Faculty of computers and artificial intelligence

**Cover sheet**

**AI330 Machine Learning Project**

**Team no.:**

|  |  |
| --- | --- |
| Name | ID |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |
|  |  |

# Numerical dataset

General information about dataset

|  |  |
| --- | --- |
| Name | Salary by Job Title and Country |
| No. of classes | 8 |
| Total no. of samples | 6684 |
| No. of samples in training\validation | 4010 \ 1336 |
| No. of samples in testing | 1336 |

## Preprocessing phase:

We converted the strings values to numeric using get\_dummies function and scaled the data using standard scaler

We spited the data to train, cross validation and test (60,20,20)

We used the cross validation to decide the best Hyperparameters

Linear regression model:

We create polynomial features with degree =2

## We used Ridge model with alph =0.8

## training accuracy 0.94

## cross valedation accuracy 0.91

## test accuracy 0.91

## 

## Knn model:

We also created polynomial features with degree =2

## There is a plot showing model performance depending on the number of neighbors:

## 

## And we saw that the best number of neighbors =7

## The results:

## Accuracy on training = 0.935

## Accuracy on cross validation= 0.91

## Accuracy on training = 0.896

## 

# Image dataset

Implementation Details

|  |  |
| --- | --- |
| Name | Food Recognition |
| No. of classes / Their labels | 3 / 0,1,2 |
| Total no. of sample | 3000 |
| Size of image | 28x28 |
| No. of samples in training\validation | 1650 |
| No. of samples in testing | 1350 |
|  |  |

(How many features were extracted, their names, the dimension of resulted features)

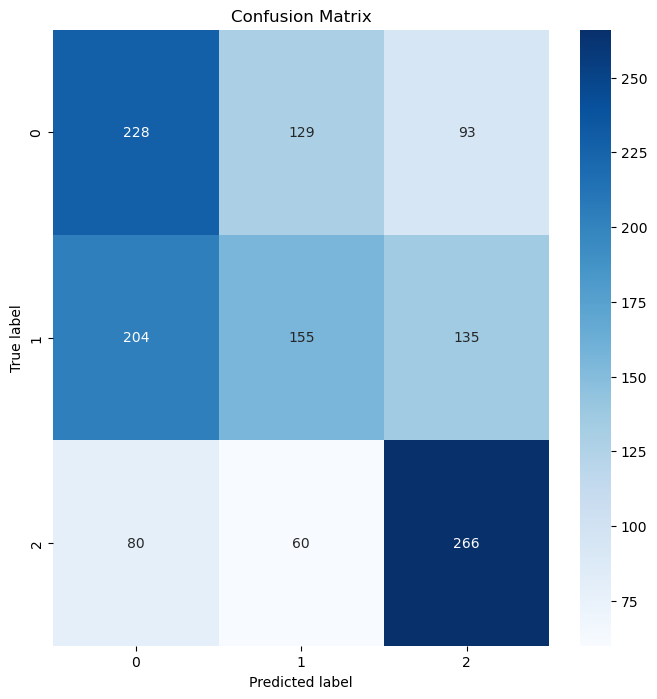
Number of extracted features: 1

Extracted feature: Mean Pixel Of RGB

Dimensions: (28x28)

Logistic Regression

* Confusion Matrix Before HOG:



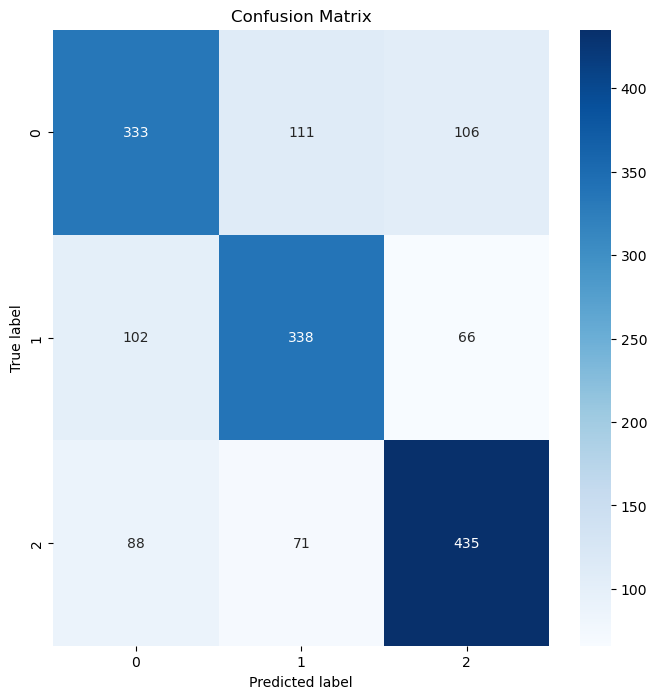
* ROC Before HOG:

A graph of a curve

Description automatically generated with medium confidence

Accuracy of logistic regression without HOG features: 0.48074074074074075

* Confusion Matrix After HOG:



* ROC After HOG:

A graph with different colored lines

Description automatically generated

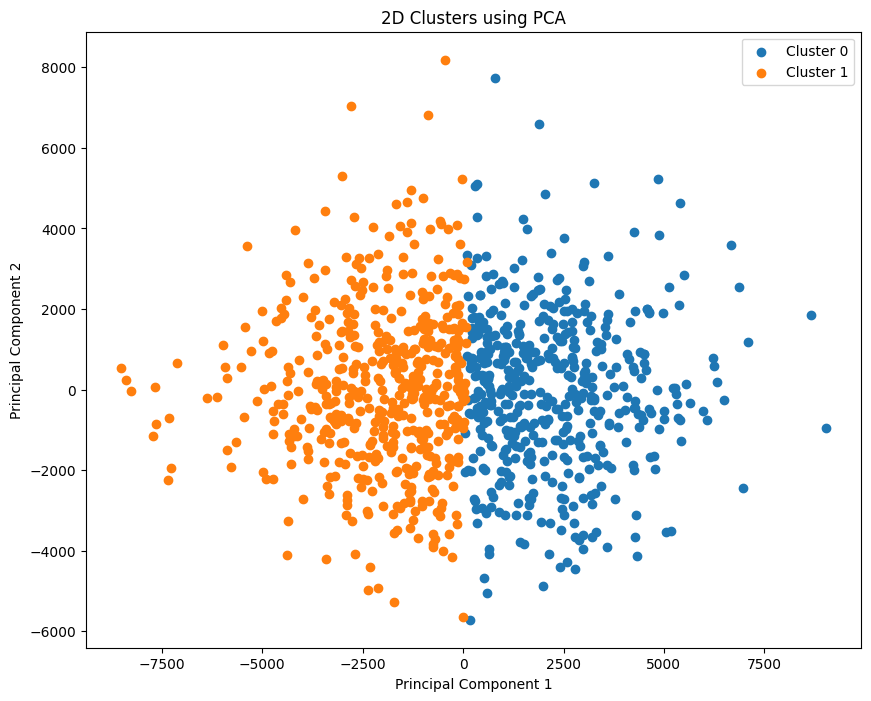
Accuracy of logistic regression with HOG features: 0.6703030303030303

K\_means

Sample from each cluster :(comparing predicted label with the actual label)

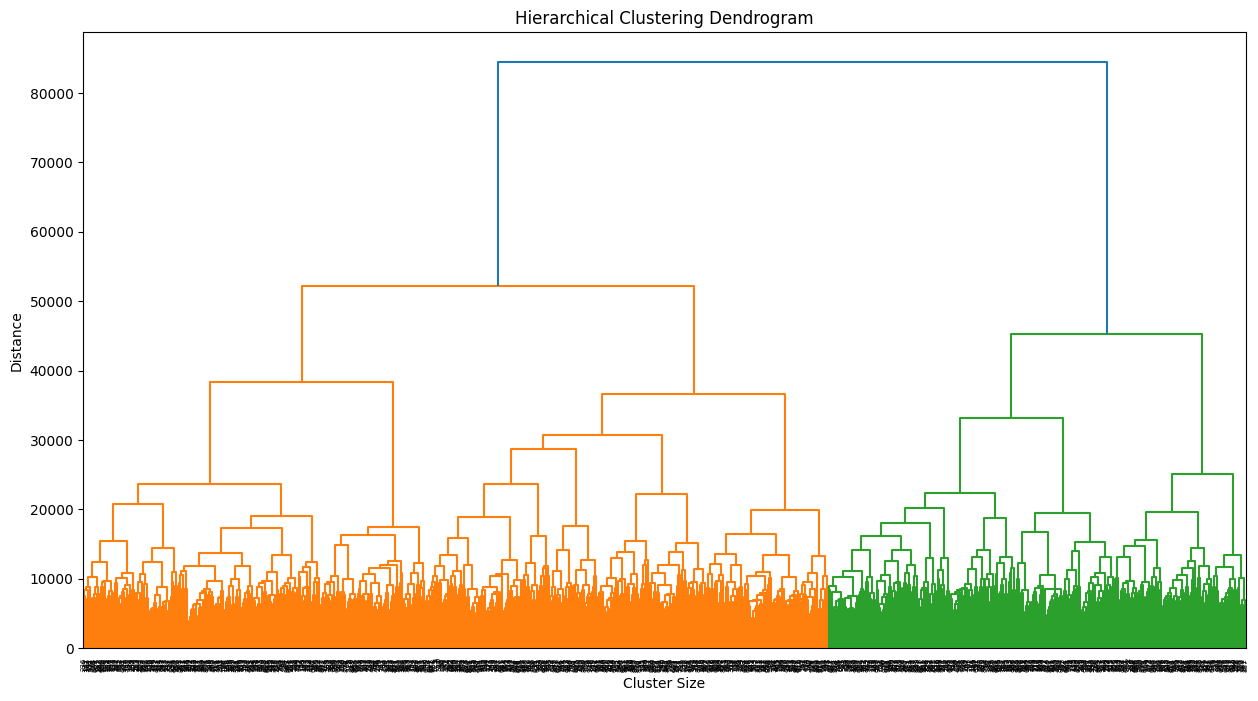


2D CLUSTERS USING PCA:



Accuracy: 50.10%

HIERARCHICAL CLUSTERING DENDROGRAM:



ELBOW :

