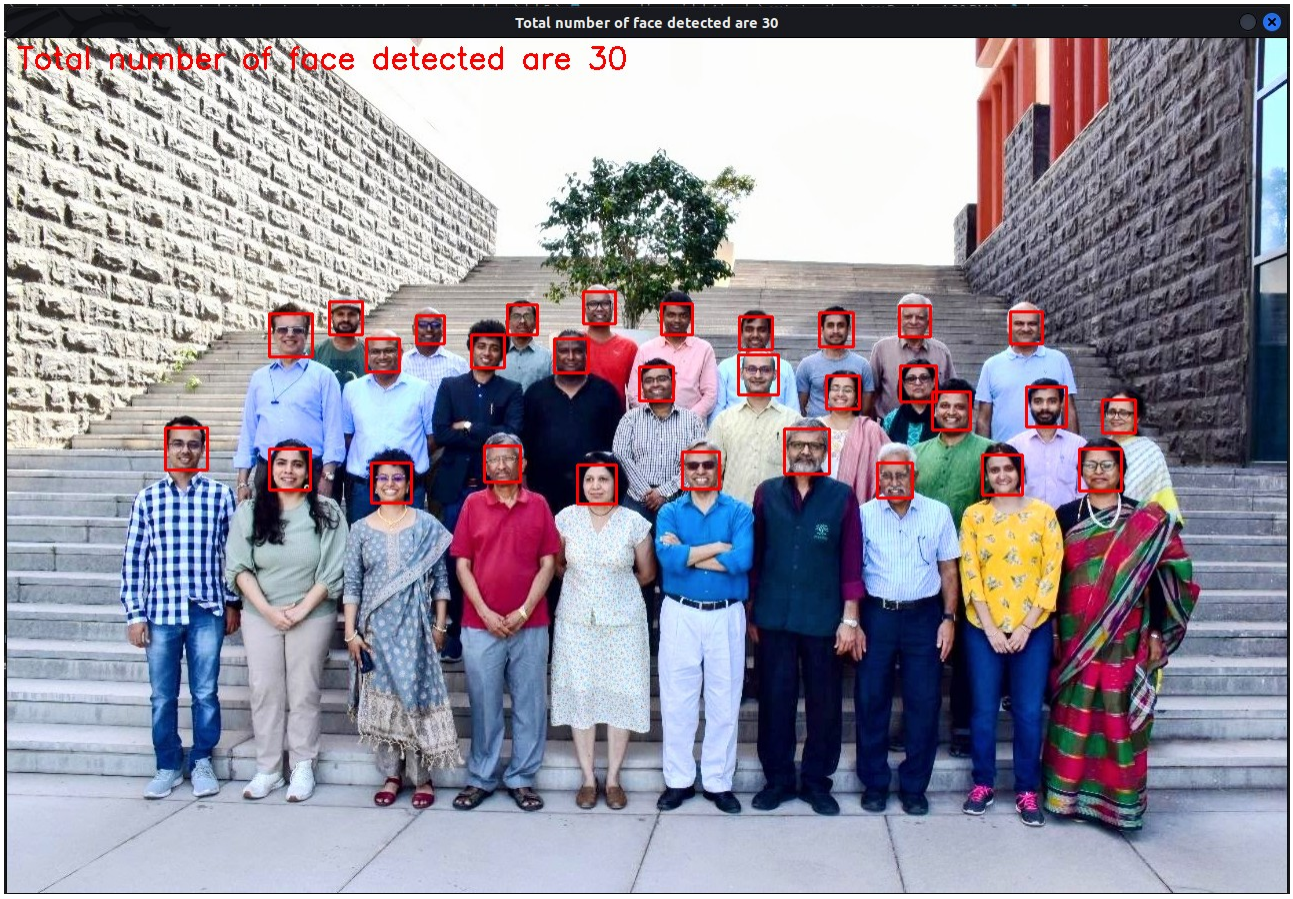
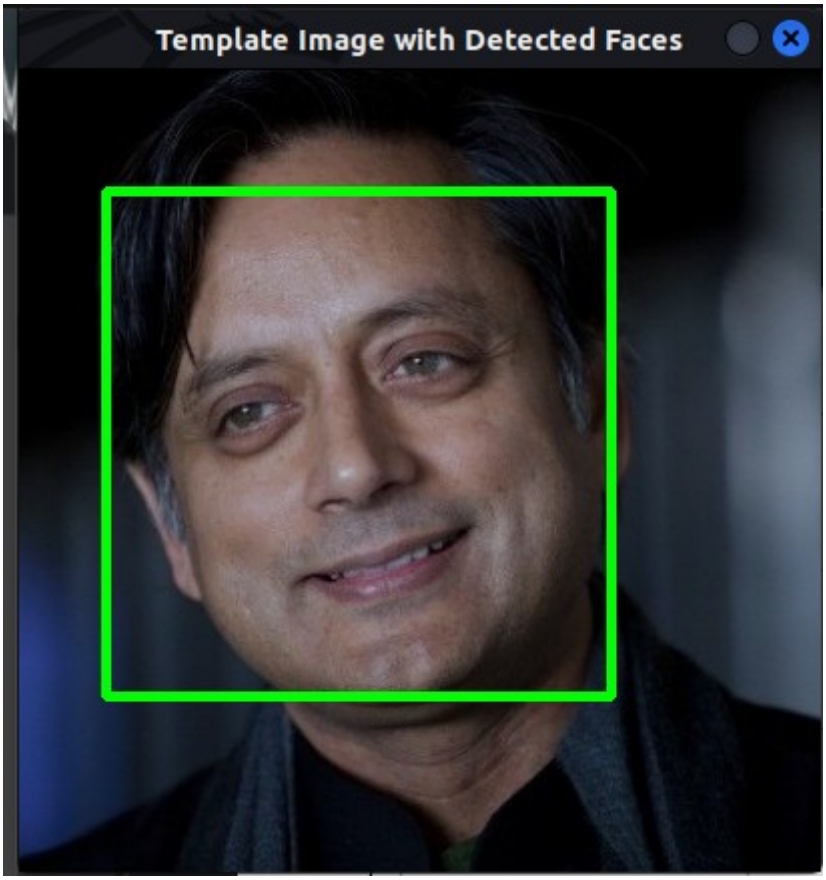
Name: Usman Akinyemi
U20220090





Questions Lab 5 MLPR

1. What are the common distance metrics used in distance-based classification algorithms?
Euclidean Distance, Mahalanobis Distance, Manhattan Distance, Chebyshev Distance, Minkowski Distance, Cosine Distance, Hamming Distance.

2. What are some real-world applications of distance-based classification algorithms?
Customer Behavior Prediction • Document Classification • Spam Filtering • Image Classification • Face Recognition • Product Categorization • Malware Categorization • Image Sentiment Analysis • Emotion Classification

3. Explain various distance metrics.
Euclidean distance: Measures the straight-line distance between two points.
Manhattan distance: Measures the sum of the absolute differences between the coordinates of two points.
Minkowski distance: Generalization of Euclidean and Manhattan distances, where the power parameter p determines the distance type.
Mahalanobis distance: Accounts for correlations between features and is sensitive to the scale of the data.

4. What is the role of cross validation in model performance?
It helps to estimate and improve the performance of a model during training. It also helps to reduce overfitting and underfitting.

5. Explain variance and bias in terms KNN?

Variance :- Whenever there is a smaller value of k which leads to a complex model that fits the taining data, we said there is a high variance in the model. This leads to overfitting of the models and high noise in the model.

Bias :- In the other hand, bias arises in KNN when the value of K is high leading to a simple model and underfitting though with less noise data points.