**MAINTENANCE AND UPDATING THE DATASET AND MODELS**

**Tools Used/Required**

The tools that are being used for the deliverables are mentioned below:

1. Jupyter Notebook(Anaconda 3 package)
2. Tensorflow( Version 2.4)
3. Scikit Learn package (Version 0.23)

The version needs to be checked and updated with time if necessary else there will be some issues in model training.

**Datasets**

With keeping in mind the future expansion of the project, there will be addition of more images from various videos for which we have two scenarios in scope:

1. For introduction of any new feature apart from the exsiting one done , create new data set as per the steps involved in creating the existing one , that is , segrgate the images and separate them into the subfolders named as the feature that needs to identified . This segregation of the images will involve manual intervention as per Civita Laurea scope of decision to assign the feature for the images. This will help in preparation of Training, Validation and Test data in future . Once done and trained the model can predict on any unseen, unlabeled data.
2. For the existing feature analysis , The training data should be audited regularly because if we consider the millions of videos available or made for educational purpose. There will be a need to update the training data for existing feature with changing time to make the model incorporate and adapt to the variations of videos. The new data needs to extracted from the new videos and added to the trainig data folder.

**Models**

With the incorporation of new data in the future , there may be a scenario where the models needs to be tuned for efficiency. Such as the current model InceptionV3 behaves good with the current feature analysis of Teaching style and presence of Visualizations . But with upgrading of features or large data in future , the models need to retrained and audited , if it seems that the accuracy is decreasing or the model is showing more signs of overfitting and underfitting than the model should be adjusted by changing the below parameters:

1. Sequential layers of the models used , they can be increased
2. Changing the value od dropout parameter or adding more dropout parameters
3. Adding more neurons in Sequential layers such as the value in the Dense layer can be changed from 512 to 1024 or so on.

All the changes may vary with the Models outcome , it is an iterative process of training the data, checking the outcome, hypertuning for a better outcome. Different model may behave differently in respect to data and features

**Data Tagging and Prediction Merging codes**

The code for tagging the dataset is a simpler and does not need a frequent update until there is a significant or mandatory change in the Scikit Learn package espacially os library. Same condition goes with the merging code for predictions.