

# [1]UTKFace dataset [dataset link](#)

## About dataset :

UTKFace dataset is a large-scale face dataset with long age span (range from 0 to 116 years old). The dataset consists of over 20,000 face images with annotations of age, gender, and ethnicity.

The images cover large variation in pose, facial expression, illumination, occlusion, resolution, etc.

This dataset could be used on a variety of tasks, e.g., face detection, age estimation, age progression/regression, landmark localization, etc.

## Labels:

The labels of each face image is embedded in the file name, formatted like [age]\_[gender]\_[race]\_[date&time].jpg

**[age]** is an integer from 0 to 116, indicating the age.

**[gender]** is either 0 (male) or 1 (female).

**[race]** is an integer from 0 to 4, denoting White, Black, Asian, Indian, and Others (like Hispanic, Latino, Middle Eastern).

**[date&time]** is in the format of yyyyymmddHHMMSSFFF, showing the date and time an image was collected to UTKFace.

**total number of samples total number of samples:**

**part1 = 10137 samples**

**part2 = 10719 samples**

**part3 = 3252 samples**

**we used part 3**

**train and test size:**

**training : 2601**

**testing : 651**

**Implementation details :**

We used two algorithms K-Means and Logistic Regression

**Pre-Processing :**

Extracts features by HOG algorithm , And reading age, gender, and race in lists, Compile a list of age, gender, race and photo features in Dataset.

We got 2190 features per each image , but we sum all this features in one new feature called features

Save this dataset into file .csv

Then normalize dataset by Min-Max Normalization algorithm.

Split data into features and target

**features** : age , race and image

**target** : gender

And split dataset into 20% test and 80% train by train\_test\_split algorithm.

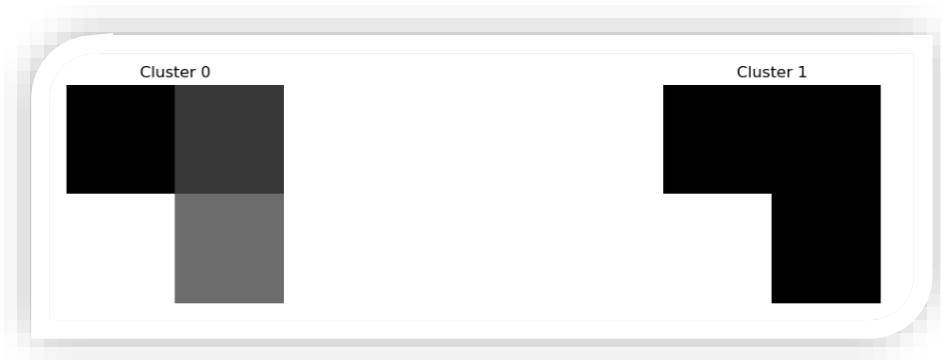
**A) K-Means :**

**Goal** : Classify humans into Male and Female (0 or 1)

**Explain Code :** select K randomly and loop from 2 to 11 , in each iteration save inertia in list after end loop compare all inertias and plot **Elbow** to select right K.

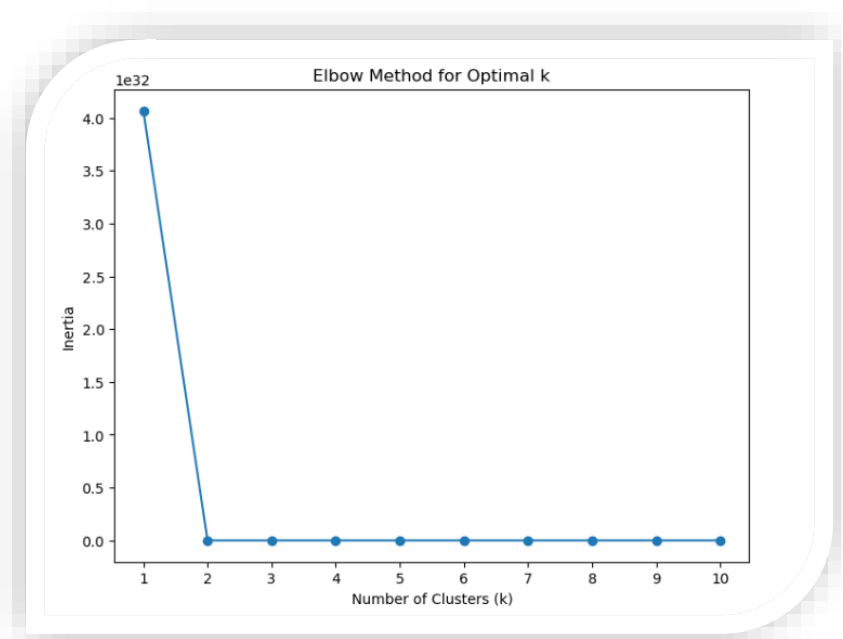
## Result:

- Not suppose Cross validation
- The Inertia: 15865.289326214694



- The silhouette score is: 0.9999067077152719
- Visualize the cluster :

## B) Logistic Regression :



**Goal :** Detect Gender

**Explain Code:**

-We don't need to sum image features , we used dataset with all image features

-Cross Validation :

"k-fold cross-validation" with k=5.

**Result :**

**Cross-Validation Scores:**

[0.74088292 0.77884615 0.73461538 0.76923077 0.73461538]

**Mean CV Accuracy:** 0.7516381219548206

**(Cross-Validation) Train Score :** 0.8796616685890042

**(Cross-Validation) Test Score :** 0.7741935483870968

**Train Score :** 0.8796616685890042

**Test Score :** 0.7741935483870968

**Classes is :** [0 1]

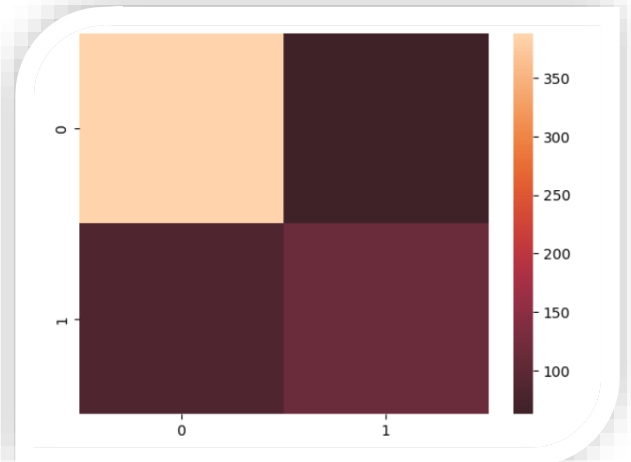
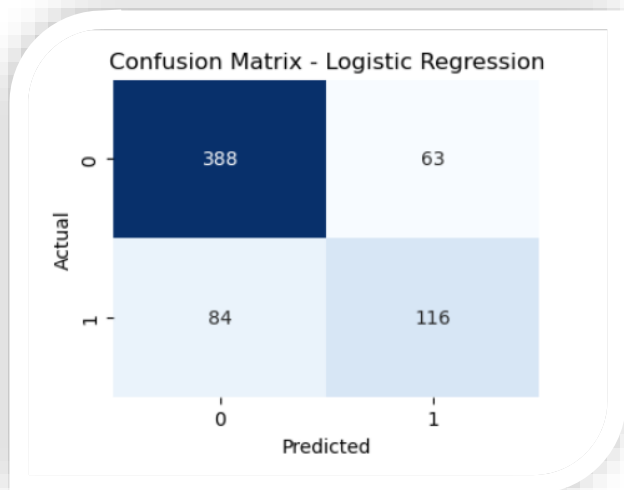
**Iterations is :** [100]

**Intercept is :** [-4.80416957]

**Confusion Matrix :**

**[[388 63]**

**[ 84 116]]**



**Accuracy Score** : 504

**F1 Score is** : 0.7741935483870968

**Recall Score is** : 0.7741935483870968

**Precision Score is** : 0.7741935483870968

**Precision Recall Score is** : (0.7741935483870968, 0.7741935483870968, 0.7741935483870968, None)

**Precision Value is** : [0.30721966 0.64804469 1. ]

**Recall Value is** : [1. 0.58 0. ]

**Thresholds Value is** : [0 1]

**AUC Value** : 0.7201552106430156

**fpr Value** : [0. 0.13968958 1. ]

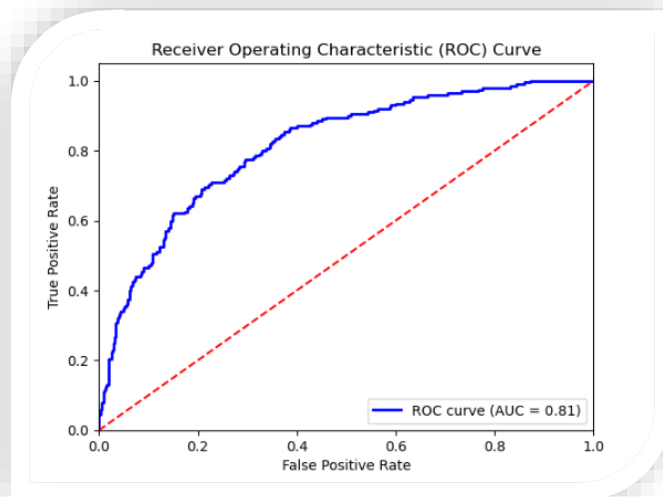
**tpr Value** : [0. 0.58 1. ]

**thresholds Value** : [inf 1. 0.]

**ROCAUC Score** : 0.7201552106430156

**Zero One Loss Value** : 147

**RCO curve** :



## Team members

Students of the [Faculty of Computers and Artificial Intelligence](#) at [Helwan University](#), Department of CS & AI

Course : Machin Learning (fall 2023)

Name	ID	Dept.
Ossama Samir	20210140	CS
Abd-ulla Mohamed	20210556	CS
Marowa Omar	20210900	CS
Rana Essam	20210335	CS
Toka Mohamed	20210242	CS
Ahmed Mohamed	20210102	AI
Omar Nasser	20210628	AI

## References :

[Kaggle: Your Home for Data Science](#)

[Machine Learning Tutorial \(geeksforgeeks.org\)](#)

[What is Machine Learning? | Google for Developers](#)

[Python Machine Learning \(w3schools.com\)](#)

[Supervised Machine Learning: Regression and Classification | Coursera](#)

[\(478\) 10 مكتبة سايكتايرن : القسم العاشر : Sklearn Library - YouTube](#)

[Natural Language Processing With Python's NLTK Package – Real Python](#)

[\(478\) Machine Learning Algorithms بالعربي - YouTube](#)

[\(478\) Machine Learning New 2023 - YouTube](#)

[\(478\) Tabular Data Preprocessing For Machine Learning Models\(in Arabic\) - YouTube](#)