**+Ideation Phase**

**Defining the Problem Statements**

|  |  |
| --- | --- |
| **Date** | **26-09-2023** |
| **Team ID** | **3933** |
| **Project Name** | **Water Quality Analysis** |

**Water Quality Analysis**

**Problem Definition and Design Thinking**

**Introduction**

This document is about to develop a machine learning model that can accurately water quality analysis based on a set of relevant features. Water quality analysis is a common problem in the public health and for helping **water service authorities, offices, restaurants and residential users**

In this document, we will outline the problem statement, the steps involved in solving it, and the design thinking approach that will guide our project.

**Problem Statement**

Objective: To identify any potential issues or deviations from regulatory standards, and determine the potability of the water based on (PH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, Turbidity, Potability) parameters.

Data: We have a data represented by the given columns (pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic carbon, Trihalomethanes, Turbidity, Potability) to assess the suitability of the water for specific purposes such as drinking. This data will be used to train and evaluate our machine learning model.

**Key Challenges:**

1. Data Quality: Ensuring the dataset is clean, complete, and free of errors.

2. Feature Selection: Identifying the most relevant features for accurate potability predictions.

3. Model Selection: Choosing the appropriate machine learning algorithm(s) for the task.

4. Model Evaluation: Evaluating the model's performance using appropriate metrics.

5. Deployment: Creating a user-friendly interface or API for end-users to make predictions.

**Design Thinking Approach**

**Empathize:**

Before diving into solving the problem, it's crucial to empathize with the users and understand their needs. In this case, our primary users **water service authorities, offices, restaurants and residential users**. We need to gather insights into what factors are most important to them analysis the water quality

**Actions:**

- Conduct water quality testing and maintains record of treatment processes in public water utilities such

- Analyse historical water quality data of drinking water source and water treatment plant records

- Seek reports of water quality issues and complaints from consumers are valuable sources of data.

**Define:**

Based on our understanding of the problem and the users' needs, we will define clear objectives and success criteria for our project.

**Objectives:**

- Develop a machine learning model that achieves a Mean Absolute Error (MAE) of less than $X on the test data.

- Create a user-friendly web application for users to input water quality parameters and receive water quality predictions.

**Ideate:**

Brainstorm potential solutions and approaches to address the problem. This phase involves thinking creatively and considering various algorithms and techniques for water quality analysis

**Actions:**

- Explore different water quality algorithms such as parameter measurement techniques, water quality indices, data analysis algorithms.

- Experiment with feature engineering techniques to enhance model performance.

- Consider incorporating external data sources (e.g., climate and weather data, Historical data) to improve predictions.

**Prototype**

Create a prototype of the Hydrological models and the user interface for water quality prediction.

**Actions:**

- Develop a IBM cognos, Jupyter Notebook or Python script for data pre-processing, model training, and evaluation.

- Create a simple web interface using tools like Flask or Django to allow users to input water quality details(like PH, Hardness, Turbidity and so on).

- Test the prototype with a subset of the dataset to ensure it meets performance objectives.

**Test**

Evaluate the model's performance using appropriate metrics and gather feedback from users.

**Actions:**

- Split the dataset into training and testing sets.

- Train the model on the training set and evaluate it on the testing set.

- Use metrics such as Data Quality, Root Mean Square Error (RMSE), and R-squared to assess model performance.

- Collect user feedback on the web interface for usability and accuracy.

**Implement**

Once the prototype meets the defined objectives and receives positive feedback, proceed with full implementation.

**Actions:**

- Train the final Hydrological model on the entire dataset.

- Deploy the model as part of a production-ready web application.

- Conduct thorough testing to ensure the application is robust and user-friendly.

**Iterate**

Continuous improvement is essential. Gather user feedback and iterate on the model and interface to enhance accuracy and usability.

**Actions:**

- Monitor the model's performance and retrain it periodically with updated data.

- Address user feedback and make necessary improvements to the web interface.

- Stay informed about advancements in machine learning

**Conclusion**

This document which outlines how to solve the problem of water quality analysis by using machine learning. It defines the problem, identified key challenges, and laid out a design thinking approach that involves empathizing with users, defining objectives, ideating potential solutions, prototyping, testing, implementing, and iterating.

The ultimate goal is to develop an accurate and user-friendly solution that provides valuable insights for potable water. By following this structured approach, we have to create a reliable tool that contributes for the best water quality analysis.