Title: Group 7_Face Verification

Group Members:

Narasimha Reddy Nalamalapu (11920025)
Srinath Veluri (11920004)
Vamshi Krishna B (11920065)
Satya Arun Manohar K (11920032)
Somanchi Sai Anirudh (11920042)

Abstract: Face verification can be defined as the task of comparing a person's face with another and verifying whether both images match and not.

Introduction and Approach:
We have performed face verification project as a part of Foundational Project – 2. We have decided to pick up this work to understand the significance, importance and its usage in Machine Learning.

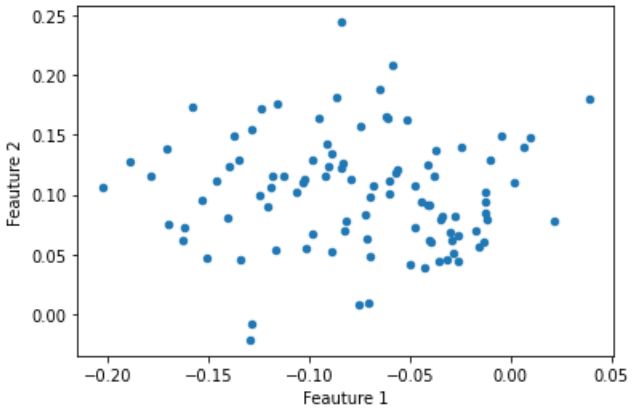
Step 1: Encoded the image data into feature space
Step 2: Performed EDA to understand the significance of data
Step 3: Identifying the right classifier for face verification

Problem Definition: Being able to match person's image with name

Dataset Description: The dataset consists of 9000+ images out of which about 25 images could not be encoded into feature space

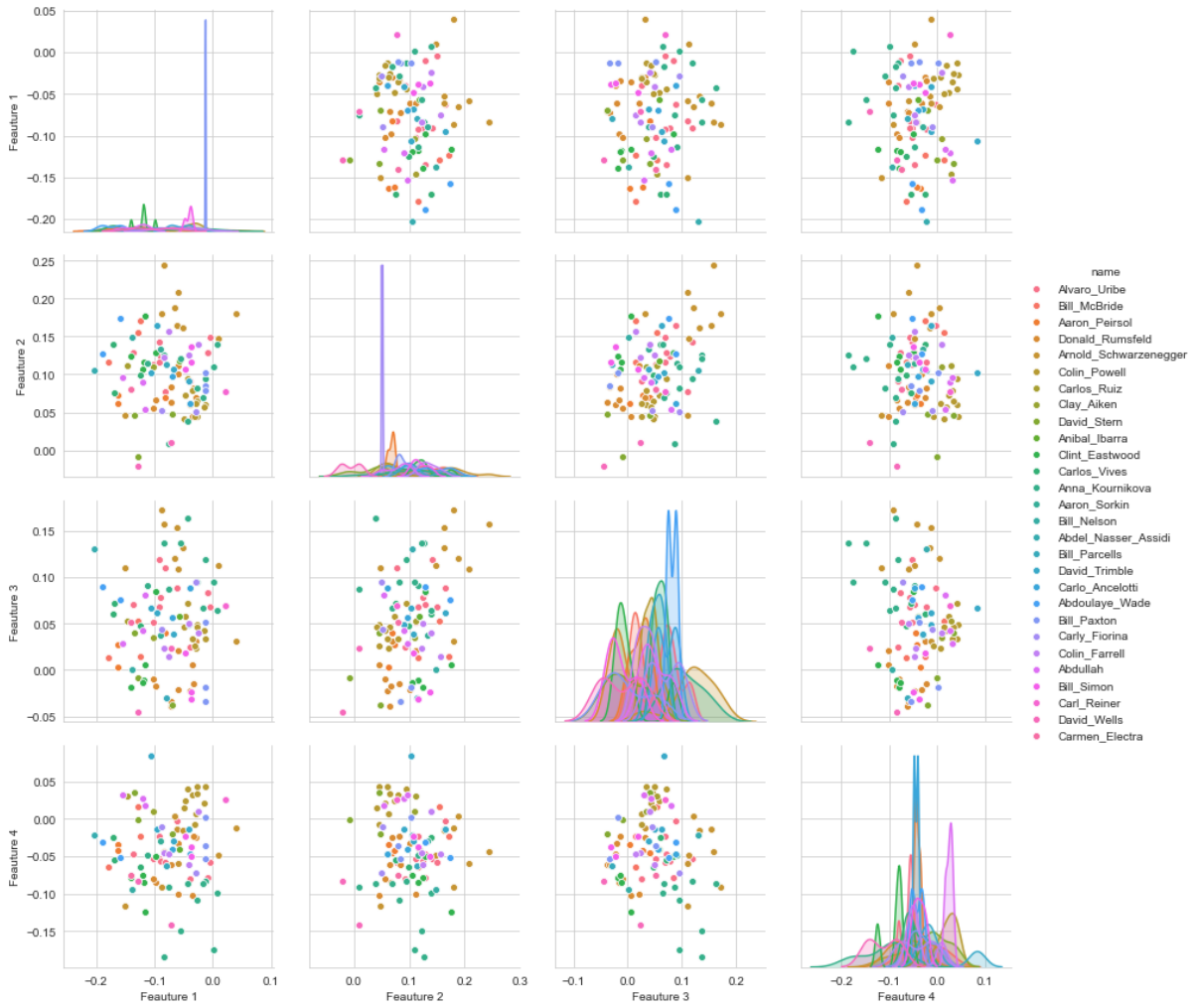
Exploratory Data Analysis: We performed EDA to highlight insights from the data.

Basic EDA:



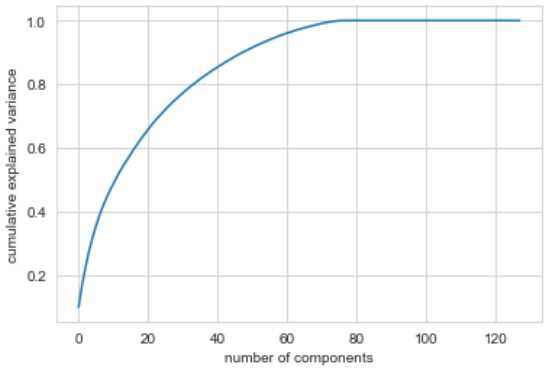
Scatterplot for feature 1 and 2 does not provide any immediate insight as grouping in the data cannot be seen.

Pair plot for 4 features on sample dataset can be seen below:



The above pair plot does not give much insight into the data. This is due to high number of labels present within the data.

EDA using PCA:



From the above graph, we can observe that the entire variance can be explained by using 70 principal components.

Using these 70 principal components – we formed 2 and 3 clusters with K means.

- With 2 clusters = We were able to segregate images into two folders (Majority females [1] and majority males [2])
- With 3 clusters = We were successful in segregating the images into three folders (Majority females [1], majority males [2], Asian/Brown/Black [3])

Experimental results:

Below table represents the respective accuracies with different clustering methods used in the model:

Model	Accuracy
Logistic Regression	5.33
Decision Tree Using Enropy	10.60
SVM using RBF	11.92
SVM Linear	14.55
KNN using Euclidean and 1 neighbor	92.10
KNN using Cosine and 1 neighbor	91.98
KNN using Manhattan and 1 neighbor	91.56
KNN using 2 neighbor	70.84

- We downloaded some images from Google for testing purposes
- All the persons available in the training dataset were identified correctly
- For some of the persons which are not present in the training dataset were shown as “Person not found”

{Matched} { Matched} {Matched} {No Match} {Wrong Match} {Matched} {Matched}

- However, some persons who were not present in the training dataset were still matched to the wrong persons
- This issue can be due to facial features matching/closely related to another person

Conclusion:

KNN-Classifier performed well on majority of new persons which were downloaded from Google