**Preprocessing**

Various signals and facial features are extracted from videos using Py-feat, particularly the AU (Action units) and emotions were extracted as they can be of importance in deception detection. The features are extracted every 30 frames (for computational reasons, a more frequent extraction will take so much more time) then the resulting records are combined into one mean feature vector.

**Model Training**

After preprocessing, the mean feature vectors are passed to ML models like SVMS and XGB and RF etc.. with XGB yielding the highest accuracy. Deep learning models like ANN performed poorly.

**Important remarks and discoveries**

Highest accuracy achieved was 75% using XGB.

Using a complex technique such as extracting the features more frequently across the video frames and then passing the sequence of features to an RNN might yield better results but it is so heavy regarding the computational aspect so we decided to skip this part for now as we want a model that is both accurate and efficient.

It is possible to use the annotations that are provided with the dataset (manually annotated by experts) but we want our application to be fully automated so that when a user uploads a video the video would be annotated automatically without human intervention.