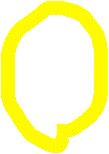
In the experiment above, we wanted to see which instances are predicted differently in the models trained on original dataset vs expanded dataset. In the excel file, Sheet 1 is the instances predicted differently, Sheet 2 is the models’ predictions and Sheet 3 is the distribution of the features in the X\_test and distribution of the instances predicted differently. I used Simple RF with 70 estimators and 10 depth for these results.

In the distribution of the instances predicted differently (Sheet 3), we can see that there are higher number of wrong prediction on the instances where martial\_status is ‘married-civ-spouse’ and relationship is ‘husband’.

A screen shot of a graph

Description automatically generated



A screen shot of a graph

Description automatically generated