**IMAGE CAPTIONING REPORT**

The PS is that for a given input of an RGB image, the final output expected is the caption describing the fed input image.

First I analysed the dataset. It consists of image ids, the comment number and the actual comment for each image. There are a total of 5 different styled comments for each image. So we can randomly choose 1 comment for each and map it to that image. After doing some data science and cleaning(like altering columns and changing shapes), I then proceeded to manually create the json file containing 2 keys ‘image’ and ‘comment’ for each image for feeding in the load\_dataset package. Then I created separate train and test splits of the dataset.

My training set contains about 1280 images and test set has 320 images.

Then I have selected a hugging face model called git-base (other options are blip image captioner large and nlpconnect) to fine tune with my training dataset. I preprocessed the images and comments i.e tokenized and normalized the tensors for easy feeding during training.

Next I created an object to the processing class. I have then loaded my training set object into DataLoader in batches of 2.

I have then imported the pretrained model from transformers and then set my output parameters and loss values to be printed for each epoch. I’m using Adam as optimizer for grad descent here. The training has been done through GPU acceleration. For each epoch the loss values get printed continuously as they get updated (reduced) in back propagation.

Then I have tested the model on the test dataset.

At this point it is generating some bad outputs which can be due to lesser training data and also lesser epochs or can also be due to overfitting.

I HAVE NOT INCLUDED METRICS AS THEY ARE VERY COMPLICATED FOR IMAGE CAPTIONING PROBLEMS