Age, Gender and Ethnicity Detection

K. Jayakar Sai Sushanth Reddy (18CS01004)

K. Kishorereddy (18CS01007)

# Objectives of the Project:

Analyze the Dataset and build Machine Learning models to detect age, gender and ethnicity of a given facial image using Classification techniques. Evaluate and compare the accuracies of the models used and arrive at a conclusion.

# Data Source:

The dataset is obtained from Kaggle at the following link <https://www.kaggle.com/nipunarora8/age-gender-and-ethnicity-face-data-csv>. It includes a CSV of facial images that are labelled on the basis of age, gender, and ethnicity. The dataset includes 27305 rows and 5 columns.

# Attributes:

|  |  |  |  |
| --- | --- | --- | --- |
| Attribute Name | Description | Attribute Type | Attribute Values |
| age | Age of the person in the image | Continuous, Integer | Range (1, 116) |
| gender | Ethnicity of the person | Categorical, Integer | {0, 1} |
| ethnicity | Gender of the person | Categorical, Integer | {0, 1, 2, 3, 4} |
| img\_name | Name of the image | String | \*.jpg |
| pixels | Array to String of the image pixels | String | String of 2304 space separated pixel values |

# Data Pre-Processing:

* The img\_name attribute was dropped since it was simply the file names of the images used when the data set created and does not signify as a training attribute.
* For the target variables age, gender and ethnicity, the corresponding columns in the dataset were selected.
* For the input variables, the strings in ‘pixels’ column of the dataset were split into 2304 columns of pixel values, each row representing an image of 48\*48 pixels.

## Cleaning:

Chosen Dataset was already clean and no further operations were required.

## Data Transformation:

Z-Score Normalization was applied on the 2304 columns of pixel values.

## Data Reduction:

Principal Component Analysis (PCA) was applied replacing 2304 input variables by 37 PCs which explained 85% of the variance.

# Methods:

Following methods were used in building models:

* Gaussian Naïve Bayes:
* Logistic Regression:
* Convolutional Neural Networks:

# Results and Analysis:

# Appendix: