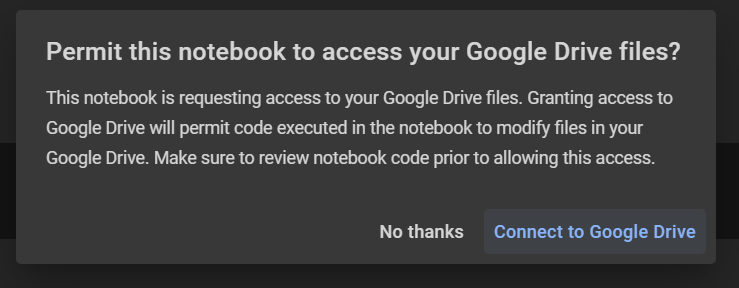
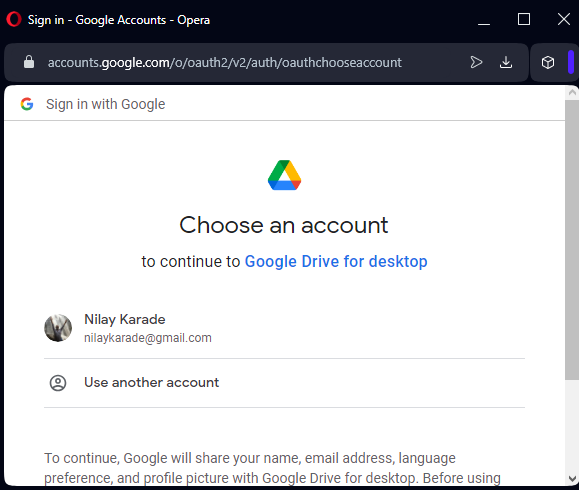


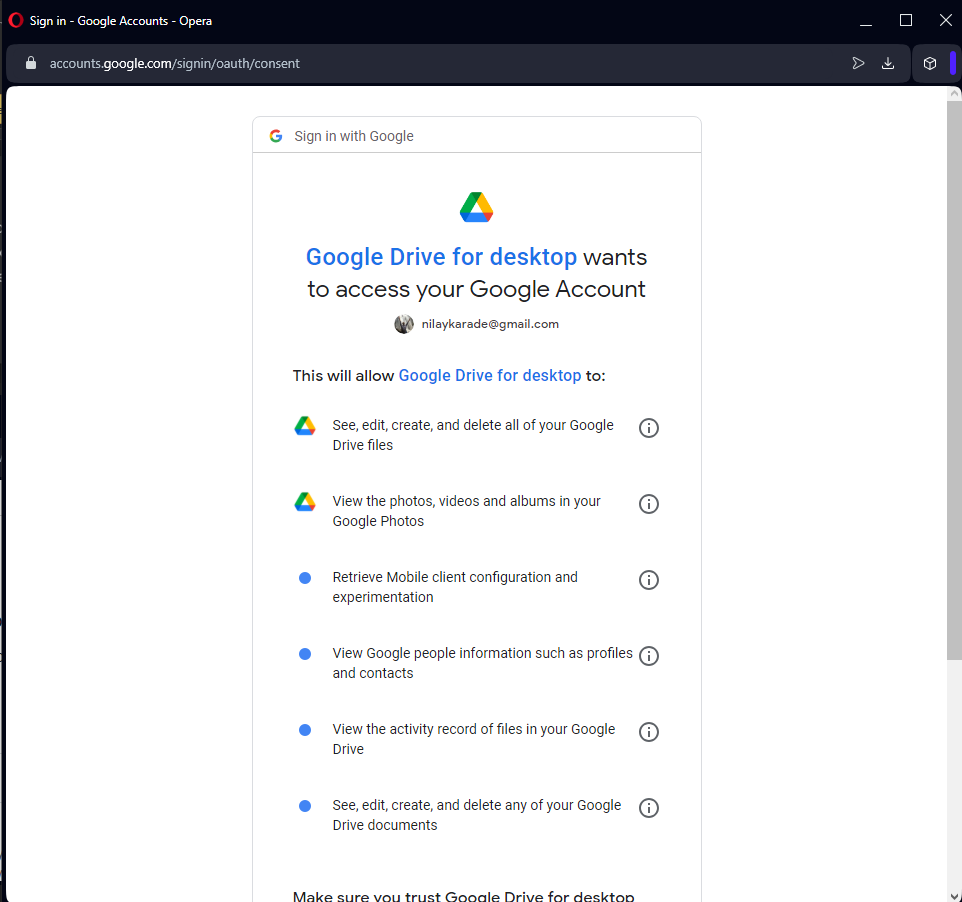
Source: https://github.com/JohnSnowLabs

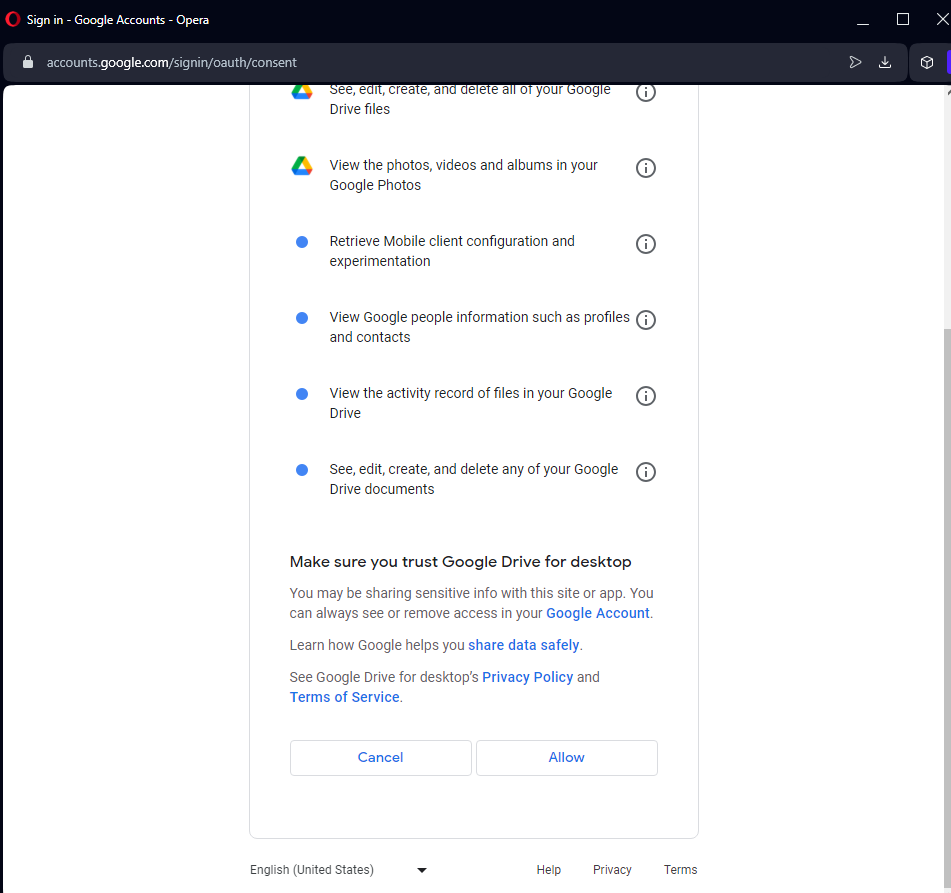
Text normalizer – removes special symbol(punctuation marks, html tags), converts the text to lower case

Finisher converts text to array format that can be used by ML models

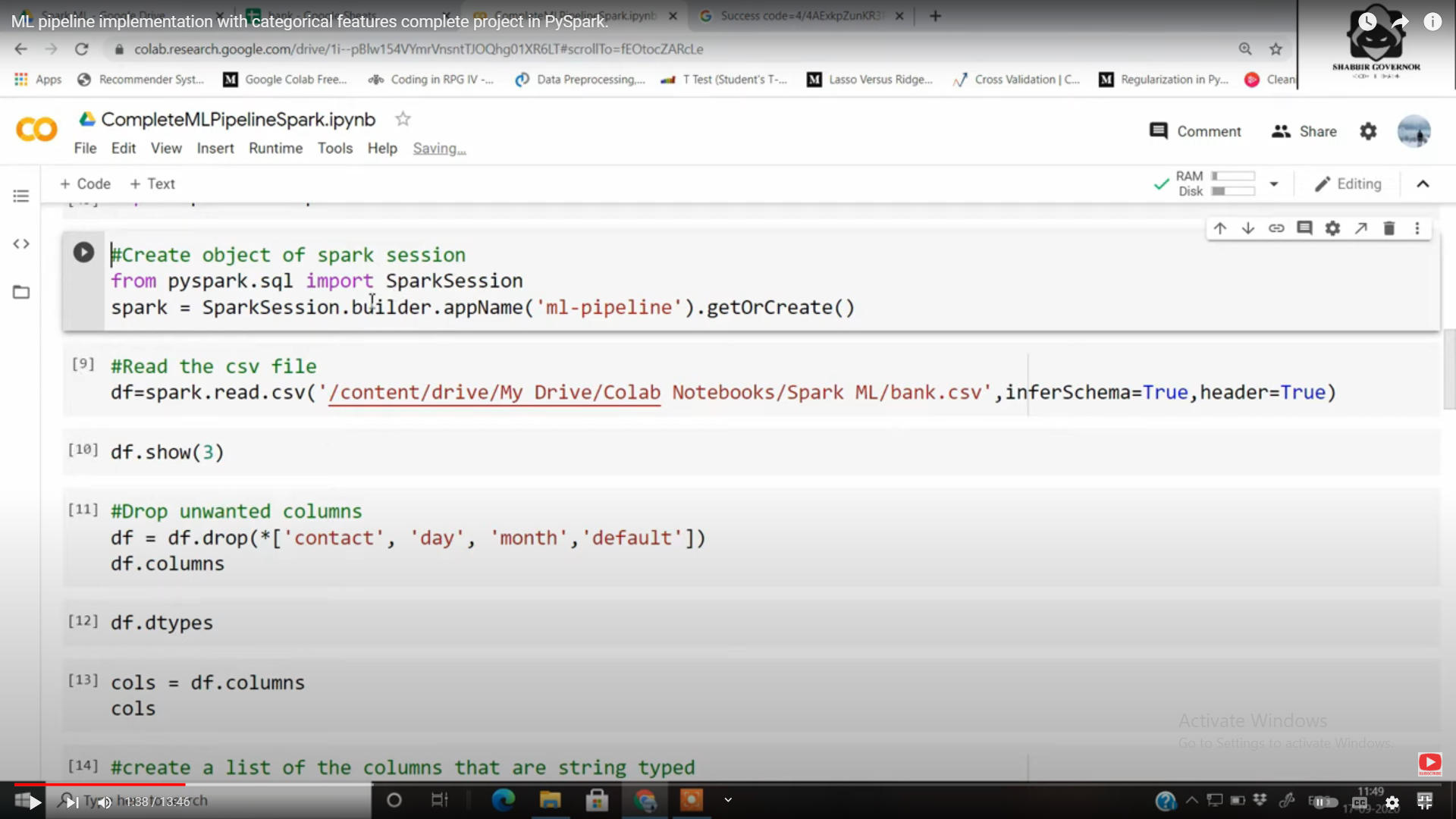






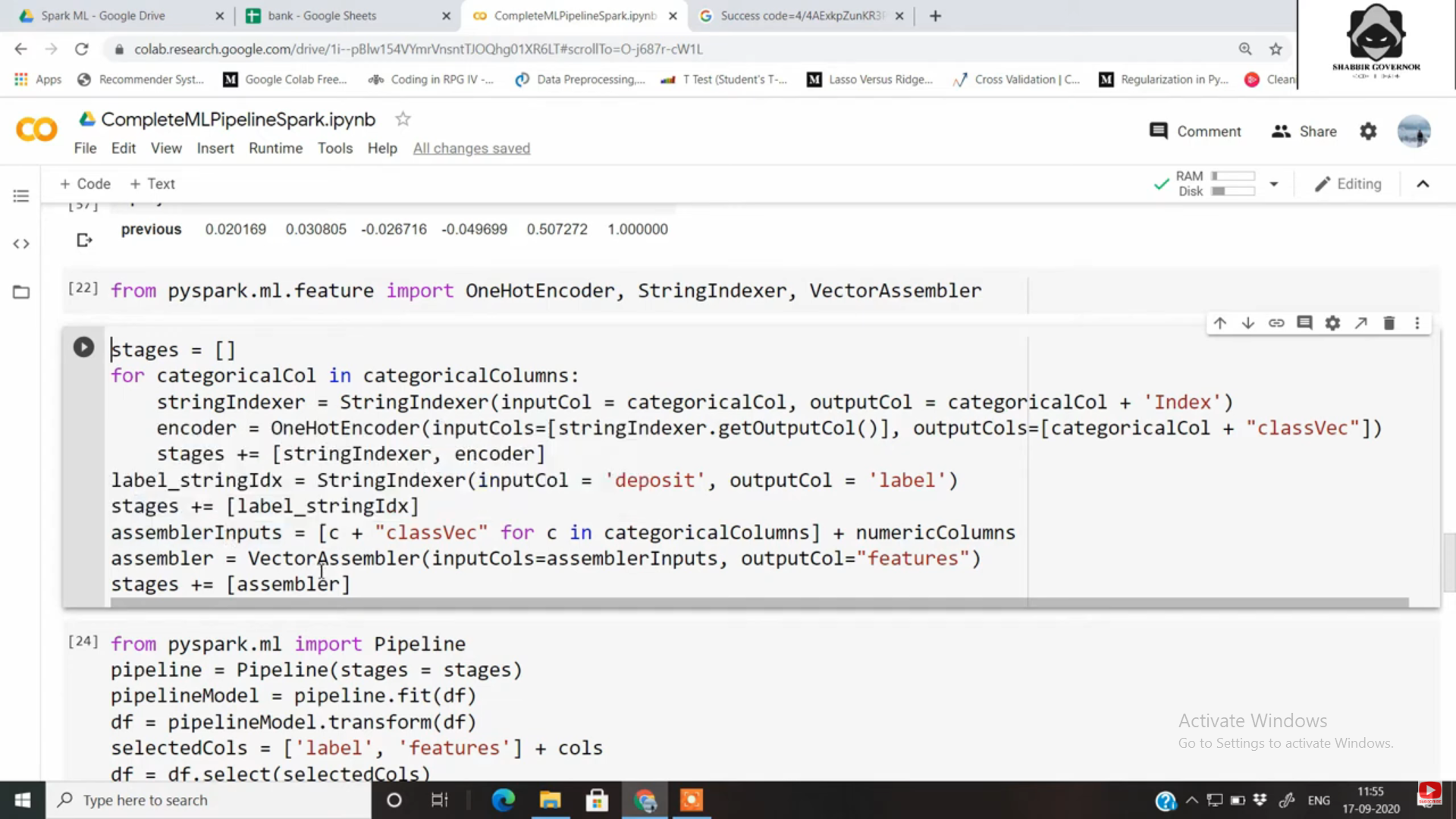


What is pyspark session



Inferschema?

Stages is list of operations to be performed on the data cols (preprocessing)



Fit and transform?

maxIter? – no of times training will happen 