**Approach 4:**

After a lot of trials to make the model change “a w a y a w a y i f t h e y”

{ we still notice that the training is too fast for the model} we notices that the function load data does not load all the data set because of the If condition so by removing it “InD” length change from 9007 to 10500

When “InD” length was equal to 9007 it was Including a lot of empty cells after removing it its length become 318 (because of this we noticed the problem)

So, we have change the function load data to maximum amount of the data to be used in the model

We starts loading data by its path :

A screenshot of a computer

Description automatically generated

After this we have read the lines and start to separate the Speakers and the conversation by “:” then we took all conversation data and start splitting it into inputs{starts from index 0 and end at index i} and outputs {starts from index 1 and end at index i+1} and we have made a condition to check if there any empty cell to get ride of them before appending them to inputs and outputs and also before appending the conversation lines is passed to the function preprocess\_sentence to get clean it and to get ride of any unnecessary symbols or space

A computer screen shot of a black screen

Description automatically generated with low confidenceA screen shot of a computer

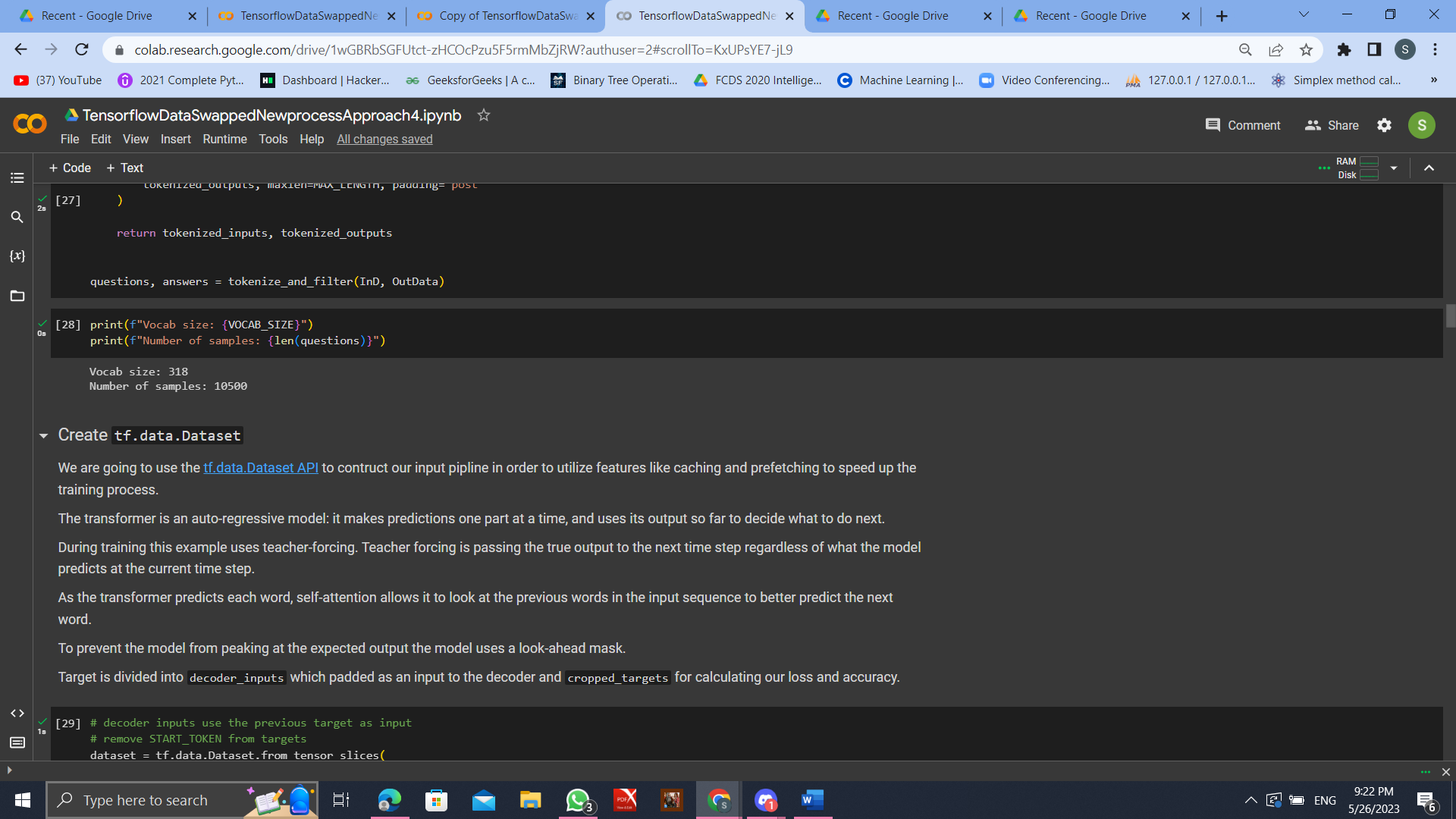
Description automatically generated

A picture containing text, screenshot, software, multimedia software

Description automatically generatedAfter this we noticed that the length of “InD” and “OutD” increase and became more clean

A screenshot of a computer

Description automatically generated

And after using tokonizer we find that VOCAB\_SIZE increased and of curse number of samples

A screenshot of a computer

Description automatically generatedSo, After All these change we have to start training the model on the New Processed Data

Experiment#1:

* Optimizer AdamA
* Num\_layer=2
* Epochs =50
* DropOut=0.1
* Num\_heads=8
* Max\_Lenght=316
* Max\_Samples=10500

A screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated

A screenshot of a computer program

Description automatically generated with medium confidence

A screenshot of a computer program

Description automatically generated with medium confidence

A screenshot of a computer

Description automatically generatedWe noticed that we have to increase no\_of\_epochs and decrease dropout layer

Experiment#2:

* Optimizer AdamA
* Num\_layer=2
* Epochs =250
* DropOut=0.0001
* Num\_heads=8
* Max\_Lenght=316
* Max\_Samples=10500

A screenshot of a computer

Description automatically generatedA screenshot of a computer

Description automatically generated

A screenshot of a computer

Description automatically generated