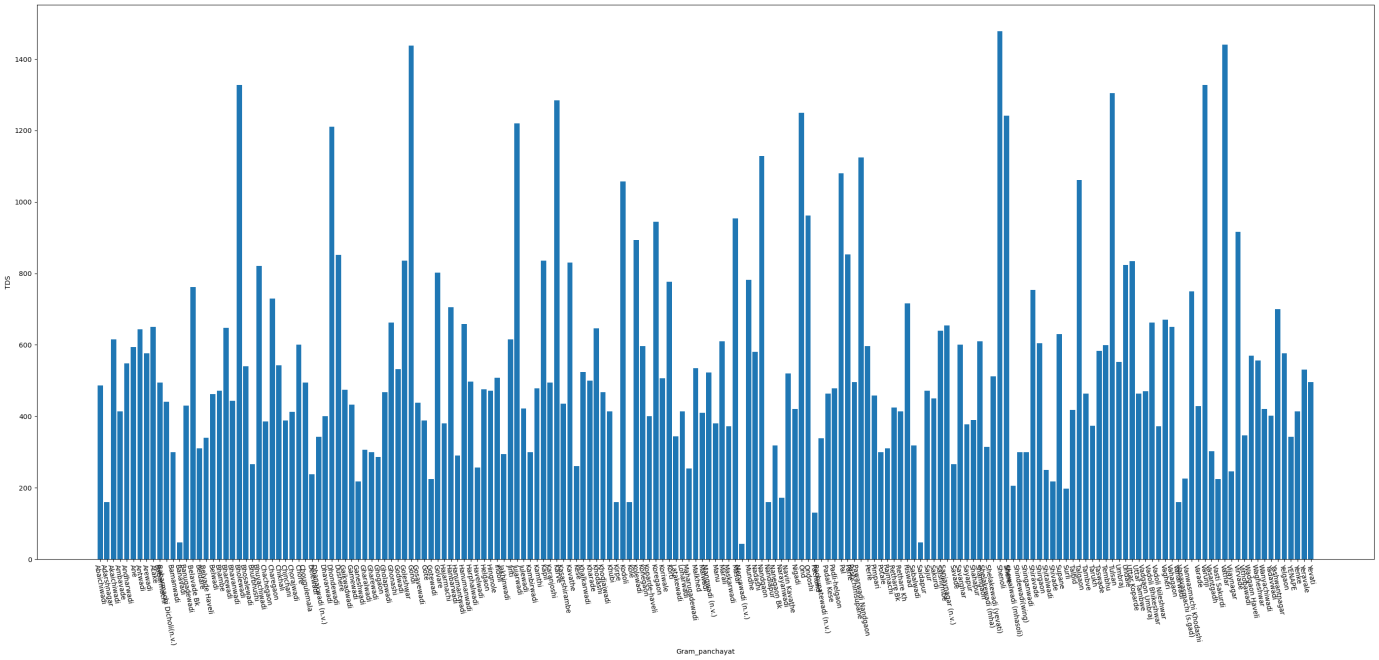
To predict whether the water is drinkable or not WHO provides some standard values for these parameters of water which is as follows:



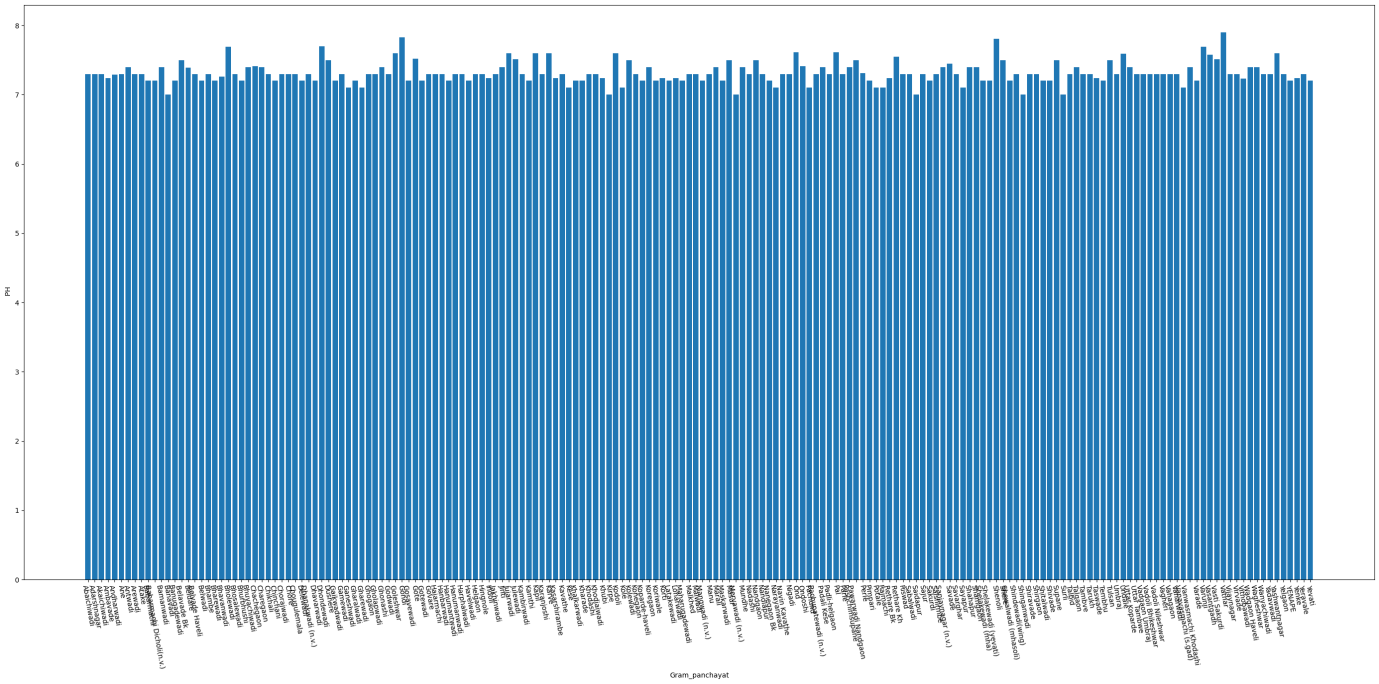
.

**3.1 Data Visualization:**

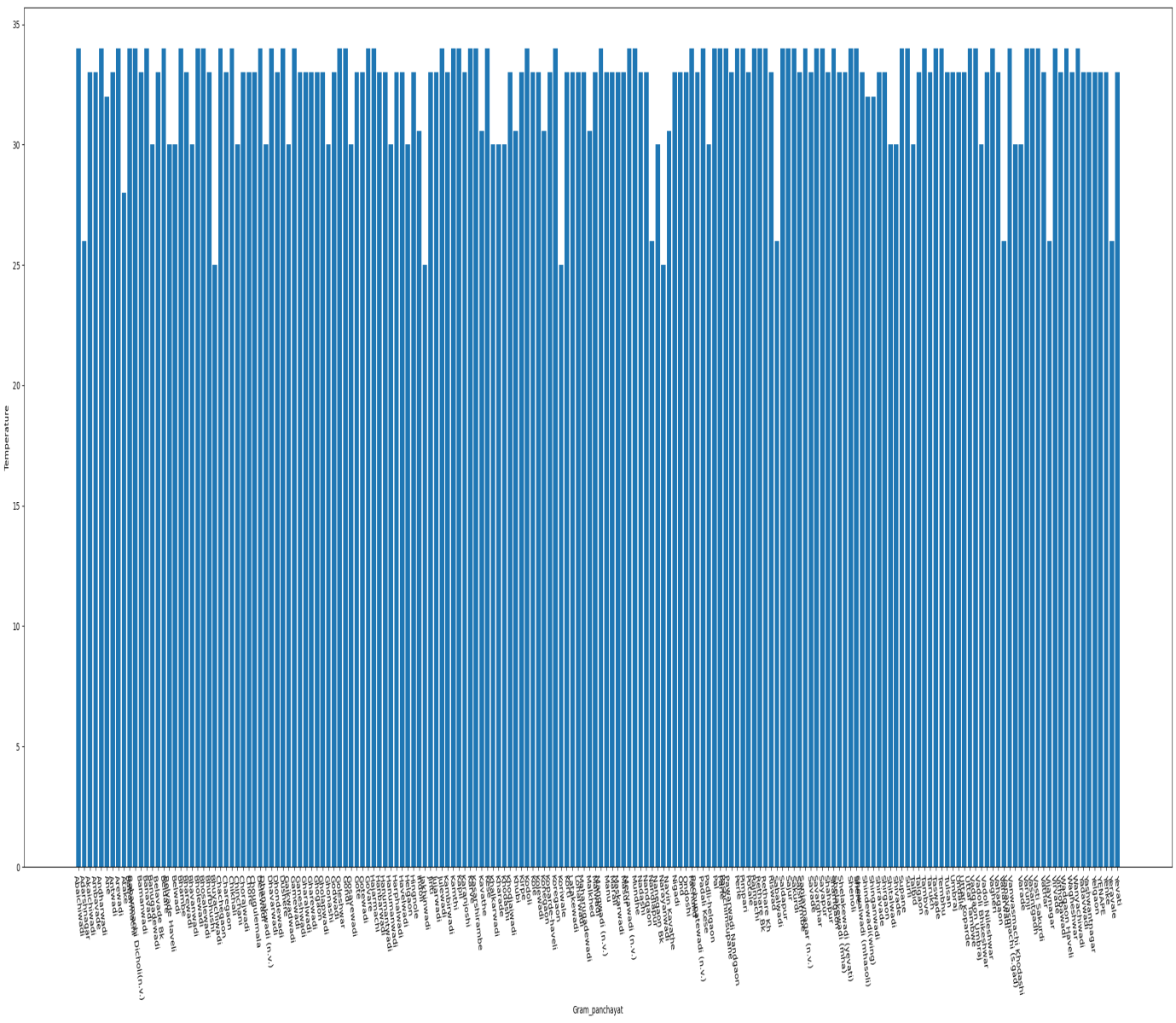
**1)TDS**

****

**2)pH**

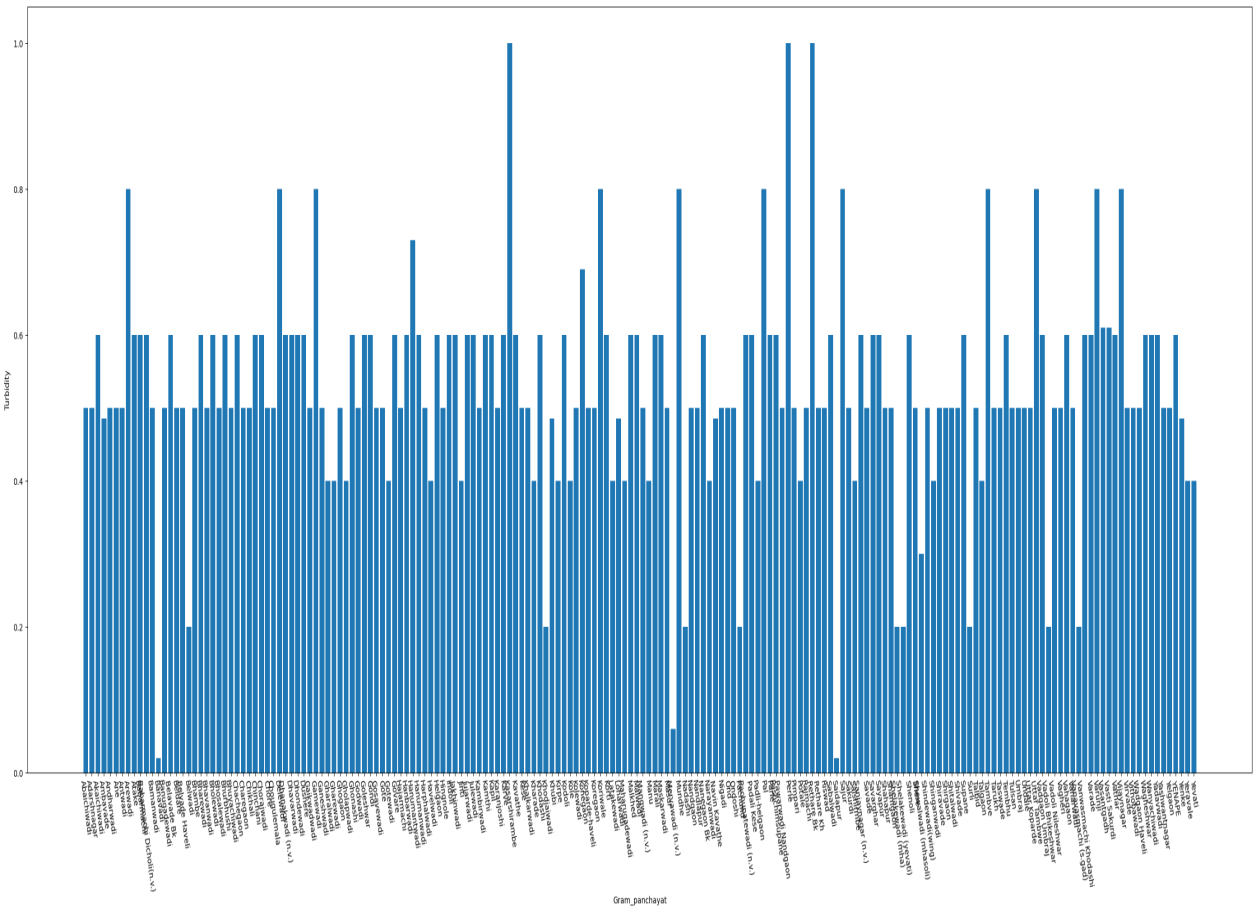
****

**3)Temperature**

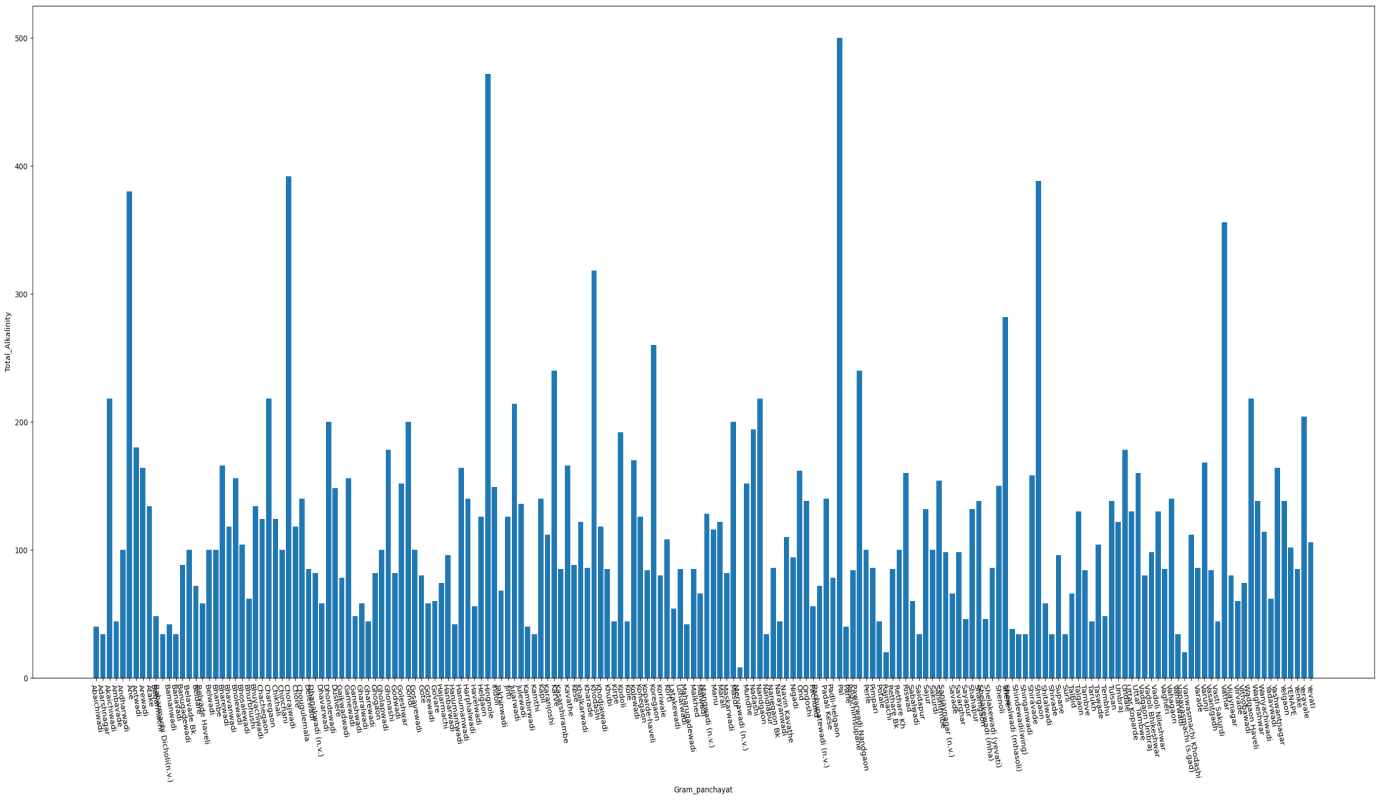
Top of Form

Bottom of Form

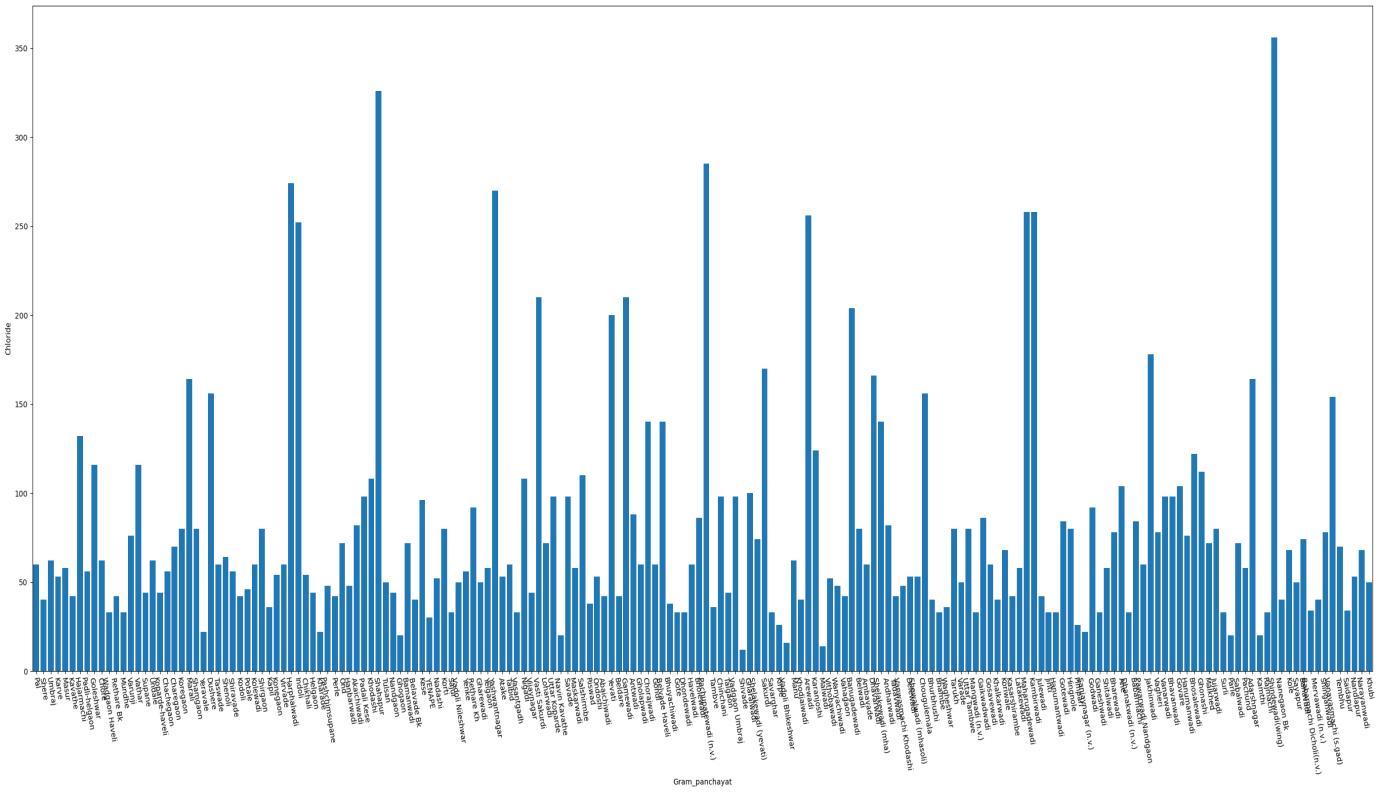
**4)Turbidity**



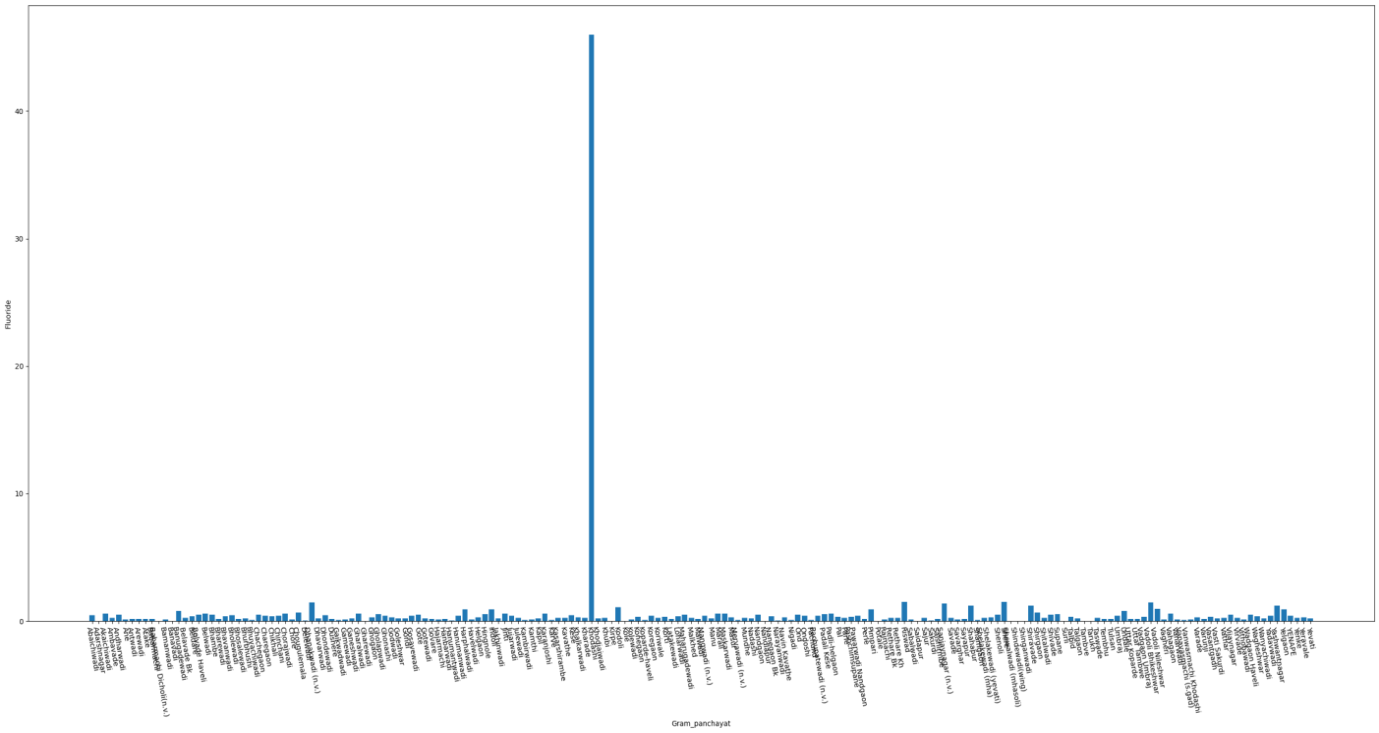
**5)Total\_Alkalinity**



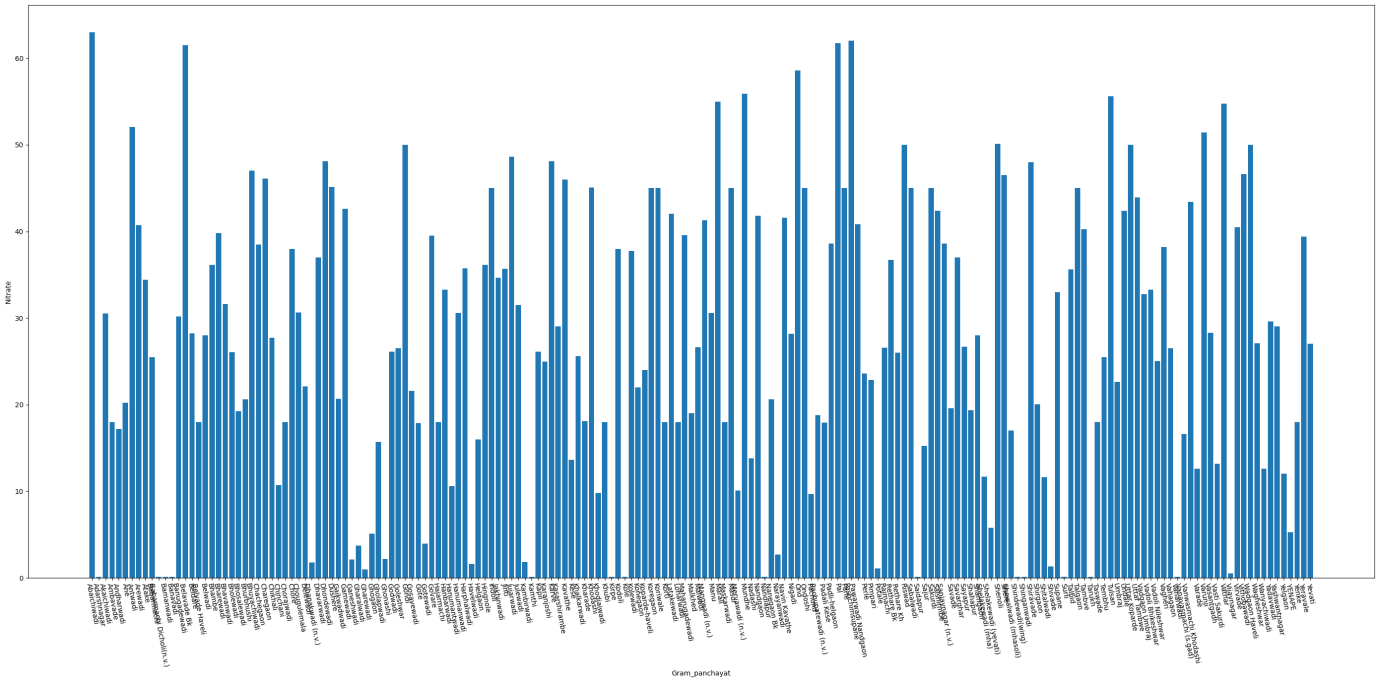
**6)Chloride**



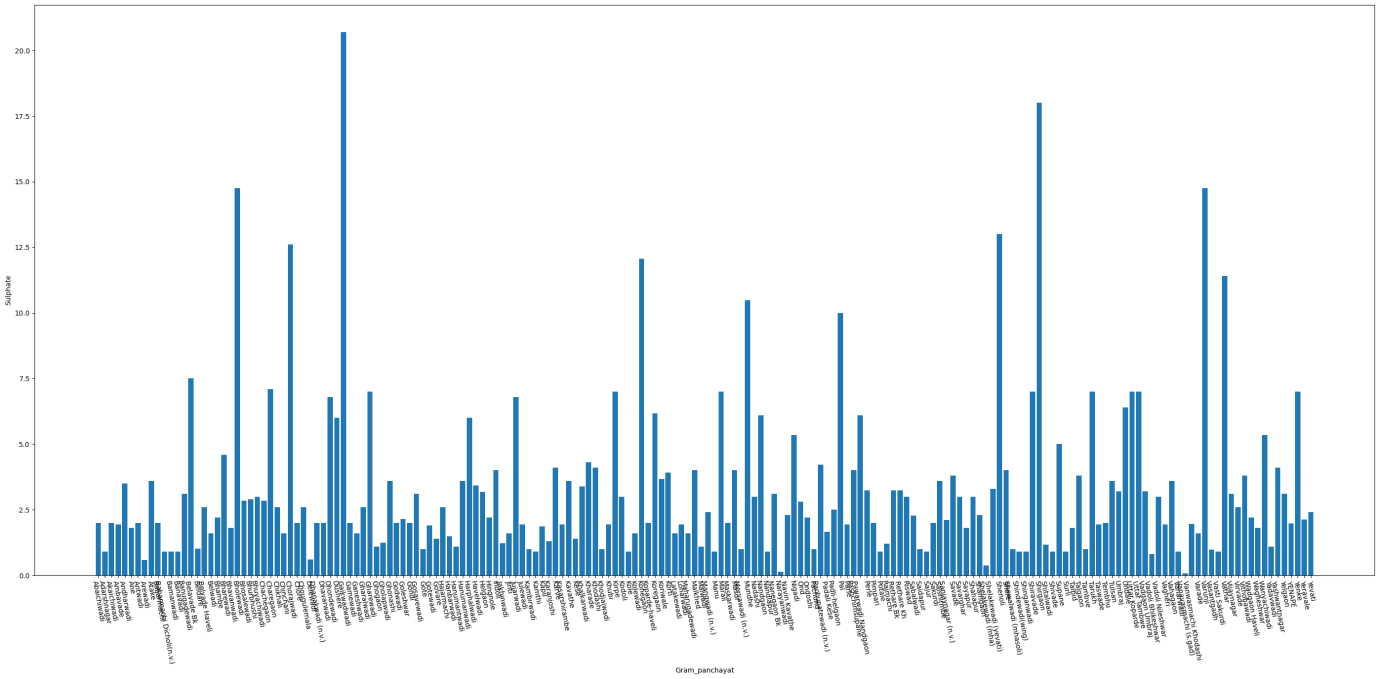
**7)Fluoride**



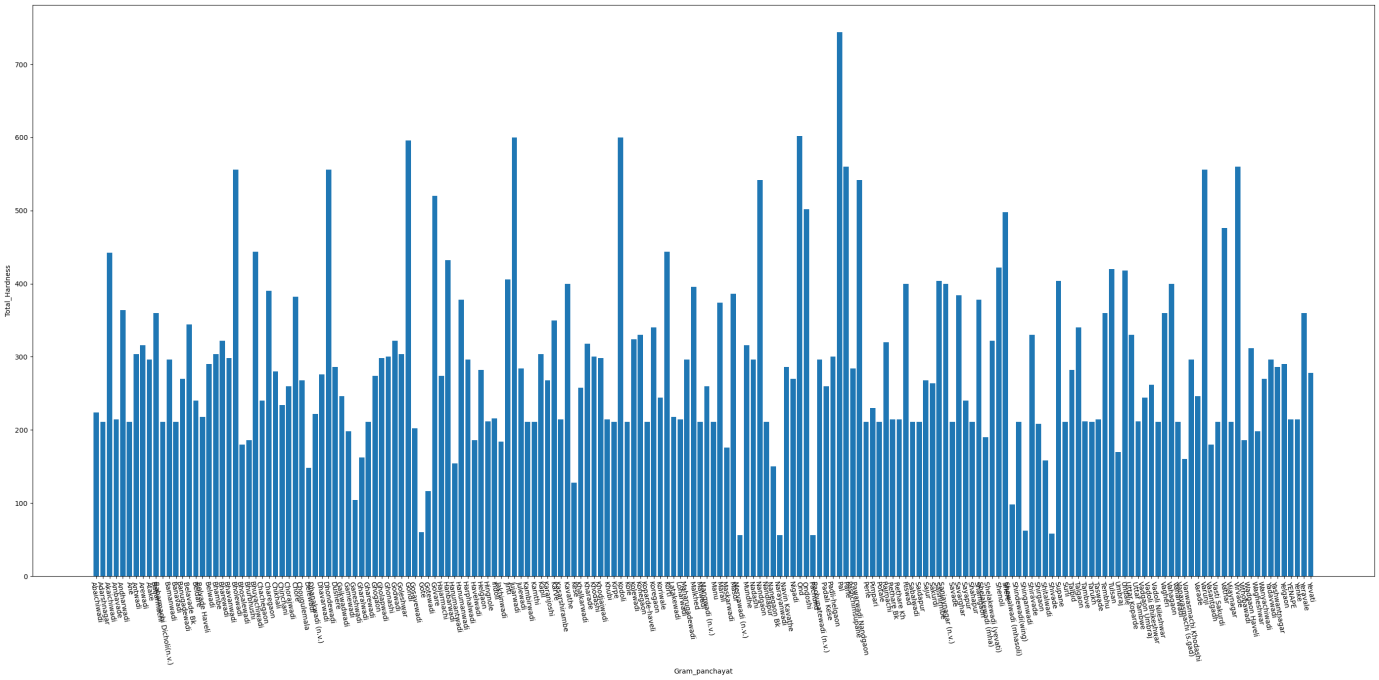
**8)Nitrate**



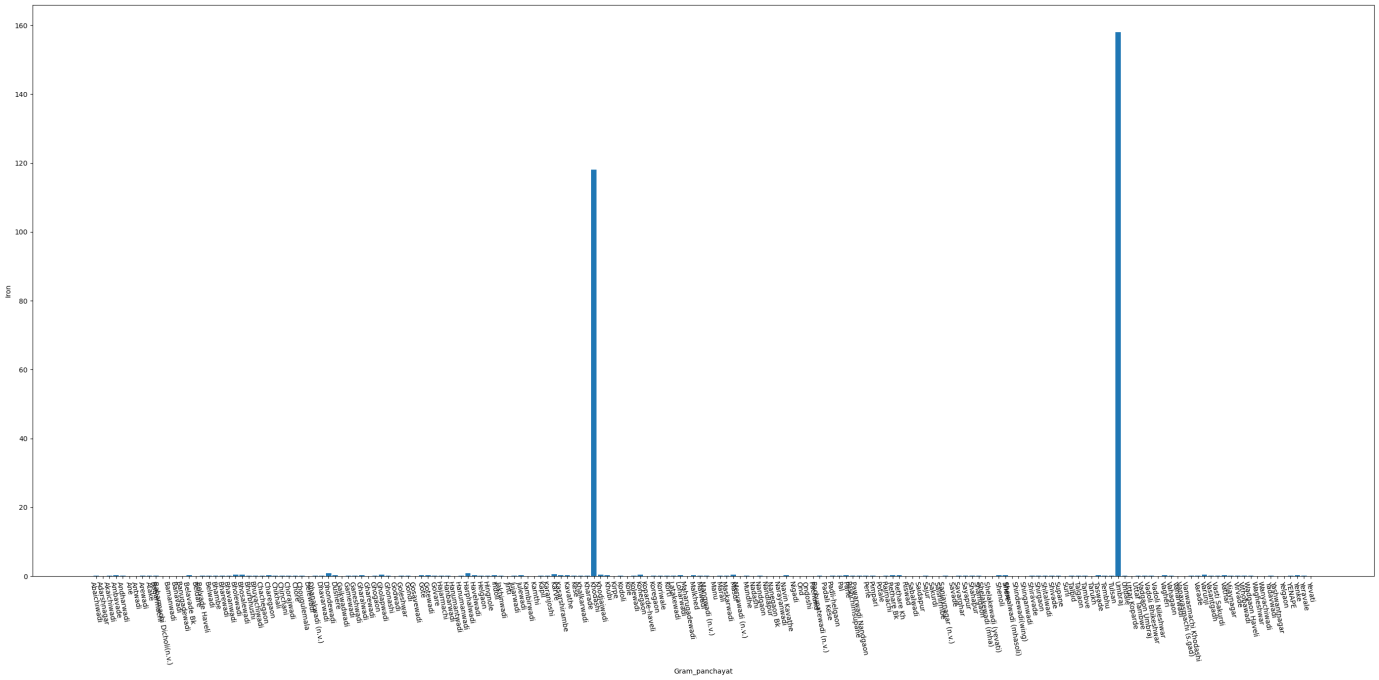
**9)Sulphate**



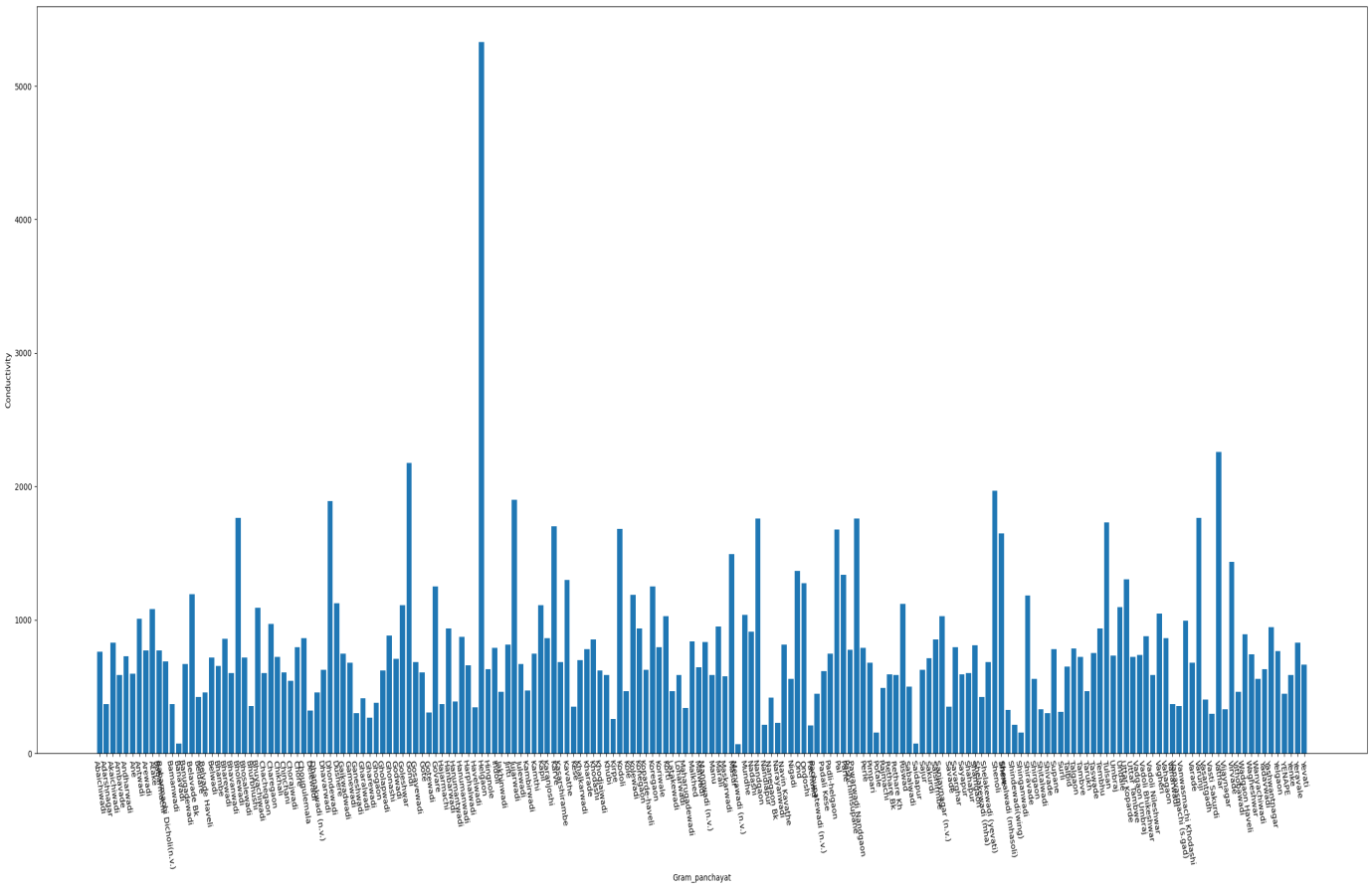
**10)Total\_Hardness**



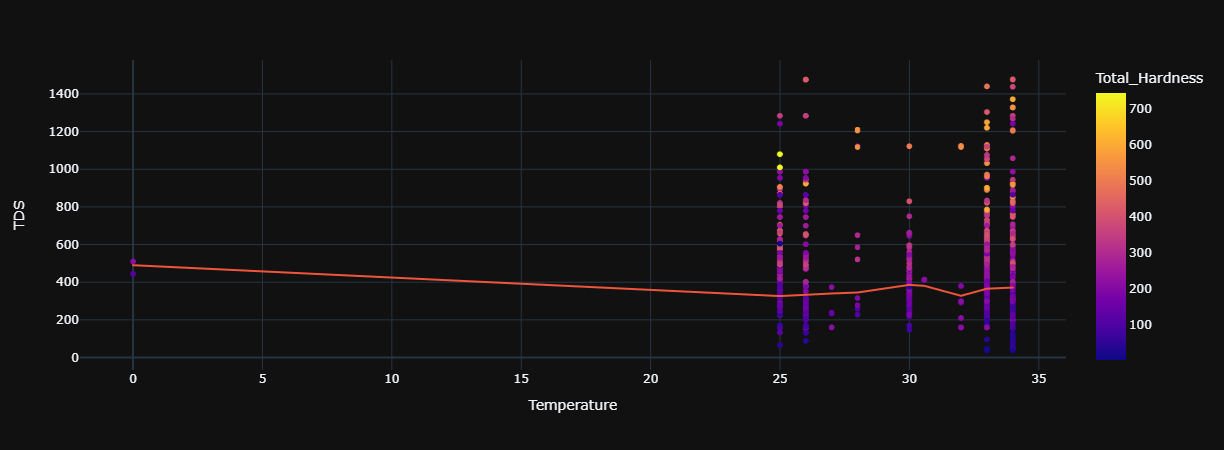
**11)Iron**



**12)Conductivity**



**Comparison between Temperature, TDS and Total\_Hardness**

****

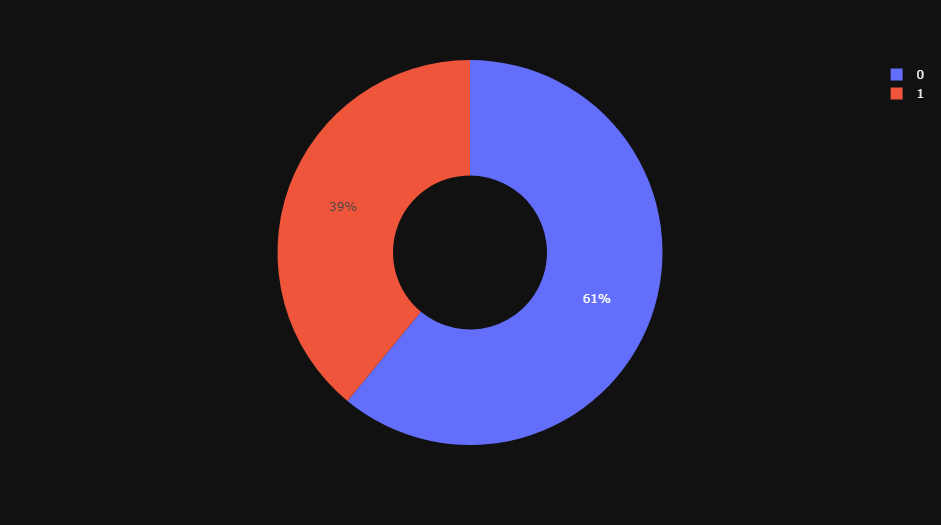
**Tools for implementation:**

* Anaconda
* Jupyter Notebook
* Python

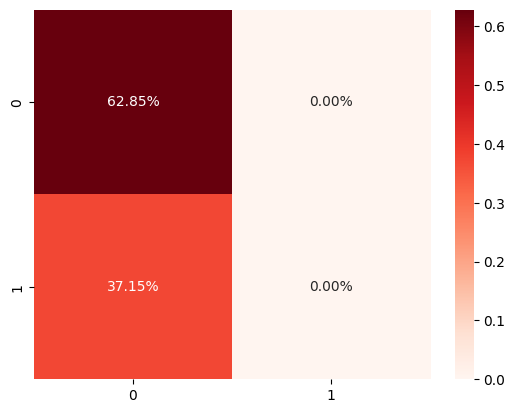
**Libraries Used:**

* Numpy : To perform wide variety of mathematical operations on arrays.
* Pandas : Provides various data structures and operations for manipulating numerical data.
* Matplotlib and seaborn:For creating static, animated and interactive visualization.
* Sklearn : Tool for predictive data analysis. Features various classification, regression and clustering algorithms.

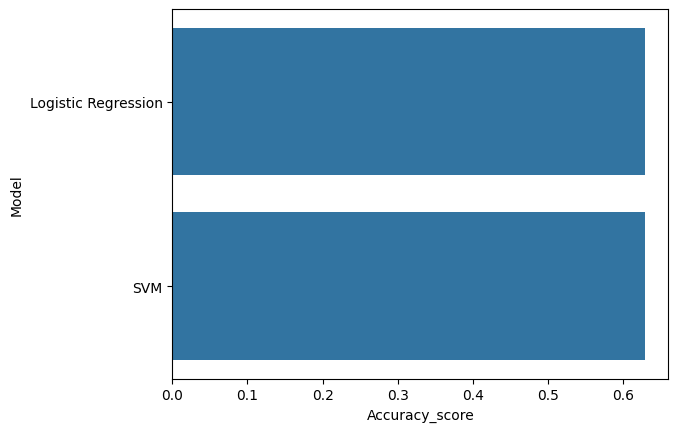
**4.1 Percentage of Potablity and Not-Potablity**



**5.1 Confusion Matrix for the Support Vector Machine**



**6.1 Accuracy of the support vector machine and Logistic regression**



**Advantages:**

1.Improved Accuracy: Machine learning algorithms can capture complex nonlinear relationships and patterns in water quality data, leading to more accurate predictions compared to conventional statistical approaches.

2.. Real-Time Monitoring: ML models can provide real-time or near real-time predictions of water quality parameters, enabling timely interventions and proactive management strategies to mitigate potential risks to human health and the environment.

3. Scalability and Generalizability: Machine learning techniques are scalable and can be applied across different geographic regions and environmental conditions, accommodating variations in data availability and quality.

4. Cost and Time Efficiency: Once trained, machine learning models can automate the prediction process, reducing the need for labor-intensive and time-consuming manual monitoring efforts, thereby saving resources and increasing operational efficiency.

Overall, water quality prediction using machine learning holds great promise for improving our understanding of aquatic ecosystems, facilitating informed decision-making, and safeguarding the health and sustainability of water resources for future generations.

**Conclusion:**

As we all know the importance of water for the human body. So knowing the Quality of the water is very much necessary because if we drink water without knowing that it is safe for drinking we could get sick. There are plenty of water-borne diseases like Cholera, Typhoid, Giardia, E. Coli, Hepatitis A, and so on. These types of diseases happen if we drink non-drinkable water. So knowing the quality of the water is the most important thing. But the main problem lies here. For testing the water quality we have to conduct lab tests on the water which is costly and time-consuming as well.

The Machine learning Algorithms have been used to design the water quality analysis model which predict wheather the water is safe for drinking or not.

In summary, water quality prediction using machine learning holds immense promise for addressing the challenges of water resource management in the 21st century. By harnessing the power of data-driven modeling techniques, we can strive towards a future where clean and safe water is accessible to all, ensuring the sustainability of ecosystems and the well-being of communities worldwide.

Top of Form

Bottom of Form

**REFERENCES:**

[1] Haghiabi, A. H., Nasrolahi, A. H., &Parsaie, A. (2018). Water quality prediction using machine learning methods.

[2] Umair Ahmed, Rafia Mumtaz, Hirra Anwar, Asad A. Shah, Rabia Irfan and Jose Garcia-Nieto (2019). Efficient Water Quality Prediction Using Supervised Machine Learning.

[3] Navideh Noori, Latif Kalin and Sabahattin Isik (2020). Water Quality prediction using SWAT-ANN coupled approach.

[4] Jin Won Yu, Ju-Song Kim, Xia Li, Yun-chol Jong, Kwang-Hun Kim and Gwang-Il Ryang (2022). Water quality forecasting based on data decomposition, fuzzy clustering and deep learning neural network.

[5] Manisha Koranga, Pushpa Pant, Tarun Kumar, Durgesh Pant, Ashutosh Kumar Bhatt and R.P. Pant (2022). Efficient water quality prediction models based on machine learning algorithms for Nainital Lake, Uttarakhand.