

# Project-1: Self-Kaggle dataset

**Kaggle dataset:**

<https://www.kaggle.com/datasets/cpluzshrijayan/milkquality>

- **Data Preprocessing and Model Building:**

- **Data Preparation:**

- Conversion of categorical 'Grade' values ('low', 'medium', 'high') to numerical values (0, 1, 2) using mapping.
    - Splitting the dataset into features (X) and the target variable (y).
    - Further splitting the data into training and testing sets using `train_test_split`.

- **Imbalanced Data Handling:**

- Employing Synthetic Minority Over-sampling Technique (SMOTE) to balance the imbalanced data in the training set.

- **Standardization:**

- Standardizing the numerical data in the training and testing sets using `StandardScaler`.

- **Hyperparameter Tuning:**

- Utilizing Keras Tuner to search and find the best hyperparameters for the neural network model.
    - Defining a function to build the model based on the hyperparameters specified.
    - Performing a random search for the best hyperparameters for the model.

- **Summary of Results:**

- Displaying the search space and the best trial results with different hyperparameter configurations.

- **Model Building and Training:**

- **Model Architecture:**

- Constructing the neural network model based on the best hyperparameters from the tuner.

- **Model Training:**

- Training the model on the standardized training data (X\_train, y\_train) for 10 epochs.
    - Displaying a plot of the training loss and validation loss across epochs.

- **Model Evaluation:**

- Making predictions on the test data and calculating evaluation metrics like accuracy, precision, recall, and F1 score.
  - Displaying the confusion matrix to visualize the model's performance in classifying the test data.

- Here are the performance metrics achieved by the model:

- Accuracy: 91.98%
  - Precision: 93.07%
  - Recall: 91.98%
  - F1 Score: 92.22%