# ***Machine Learning Project***

**Image Dataset Documentation**

|  |  |
| --- | --- |
| **ID** | **Name** |
| **20210065** | **Ahmed Tarek Abdelwali Ahmed** |
| **20210080** | **Ahmed Alaa Abassi** |
| **20210058** | **Ahmed Said Mohamed Said** |
| **20210044** | **Ahmed Khaled Amin Kadi** |
| **20210119** | **Ahmed Medhat Abobakr** |
| **20210590** | **Omar Ahmed Elsaid Ahmed** |

# **Numerical dataset :**

General information about dataset

|  |  |
| --- | --- |
| Name | Food101 |
| No. of classes | 5 |
| Total no. of samples | 1000 / class |
| No. of samples in training\validation | 500 / 100 |
| No. of samples in testing | 300 |

**Data Loading and Preprocessing :**

This function loads and preprocesses image data for the Food101 dataset. It reads images, converts them to grayscale, resizes them to 64x64 pixels, and extracts Histogram of Oriented Gradients (HOG) features.

A screenshot of a computer code

Description automatically generated

**Scaling :**

Standard scaling is applied to the training, validation, and test data using StandardScaler. This step ensures that the features have a mean of 0 and a standard deviation of 1, which is essential for certain machine learning algorithms.

A screenshot of a computer code

Description automatically generated

**Implementation Details :**

This section includes details about the dataset, such as the dataset name, number of classes, class labels, total samples, and sample size. Additionally, information about the feature extraction phase is provided, including the number of features extracted which is = 4096 , feature names, and the dimension of the resulted features.

A screenshot of a computer code

Description automatically generated

**Logistic Regression Model Training**

Logistic Regression model is trained on the preprocessed training data. Hyperparameters such as learning rate, optimizer, regularization, batch size, and number of epochs are defined. The model is then evaluated on the validation set, and accuracy is printed.

A screenshot of a computer program

Description automatically generated

**K-Means Clustering**

K-means clustering is applied to the training data using KMeans with the number of clusters set to the number of classes. The model is then used to predict cluster assignments for the validation set.

A screenshot of a computer code

Description automatically generated

**Clustered Logistic Regression Model Training :**

A logistic regression model is trained on the cluster assignments obtained from K-means. The combined model (clustering + logistic regression) is evaluated on the validation set, and accuracy is printed.

A screenshot of a computer program

Description automatically generated

**Model Evaluation on Test Set :**

The combined model is evaluated on the test set, and accuracy is printed. Additionally, confusion matrix for the test set is printed.

A screenshot of a computer

Description automatically generated