Human Gender Classification Using Machine Learning

Taghreed Alamri

Date

10-10-2021

Course title

Data Science

Introduction

Gender classification has become an essential task in human computer interaction (HCI). It is used in immense number of applications like passive surveillance, control in smart buildings (restricting access to certain areas based on gender) and supermarkets, gender advertising and security investigation.
Gender classification is to determine a person's gender, e.g., male or female, based on his or her biometric cues. Usually, facial features are extracted and then a classifier is applied to the extracted features to learn a gender recognizer. It is an active research topic in Computer Vision and Biometrics fields. The gender classification result is often a binary value, e.g., 1 or 0, representing either male or female. It is essentially a two-class classification problem. Although other biometric traits could also be used for gender classification, such as gait, face-based approaches are still the most popular for gender discrimination.

Objective

The goal of this project is to use a machine learning model to classify whether a person is male or female considering the facial features (such as nose width, Forehead length, etc.) of that person. Then apply other machine learning models and compare their performance.

Data

The data I will use is the Gender Classification dataset from Kaggle. Although the dataset is made-up, it is created considering real scenarios. It has 5001 samples that consist of 8 columns (7 features and 1 label/target column).

long_hair	Indicates whether this person has a long hair (1) or not (0).
forehead_width_cm	Width of the forehead from right to left given in cm.
forehead_height_cm	Height of the forehead in cm from where the hair grows to the eyebrows.
nose_wide	Whether the nose is wide or not. 1 represents wide and 0 not.
nose_long	Whether the nose is long or not. 1 represents long and 0 not.
lips_thin	Whether this person has a thin lip or not. 1 represents thin and 0 not.
distance_nose_to_lip_long	Is the distance from nose to lip is long? 1 represents yes and 0 not.
Gender	Either Meal or Female

Gender is the target column with 2 classes Male and Female.

Tools

The main tools I will use in addition to Python and Jupyter Notebook are:

- Numpy and Pandas for data manipulation.
- Scikit-learn for modeling.
- Matplotlib and Seaborn for plotting and visualization.

References

1- Guo G. (2015) Gender Classification. In: Li S.Z., Jain A.K. (eds) Encyclopedia of Biometrics. Springer, Boston, MA. https://doi.org/10.1007/978-1-4899-7488-4_9176

2-Kalam, Swathi & Guttikonda, Geetha. (2013). Gender Classification using Geometric Facial Features. International Journal of Computer Applications. 85. 10.5120/14855-3222.

3- https://www.kaggle.com/ortalcheriker/gender-classification-ai-project