Water Quality Classification Project

• **Abstract:**

Safe drinking water is a basic human right and a component of a good health-protection strategy. On a national, regional, and local level, this is a major health and development concern. Investments in water supply and sanitation have been demonstrated to provide a net economic benefit in some locations since the reductions in negative health consequences and health expenses surpass the price of implementing the interventions.

• **Design:**

The classification model was analyzed, drinking Safe drinking water In this project, the AKN used historical samples from various sources and constrained water quality standards based on the chemical, and biological specifications of the sample that was determined to be potable or non-potable.

• **Data:**

The data set contains 8000 rows 22 columns. The data were selected from the Kagel website and the data were Data released in January 2021
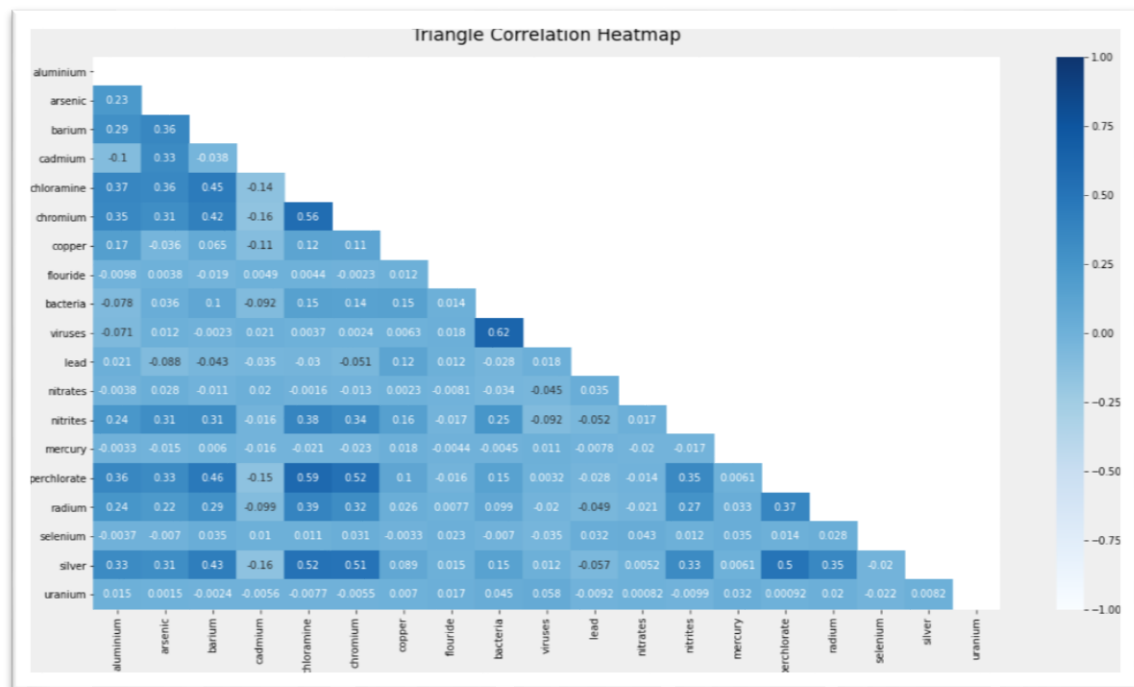
• **Algorithms:**

- Gradient Boosting Classifier
- Extreme Gradient Boosting
- Random Forest Classifier
- CatBoost Classifier
- Decision Tree Classifier
- Light Gradient Boosting Machine
- Extra Trees Classifier
- Quadratic Discriminant Analysis
- Naive Bayes K Neighbors Classifier
- Logistic Regression
- Ridge Classifier Linear
- Discriminant Analysis
- SVM - Linear Kernel

Tools:

- Data manipulation: NumPy and Pandas
- Modeling: Sklearn
- Visualization: matlibplot and seaborn
- Others:
- NullFormatter, Plotly, Missingno, Filterwarnings, Counter. Orange

• **Communication:**

This figure shows the amount of potable water (0), and non-potable water (1)

Potability Feature

89.0%

11.0%

Not safe for Human consumption

Safe for Human consumption

is_safe

Recall of The Data

Recall

Log_Reg   KNN   Naive-B   D-Tree   SVM

Models