I tried different features and add or replace them by different threshold, but the result is almost the same at 94% or so. The new feature doesn't help much. I think this is because it is from the old feature and the old feature is more preciser than the new one in the case of adding. If we replace the old feature with the new one, the accuracy will drop down a little bit, because the new feature is from the old one by thresholding so it can't totally take the place of it. By the way, from the feature importances, we can see that the model does not depend on a particular feature very much, so the effect of replacing won't change the accuracy very much. For future recommendation, we can use permutation to add or replace the features or change threshold to the mean value of the column and then compare to a baseline to get better result.