

Project Understanding and Approach Document

Assessing Water Quality for Potability

Table of Contents

1. Introduction
2. Project Objectives
3. Data Collection
4. Data Preparation
5. Analysis Approach
6. Data Visualization Strategy
7. Innovation
8. Development - Part 1
9. Development - Part 2
10. Project Documentation & Submission

1. Introduction

This project aims to delve into water quality analysis, specifically to determine its suitability for various purposes, especially drinking. Our primary goal is to identify potential problems or deviations from established standards and then decide whether the water is potable based on multiple parameters. This document provides an overview of our project's understanding and outlines our planned approach.

2. Project Objectives

Our main objectives for this project are as follows:

- Clearly define the analysis objectives, which include assessing water potability, spotting any deviations from established standards, and grasping the relationships between different water quality parameters.
- Collect relevant water quality data containing parameters such as pH, Hardness, Solids, and others.
- Develop a visualization strategy to effectively illustrate parameter distributions, correlations, and water potability.
- Select suitable machine learning algorithms and features to construct a predictive model for water potability.
- Optionally, explore the incorporation of anomaly detection techniques to identify unusual patterns in water quality parameters.

- Execute the project in two phases, commencing with data preprocessing and exploratory data analysis in Part 1, followed by visualization and predictive model creation in Part 2.
- Meticulously document the entire project to ensure it is ready for submission.

3. Data Collection

Our initial step involves sourcing water quality data from reliable datasets, ensuring that it encompasses crucial parameters like pH, Hardness, Solids, Chloramines, Sulfate, Conductivity, Organic_carbon, Trihalomethanes, and Turbidity.

4. Data Preparation

Prior to launching our analysis, it is essential to adequately prepare the data. This entails:

- Cleansing the data to rectify inconsistencies and address any missing values.
- Aggregating the data to facilitate the calculation of mean and standard deviation values.
- Grouping the data by country for a country-specific analysis.

5. Analysis Approach

Our analytical approach encompasses the following key steps:

a. Calculate Mean and Standard Deviation

- Determine the mean and standard deviation for daily water quality parameters across the entire dataset.
- Compute the mean and standard deviation for each parameter and individual country.

b. Comparison and Contrast

- Perform comparisons of means and standard deviations across parameters to assess water potability.
- Contrast means and standard deviations between various countries for deeper insights.

6. Data Visualization Strategy

To effectively communicate our findings, we will devise a comprehensive data visualization strategy, including:

- Employing line charts to depict trends in parameter values over time.
- Utilizing bar charts to facilitate comparisons of means and standard deviations.
- Employing box plots to identify any outliers.
- Creating maps for a geographic representation of water quality by country.

7. Innovation

In addition, we may consider incorporating anomaly detection techniques, an optional step, to uncover any unusual patterns within the water quality parameters.

8. Development - Part 1

The initial development phase primarily entails data preprocessing and conducting exploratory data analysis to ensure that our dataset is primed for subsequent analysis.

9. Development - Part 2

The second development phase will focus on crafting visualizations and establishing a predictive model for water potability.

10. Project Documentation & Submission

We will conclude the project by meticulously documenting the entire process. This documentation will encompass our objectives, data preprocessing, insights from visualizations, modeling methodologies, and results. The ultimate goal is to ensure that our project is well-structured and readily comprehensible for submission.

This document provides a comprehensive overview of our approach to analyzing water quality for potability, serving as our project roadmap for successful execution.