User Profiling on social media

*** tcss555**: It is a script to run the prediction algorithm and save the results in output file. This file consists of the following command:

python3.6 "ensemble.py" $INPUTPATH $OUTPUTPATH

*** public-test-data:** It is the folder where the test data is present. It is the “INPUTPATH”

*** output:** It is the folder where the xml files of all the users are stored after predicting their age, gender and personality. It is the “OUTPUTPATH”

*** Command to execute our code**:

tcss555 -i /data/public-test-data -o ~/output/

*** ensemble.py:** Is the main python file which reads the training and test data and refers different files to run prediction algorithm. It also saves the output

*** testAccuracy.py:** Similar to ensemble.py, this file gets the input, splits it into test data and train data and calls the other algorithms to get the prediction. But instead of saving the output, it calculates the accuracy of the result. It is convenient to use this file to calculate accuracy rather than modifying ensemble.py.

***Files related to text:**

*** Gender_By_Status.py:** This file predicts the gender of the user using the status messages. It uses Stochastic Gradient Descent classifier.

***Age_By_Status.py:** This file predicts the age of the user using the status messages. It uses Stochastic Gradient Descent classifier.

***Emotions_By_Status.py:** This file predicts the emotions of users based on the LIWC Features. It uses Linear Regression model.

***Files related to Likes:**

*** Gender_By_Likes.py:** This file predicts the gender of the users using the likes of the user.

*** Age_by_Likes.py:** This file predicts the age of users based on the likes

*** personalities_by_Likes.py:** This file predicts the emotions of users based on the likes

***Files related to Images:**

*** Gender_By_Image.py:** This file predicts the gender of the users using the images of the user present in the public-test-data folder. It uses Convolutional Neural network (CNN) with Keras

*** trained_cnn_image_classifier.h5:** This file consists of the weights obtained after training the model using CNN

*** cnn_image_data_processor.py:** This file consists of code to organize the data in required format. It is imported by cnn_image_model_generator.py.

*** cnn_image_model_generator.py:** This file consists of the code which is used to train the model using training data to generate a file which consists of weights (trained_cnn_image_classifier.h5). This file is later used by Gender_By_Image.py for predicting the gender.