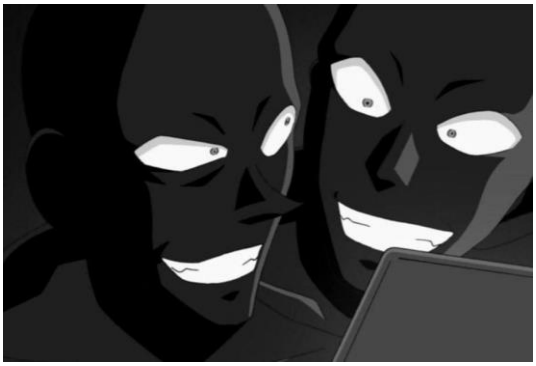
Gender Classification

By

Mashari Almuhanna

Sohaib Albakri



Data Science Bootcamp

SDAIA Academy

Riyadh – Saudi Arabia

[October 2021]

Problem

Sometimes the police have the witness description of the criminal but they do not know whether he/she is a male or female so they cannot reduce the search area.

Solution

In this project, we will develop a classification model to predict the gender of the criminal according to the description we got from the witness. Which will help the police to search for the criminal.

Dataset

The dataset we will use contains the witness description, The data are taken from the Gender Classification Dataset from Kaggle The following table explains the dataset in detail:

Column	Data type
long_hair	Integer (0 or 1), indicates whether this person has a long hair or not.
forehead_width	Integer, width of the forehead from right to left given in cm.
forehead_height	Integer, width of the forehead width in cm
nose_wide	Integer (0 or 1), whether the nose is wide or not
nose_long	Integer (0 or 1), whether the nose is long or not
lips_thin	Integer (0 or 1), whether this person has a thin lip or not
distance_nose_to_lip_long	Integer (0 or 1), is the distance from nose to lip is long
gender	String, either Male or Female

Features

We will use five models to find the best model Logistic Regression Model, which is the best for binary classification, Decision Tree, Support Vector, Random Forest Classifier, KNN.

Tools

To grab the data and build the model we will use the following tools:

Tool	Description
Jupyter notebook	Contains cells of Python code and human-readable text
pandas	Library is written in Python for data manipulation and analysis
sklearn	Software machine learning library for the Python programming language
Matplotlib	Matplotlib is a plotting library for Python

Conclusion

We expect the model will predict the gender of criminal after receiving the description from the witness. The model will help the police to reduce the search area.