Dear Prof. Fedyk

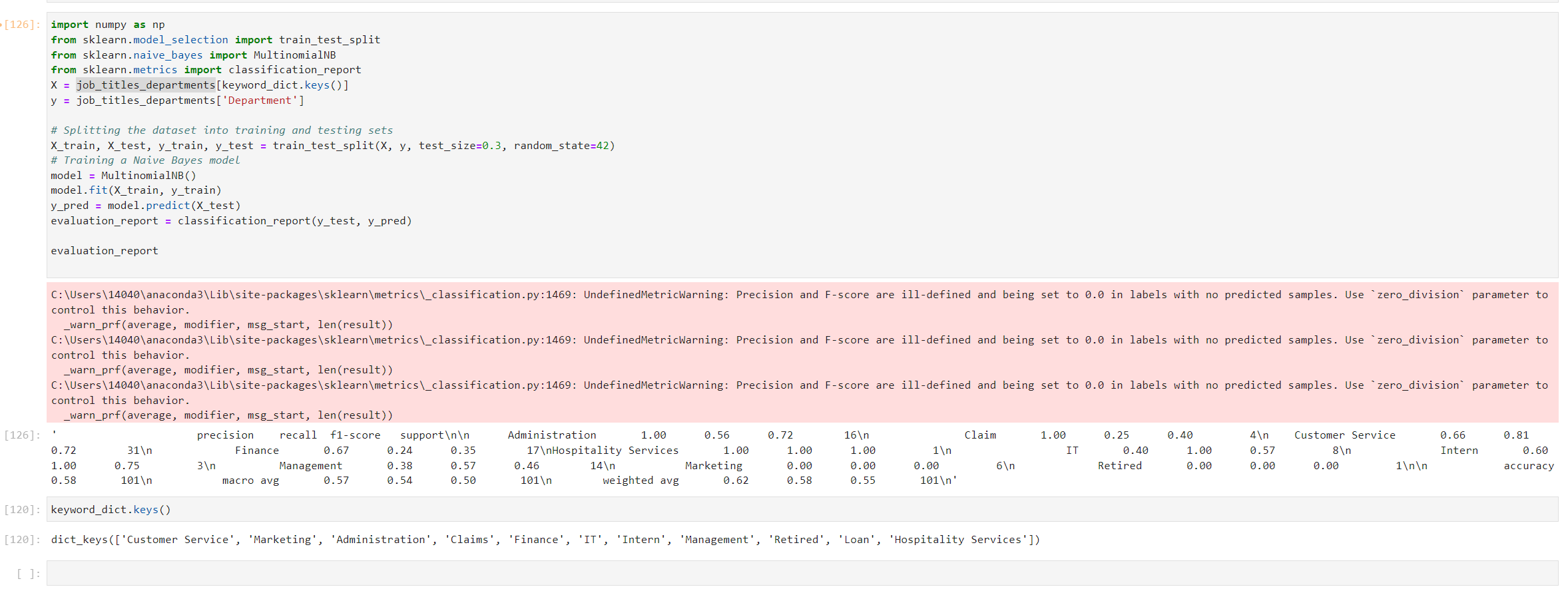
Thank you for the good news. And thank you for the great opportunity.

For this sample task, I went through 1000 rows and got 335 distinct values. I put all the 335 into a Word and Excel file named “job\_titles\_departments”. For the keywords, I have put them into a separate file named “keyword”.

For the department I was asked to classify, I personally made 11 distinct departments and assigned different keywords to each of them. Maybe the name of the departments is different from your thought. I am so sorry for it.

Although I noticed that we don’t need to actually build a classifier, I still made a simple one and tested it. I also attached an ipynb file to the email. If you want to take a look at how I preprocess data and make the classifier, it is over there and I also make annotations for every single cell I think necessary.

For the overall time, I used around 2 hours to take a look at the data sheet and do basic preprocess (like drop duplicate), and I used around 6-7 hours to go through all 335 distinct values. It is long because it took me some time to deal with the terms I am not familiar with. Finally, I used around 4 hours to build a simple classifier and test it. Here Are the results I got:



If you want, you can also access these files through this GitHub link: https://github.com/yuxuan-wei118/URAP\_Project.git

Thank you for your patience to read! Hope to hear good news from you!

Best,

Yuxuan