Description of the AI techniques used for the Estimation of Emotion

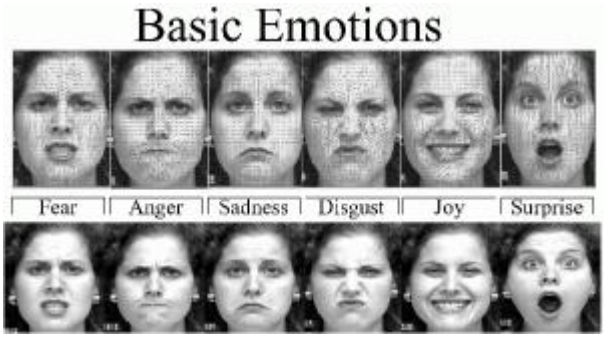
There are two main mechanisms by which emotion estimation is done:

- Emotion estimation from Face Analytics
- Emotion estimation (Valence) from Full Body Posture.

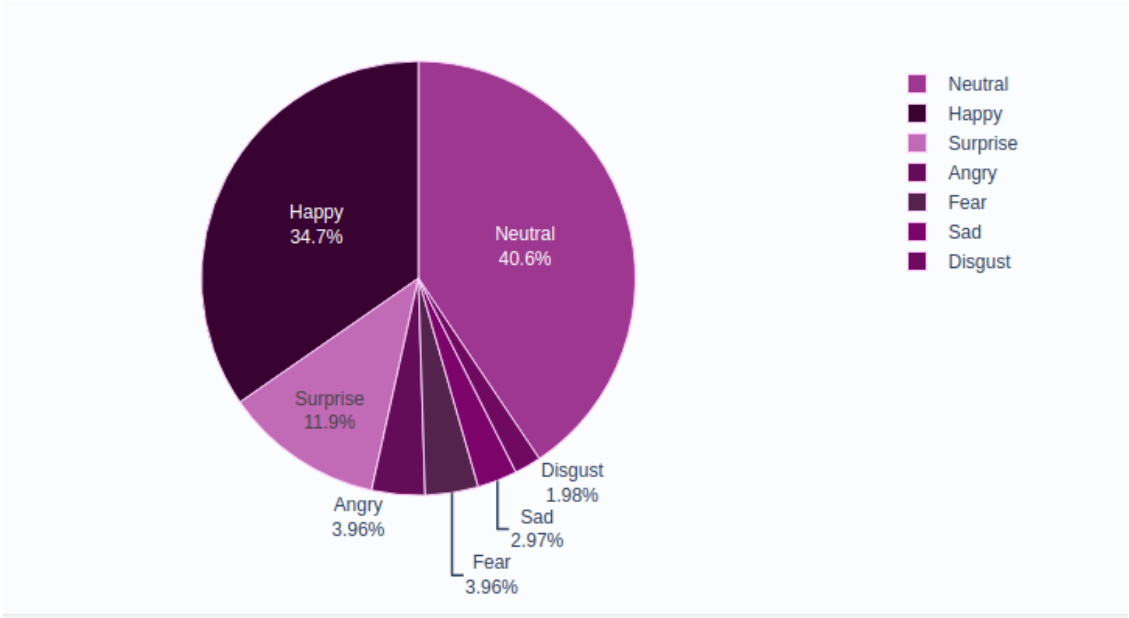
Both techniques on the above use Convolutional Neural Networks, but the data sets that they expect and use are different, as their respective categorizations.

Emotion Estimation from Face Analytics

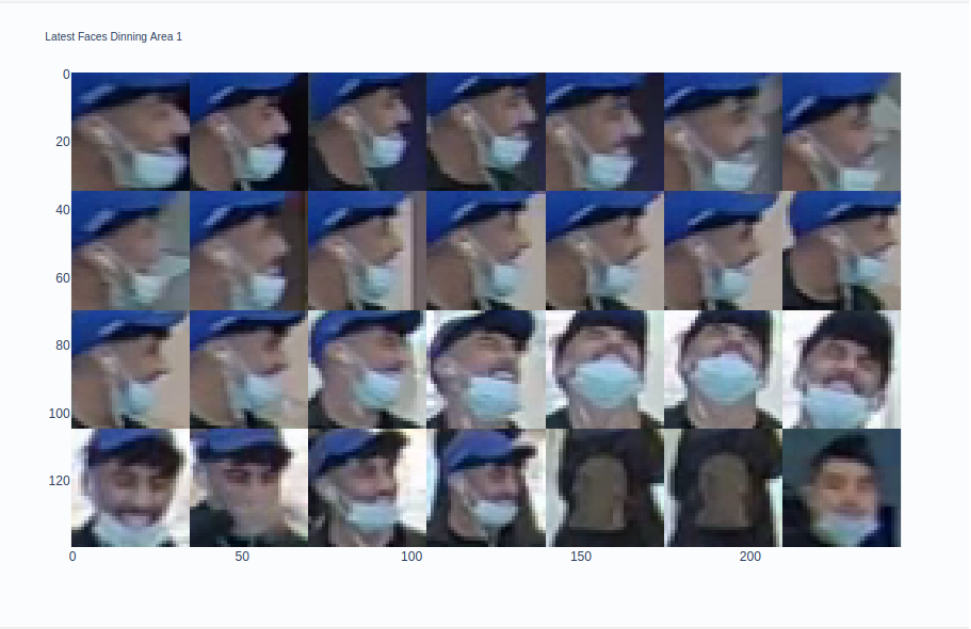
Emotion Estimation from Face Analytics is category-based and is dependent on seeing the person's face. The training samples for this Neural Network is similar to the above sample:

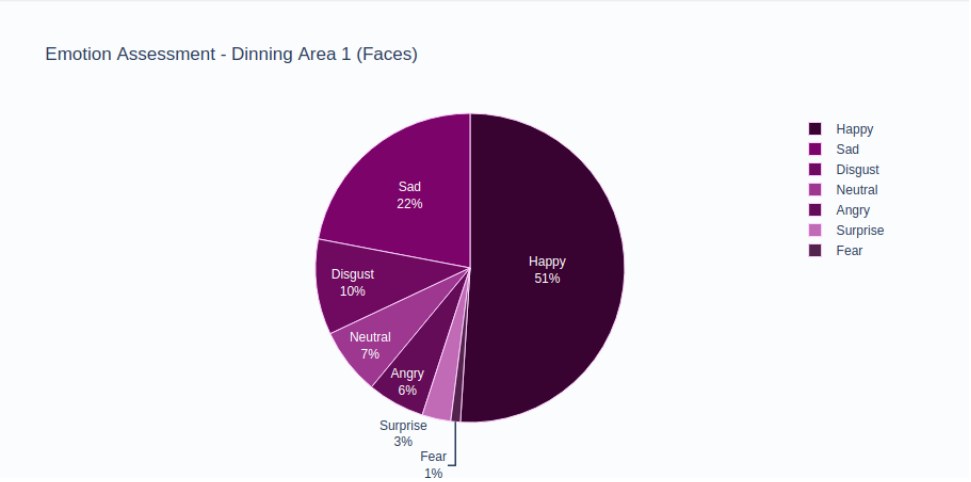


Around 3000 different faces for each basic emotion is used to train the Neural Network. The NN, when seeing a face, will try to categorize it and will provide a percentage of probability for the face seen (i.e, something like the below)



The system will accumulate for every camera the latest faces seen and will estimate a probability of emotion for all faces analyzed:





This technique does not estimate people where faces are not seen. The categories cannot be changed without retraining the entire neural network.

Due to the prevalence of masks and the natural cultural face coverings in the region, the sampling for face analysis will not cover all of the visible people on the camera shots.

Valence and estimation from full Body posture.

Our second emotion estimation technique is **Valence from full Body posture.** In this case, emotion is estimated from the full-body, which can return slightly different results than from face analysis. For example, the following face might be classified as pain:

But on further review, if the whole body is included then the same sample might be classified as exhilaration/joy:

As opposed to facing analytics, the result of full body emotion (valences) is returned not in a category with percentages bit in a valence scale from 0 to 10, where 0 is a negative emotion and 10 is very positive emotion:

For this NN a sample of 80.000 images of bodies is used. In some of these images, faces can be seen, in some others, faces are not visible but qualification is still given. For instance, the below image is tagged as 2:

While the following image is tagged as a 8:

Notice that both images don't have visibility of faces, but can still be used for purposes of training the NN to understand apparent emotion on a scale from 0 to 10 (called a valence).

The above is trained on 80,000 images.