Naan Mudhalvan

1105-Gojan School of Business and Technology

Course: Data Analytics in Process Industries

Project Title: Classification of Water Safety

Submitted by

Abinaya R-110523237001

Akash M-110523237002

Ashok kumar A-110523237003

Ashwathi R-110523237004

Divya Bhavani R-110523237005

Water Safety Classification

Introduction:

Water safety is a critical issue worldwide, addressing the need for access to clean and safe water for human consumption, sanitation, and recreational purposes. Waterborne diseases and contamination can cause significant health risks, especially in regions with inadequate water treatment infrastructure. Ensuring water safety requires a comprehensive.

Skills Gained:

- Data Pre-processing & Cleaning
- Exploratory Data Analysis (EDA)
- Feature Engineering
- Machine Learning Model Development
- Model Evaluation
- Streamlit Application Development

Problem Statement:

Despite advancements in water treatment and safety measures, waterborne diseases, contamination, and unsafe water practices continue to pose a serious threat to public health. The problem lies in the classification and effective management of water sources, as improper treatment, contamination, and neglect can cause widespread health risks. There is a need for a clear and structured classification system for water safety to enhance public awareness, regulatory practices, and intervention strategies.

Domain:

- Water Quality & Environmental Health

Motivation:

Water is a fundamental resource required for survival, yet globally, millions of people suffer from the lack of access to clean, safe drinking water. Contaminated water is responsible for many preventable diseases, especially in developing countries. The motivation for this study is to develop a robust classification system that can help in identifying unsafe water sources, improving water treatment processes, and educating the public on water safety standards. With increasing global challenges like climate change, industrial waste, and population growth, the need to ensure safe and sustainable water management has never been more urgent.

Water Safety Classification

Goal:

The goal of this study is to classify water safety based on various factors including water quality, source, usage, and potential health risks. By developing a clear, accessible, and structured classification system, the study aims to:

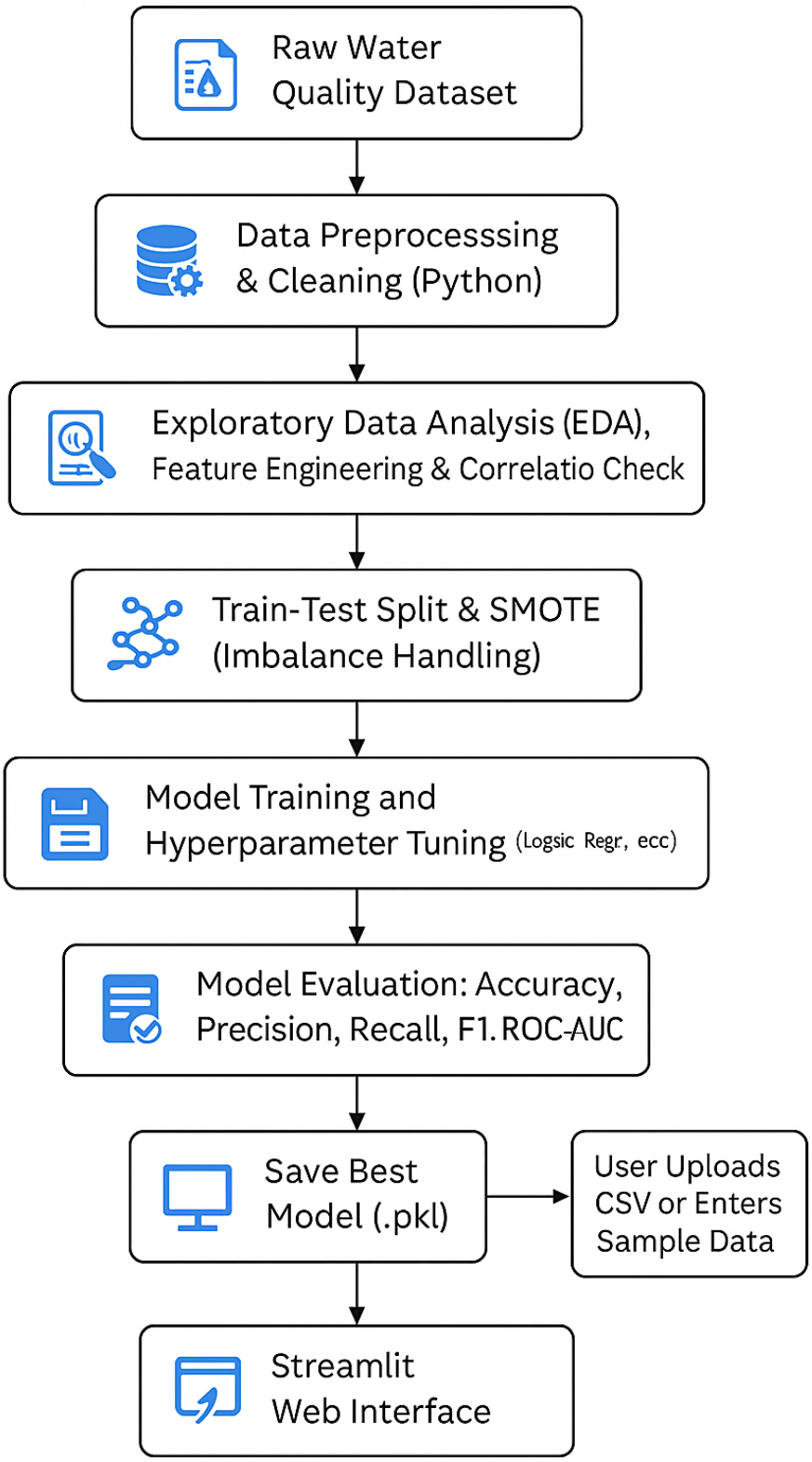
- Identify and categorize safe and unsafe water sources.
- Provide guidelines for water safety across different regions and usage types.
- Enhance public awareness about water contamination and treatment methods.
- Support policymakers in developing regulations and guidelines to ensure clean water access.

Methodology

The methodology for classifying water safety involves:

- **Data Collection:** Gathering data on water quality (e.g., chemical composition, biological contamination), sources, and geographical locations.
- **Classification Framework:** Developing a classification system based on:
  - Water quality (e.g., potable vs. non-potable)
  - Risk level (e.g., low-risk, moderate-risk, high-risk water)
  - Water source (e.g., surface water, groundwater, municipal supply)
- **Survey and Field Analysis:** Conducting field surveys and water testing to assess the safety of various water sources.
- **Regulatory Comparison:** Reviewing water safety regulations from international bodies (e.g., WHO, EPA, EU) and national standards to integrate into the classification.
- **Evaluation and Review:** Evaluating the effectiveness of the classification system in different contexts (e.g., urban vs. rural areas, developing vs. developed countries).

System Architecture:



Water Safety Classification

Technical Tags:

- Machine Learning
- Classification
- Water Quality
- EDA
- Imbalanced Data
- Streamlit
- Python
- Scikit-learn

Advantages:

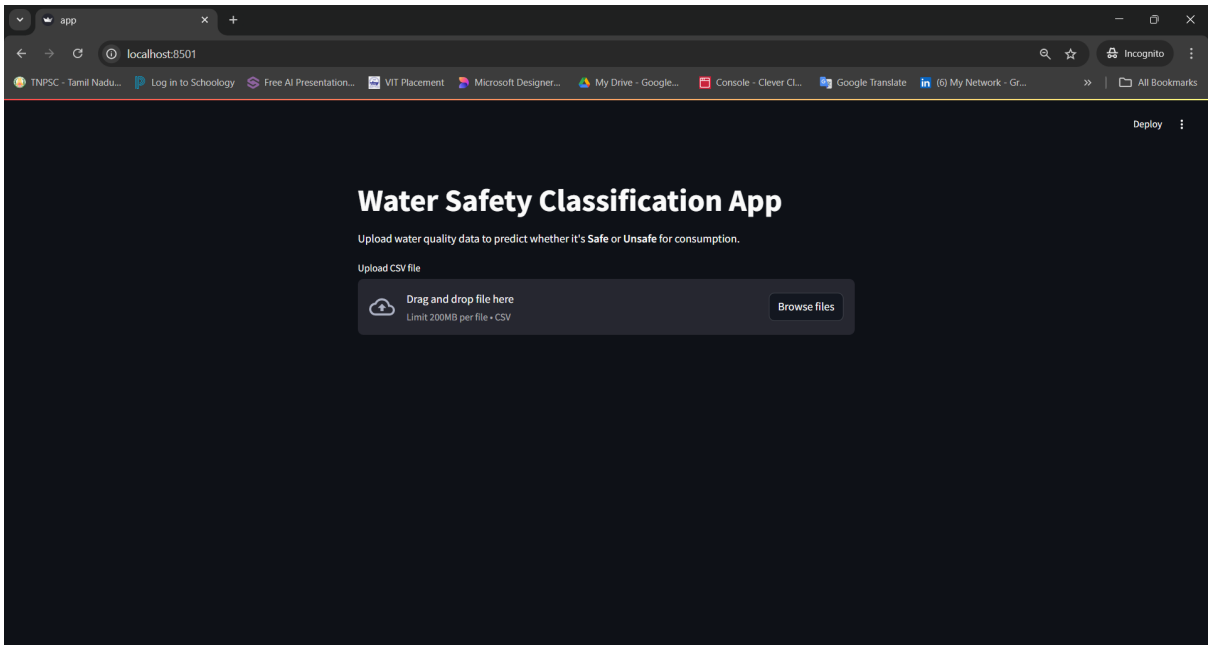
- **Improved Public Health:** By identifying safe and unsafe water sources, this classification system helps reduce the spread of waterborne diseases.
- **Better Resource Management:** Water resources can be allocated and managed more effectively, ensuring sustainable use.
- **Policy Support:** Helps governments and regulatory bodies in creating policies and standards to improve water safety.
- **Public Awareness:** Educates the general public on recognizing unsafe water and taking preventive actions, such as boiling or filtering water.
- **Early Warning:** Can help in the early identification of contaminated water sources, minimizing the risk of widespread outbreaks.

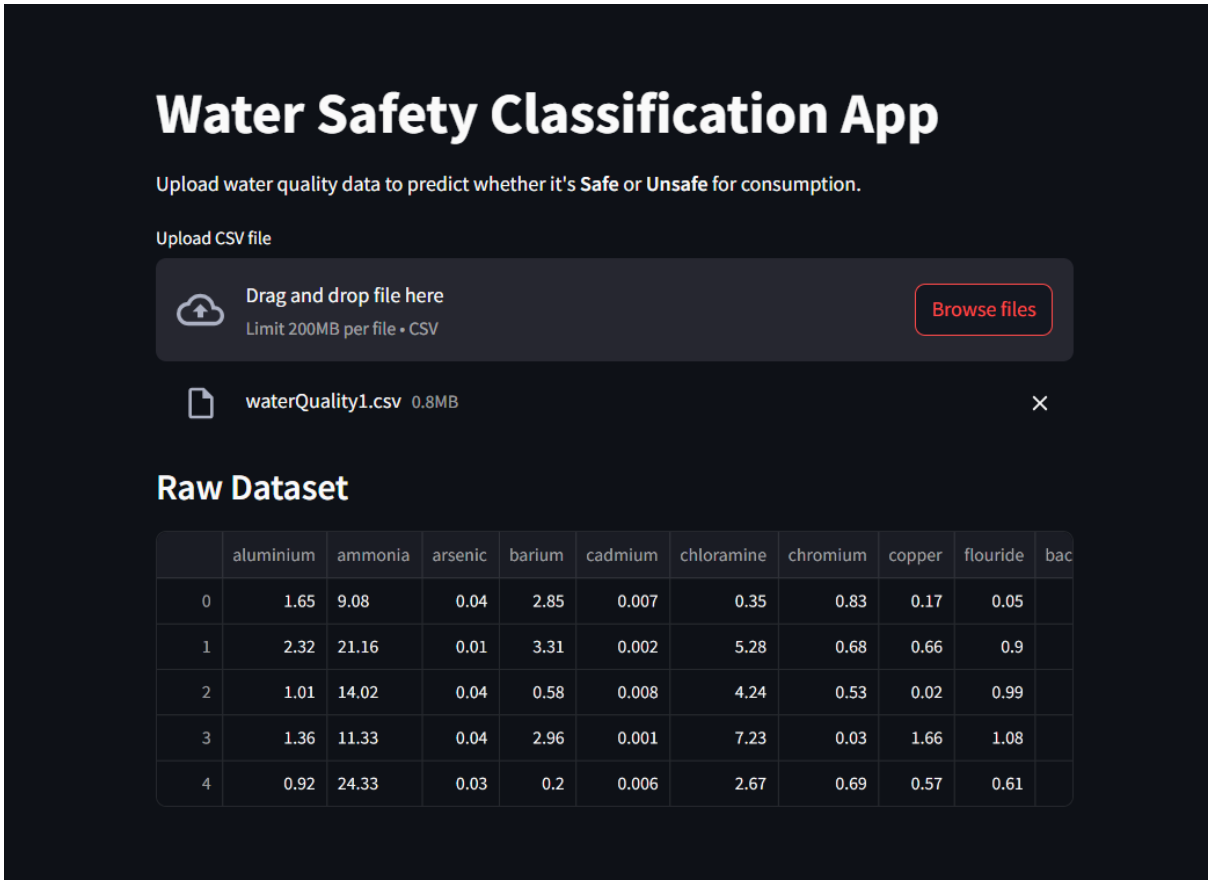
Disadvantages:

- **Data Limitations:** Inaccurate or incomplete data on water quality can hinder the effectiveness of the classification system.
- **Implementation Complexity:** In some regions, especially rural or underdeveloped areas, implementing safety standards and classifications may be challenging due to a lack of infrastructure.
- **Cost:** The process of testing and monitoring water quality regularly can be expensive for governments and institutions, particularly in resource-constrained regions.
- **Regional Variability:** Water safety classifications may not always be universally applicable due to differences in climate, geography, and local environmental conditions.
- **Resistance to Change**: Communities accustomed to certain water practices may resist new classifications or recommendations.

Water Safety Classification

Result:





	aluminium	ammonia	arsenic	barium	cadmium	chloramine	chromium	copper	flouride	bac
0	1.65	9.08	0.04	2.85	0.007	0.35	0.83	0.17	0.05	
1	2.32	21.16	0.01	3.31	0.002	5.28	0.68	0.66	0.9	
2	1.01	14.02	0.04	0.58	0.008	4.24	0.53	0.02	0.99	
3	1.36	11.33	0.04	2.96	0.001	7.23	0.03	1.66	1.08	
4	0.92	24.33	0.03	0.2	0.006	2.67	0.69	0.57	0.61	

Water Safety Classification

Model Evaluation Metrics

Accuracy: 0.96

Precision: 0.83

Recall: 0.82

F1 Score: 0.83

ROC AUC Score: 0.99

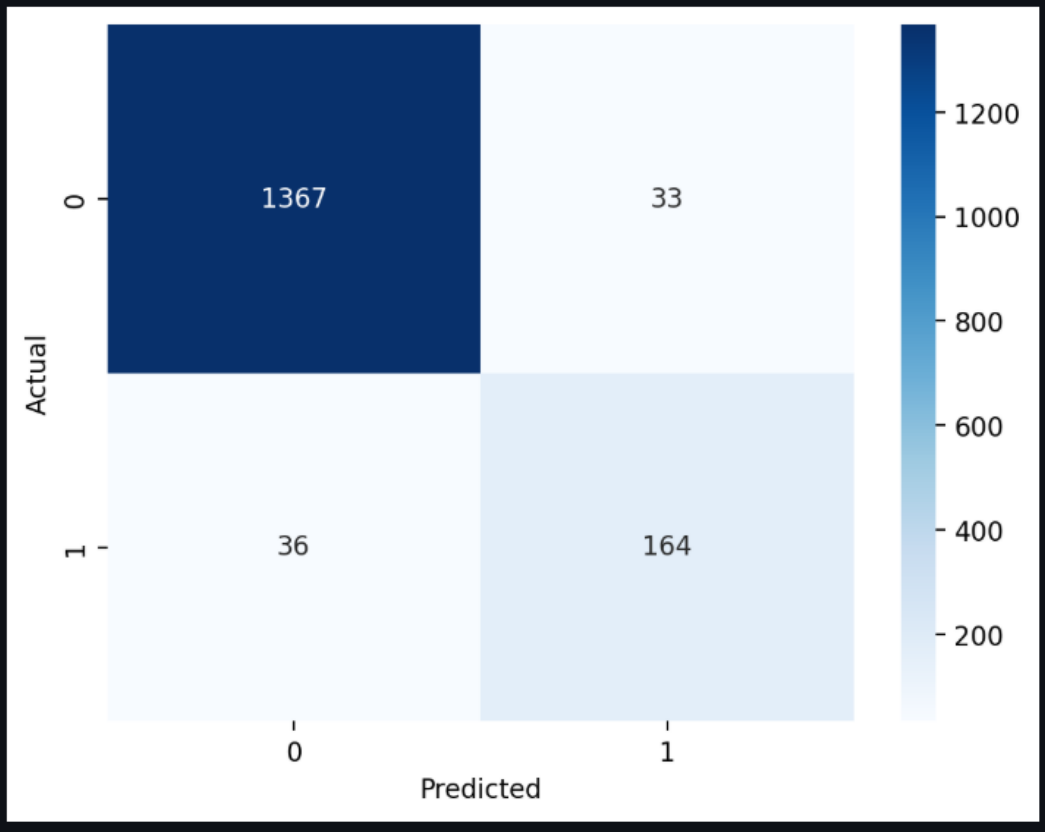
Classification Report:

precision recall f1-score support

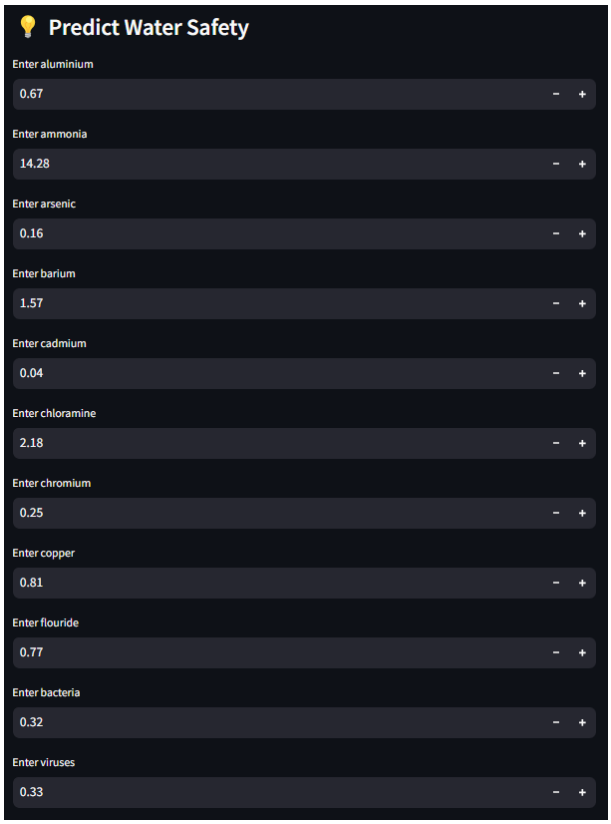
0 0.97 0.98 0.98 1400
1 0.83 0.82 0.83 200

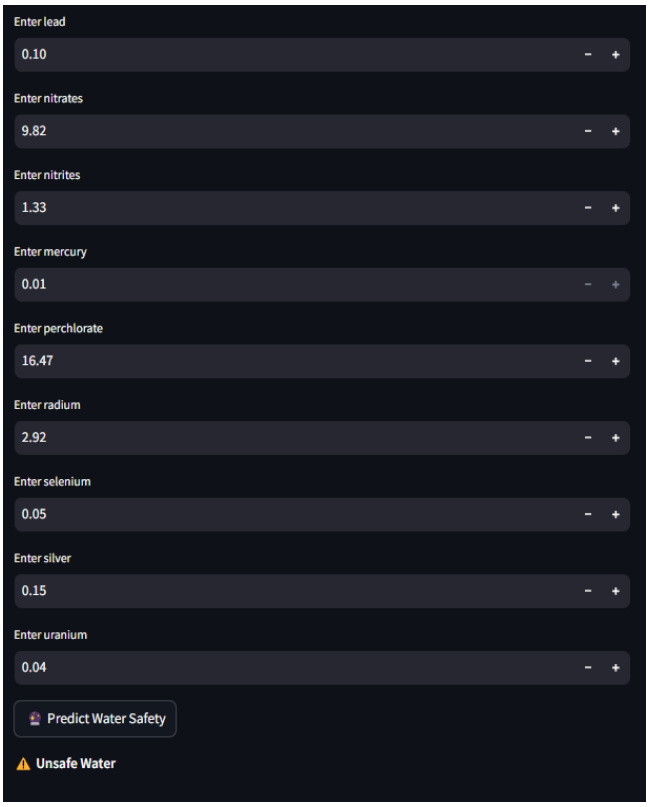
accuracy 0.96 1600
macro avg 0.90 0.90 0.90 1600
weighted avg 0.96 0.96 0.96 1600

Confusion Matrix



Water Safety Classification





Conclusion:

The classification of water safety is an essential step toward ensuring that clean, safe water is accessible to all. By categorizing water based on its quality, source, and risk levels, it is possible to identify areas requiring intervention and improve water management practices. This study's proposed classification system can help reduce waterborne diseases, facilitate effective water treatment and management policies, and promote public awareness of water safety. However, the successful implementation of this classification system will require continuous monitoring, collaboration among stakeholders, and the integration of new technologies and regulatory frameworks.