**How-To document for the Project**

Pre-Requisites and know-how for running the model🡪

**Dataset**

1. The dataset (Training, Validation and Test) has to be downloaded from the web link below, follow the subsequent path from the link: “data🡪Training/Validation/Test Data🡪files” to obtain the zip files (sign up required to begin download)

<http://chalearnlap.cvc.uab.es/dataset/20/description>

password for Test dataset unzipping 🡪 “.chalearnLAPFirstImpressionsSECONDRoundICPRWorkshop2016.”

1. The ground truth for Test dataset is not available on the webpage with the Test dataset. Upon consulting the ChaLearn contact personal (Sergio Escalera, Julio Jacques Jr.) the test dataset ground truth links were shared via email, link attached below:  
    <http://datasets.cvc.uab.es/chalearn/FirstImpressionsV2/test-annotation-e.zip>   
   password for unzipping ground truth of test dataset🡪

“zeAzLQN7DnSIexQukc9W"

1. Perform one level of manual extraction of the downloaded dataset zip files by using a suitable tool namely (WinZip or equivalent Linux command)
2. Place all the extracted files (zip files) in a folder named “data”. This folder has to be created in the same directory where the main IPYNB file is placed
3. The other levels of extraction are done in the first three blocks of the IPYNB file
4. The ground truth files of Training and Validation dataset upon unzipping manually are obtained in the form of *csv* files
5. For the Test dataset, it is a *pkl* file which is converted into a *txt* file by a python script (*Test\_gt\_extraction.py*) attached in GitHub repository
6. Run this file from the point in memory where it is located with the command “python *Test\_gt\_extraction.py”* in the command line
7. The script outputs five *txt* files for each personality trait with ground truths which is converted into a *csv* file (similar to the format of Training and Validation ground truth files) manually
8. Place all the three ground truth files as training\_gt.csv, validation\_gt.csv and test\_gt.csv respectively in the created “data” folder

The three ground truth csv files are upload in the GitHub repository

Dataset part complete, rest is handled in the code

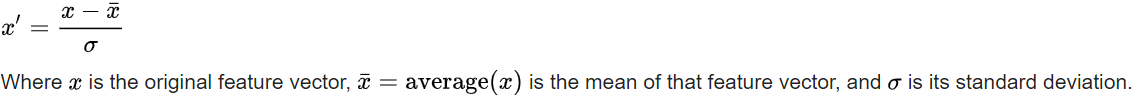
**General Instructions for running the Facial Feature model**

1. The outfile\_face (obtained after running “Training for the Facial features part of the model” block) has the output accuracy considering only the facial features as modality and considering the individual frames of the videos
2. The outfile\_ensemble (obtained after running all the blocks of the IPYNB file) has the final accuracy considering facial features only (after averaging the values obtained for all images per video) and accuracy value for each personality trait considering only facial features. It also has the accuracy values considering the ensembled result – the overall and per trait accuracy.

**OpenFace**

1. This section is used for extracting the Gaze features from all the videos using OpenFace tool
2. First Download the OpenFace binaries from the following link 🡪  
   <https://github.com/TadasBaltrusaitis/OpenFace/releases/download/OpenFace_2.2.0/OpenFace_v2.2.0_win_x64.zip>
3. The above link was used by us (64-bit CPU, Windows OS) for feature extraction with the commands as given in the IPYNB file
4. Please extract all training (6000 videos), validation (2000 videos) and test (2000 videos) videos manually into separate folders and give the locations of these folders in the “Gaze features extraction” block in the IPYNB file for successful feature extraction (Note: Location for OpenFace🡪 FeatureExtraction.exe also should be changed in IPYNB file)
5. Once the manual extraction of all videos is completed and file paths are changed, the features can be extracted by running the commands of the “Gaze features extraction” block in IPYNB
6. For every video, one csv file with gaze features will be generate. All csv files will be located inside a folder named “processed”, created in the current directory where the main IPYNB is located

**General Instructions for running the Gaze model**

1. We created a folder structure as shown below to store the gaze feature csv files in the location of our main IPYNB file   
   folder PPT🡪folder gazeData🡪folders test\_csv/training\_csv/validation\_csv🡪 place all the individual csv files in appropriate folders
2. If changes to above folder structure is made, please make note to update the corresponding folder paths in the “Consolidating the dataset to work” block in the IPYNB file
3. After running the “Consolidating the dataset to work” block we obtain 3 csv files namely training\_gaze\_features\_6k\_15.csv, validation\_gaze\_features\_2k\_15.csv and test\_gaze\_features\_2k\_15.csv which contain one row of gaze features per video (only Right eye, 15 features including ground truth) of their respective datasets
4. In the next step we perform manual normalization on each of the three files as per the formula given below to obtain training\_gaze\_features\_6k\_1\_N.csv, validation\_gaze\_features\_1\_N.csv, test\_gaze\_features\_1\_N.csv  
     
   
5. The manual normalization is performed for all the features vectors of all videos in the three files
6. Basic Excel functions were used to perform these normalizations easily
7. An alternative method to perform normalization is by using the below function from sklearn package “sklearn.preprocessing.StandardScaler”
8. After running all the blocks in the IPYNB file, the outfile\_gaze contains gaze accuracy as well as each personality trait prediction accuracy considering only gaze modality

All the csv files generated by OpenFace and created manually (after normalization) are upload in the GitHub repository

**General Instructions for running the Ensembled model**

1. After running all the blocks in the IPYNB file, the outfile\_ensemble contains ensemble accuracy as well as each personality trait prediction accuracy considering the ensembled model
2. Once the “P- value for statistical significance” block is run , the outfile\_pvalue contains the per video/per trait chi square values. All the 5 values (5 traits / video) are added to obtain a single value per video and the p-value per video is computed manually using the look up table as mentioned in the report