Dear Prof. Fedyk

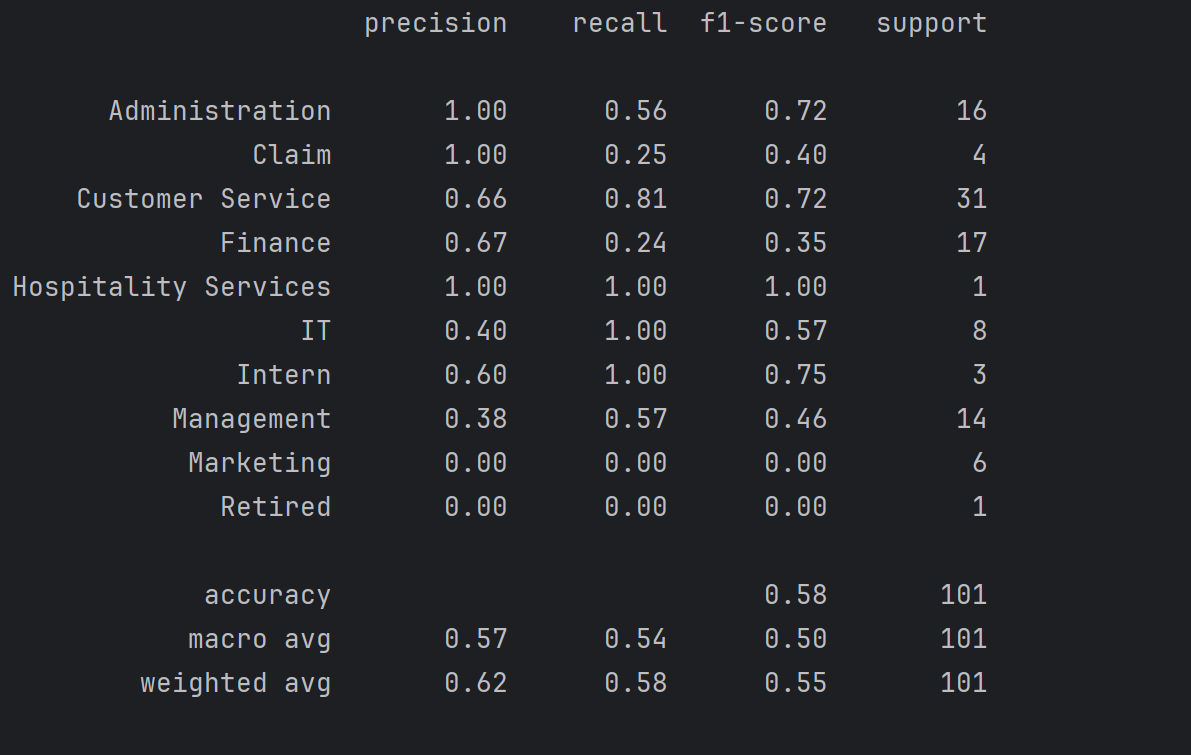
Thank you for the good news. And thank you for the great opportunity.

For this sample task, I went through 1000 rows of data and got 335 distinct values manually. All data are stored in the CSV file named “job\_titles\_departments”. The keyword that is intended to be used for classification is stored in the “Keyword” file.

Python code is used to check the accuracy of the classifier. I split the 335 value into a train set and a test set. The result has already been attached to the below.

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.

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 classified and test it. Here Is the result I got:



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

Best,

Yuxuan