Muhammad Arslan Aslam

[Bsstats138.20@gmail.com](mailto:Bsstats138.20@gmail.com)

Roll no C1-23

Project Report:

I collect dataset from kaggle here is link <https://www.kaggle.com/datasets/ashwingupta3012/human-faces> . The dataset contain male and female images . firstly I distributed data into 2 sub folder one is male images and other is female images . Both have almost 3000 images .

**Gender Detection Using Machine Learning:**

The problem of gender detection using a combination of image processing techniques and machine learning algorithms. The goal is to accurately determine the gender of individuals from facial images. Image preprocessing, feature extraction using the classifier, and training various machine learning models to achieve accurate gender predictions.

**Data Preparation and Preprocessing**

The code begins by loading and preprocessing facial images from gender-specific folders using the classifier. The images are converted to grayscale and then subjected to face detection using the pre-trained model. Detected faces are resized to a uniform image dimension to ensure consistency across the dataset. Extracted face images are flattened to create feature vectors and paired with corresponding gender labels.

**Data Exploration and Visualization**

The script provides insights into the dataset by displaying a set of randomly selected male and female face images. This visualization offers a glimpse of the data being used for training and testing the models. It helps us understand the diversity and quality of the dataset.

**Model Selection and Evaluation**

The code introduces several machine learning classifiers for gender prediction, including Logistic Regression, Support Vector Machines (SVM), K-Nearest Neighbour (KNN), Random Forest. For each classifier, the script trains the model on the pre-processed data and evaluates its performance using accuracy. Additionally, the training time for each model is measured to assess computational efficiency.

**Results:**

**1 : Logistic Regression**

Training Time: 207.02057361602783 seconds

Accuracy: 0.9506802721088435

**2 : Support Vector Machine**

Training Time: 275.6901183128357 seconds

Accuracy: 0.9183673469387755

**3 : Random Forest Classifier**

Training Time: 62.46686577796936 seconds

Accuracy: 0.951530612244898

**4 : K-Nearest Neighbour (KNN)**

Training Time: 0.22419261932373047 seconds

Accuracy: 0.8469387755102041

**Conclusion :**

Based on the provided results, the Random Forest Classifier appears to be the best-performing model among the four. It achieved an accuracy of approximately 95.15%, which is quite high. Furthermore, it trained relatively quickly in about 62.47 seconds, striking a balance between accuracy and training time. While both Logistic Regression and Support Vector Machine also demonstrated strong performance with accuracies of around 95.07% and 91.83%, respectively, the Random Forest Classifier outperformed them in terms of accuracy and training time, making it the preferred choice for this task.