**FINAL ASSIGNMENT**

**Dataset: Wastewater Treatment Plant Dataset**

The dataset contains daily measurements from a full-scale wastewater treatment plant, including various physicochemical properties that help assess plant performance. The dataset can be used to analyse the plant's operational effectiveness and predict potential faults.

**Dataset Link** : [**https://www.kaggle.com/datasets/d4rklucif3r/full-scale-waste-water-treatment-plant-data**](https://www.kaggle.com/datasets/d4rklucif3r/full-scale-waste-water-treatment-plant-data)

The objective is to build a machine learning model that predicts whether the wastewater treatment process is operating optimally based on daily measurements of the plant's operational data.

**Task :**

1. **Data Preprocessing:**

Perform comprehensive data preprocessing, including handling missing values, scaling, and feature selection, while visualizing key trends and correlations to enhance data insights and model performance.

1. **Modeling with Machine Learning Algorithms:**

Apply various machine learning algorithms to classify the operational state of the wastewater treatment process. Use the following algorithms:

* + Logistic Regression from scratch as well as from sklearn
  + K-Nearest Neighbours (KNN)
  + Decision Tree Classifier
  + Random Forest Classifier
  + Support Vector Machine (SVM)
  + Any other relevant algorithms you feel might improve performance.

1. **Model Evaluation:**

Evaluate and compare models using accuracy, F1-score, and confusion matrix, and perform hyperparameter tuning to optimize performance on the test dataset.

**Note :** With the given link for dataset, it will redirect you to Kaggle website from there download this dataset. It is a zip file containing two datasets you must use Data-Melbourne\_F\_fixed csv file and ignore the other csv file and make a little change in first row it is starting with comma so before that you must write serial No then save changes and use the updated dataset.