Due 03/03 11:59 pm

Two distributions p and q are close to each other if their KL-divergence is close to 0. DL-divergence of p to q over a set of outcomes is defined by D left parenthesis p vertical line vertical line q right parenthesis space equals space sum for i of p left parenthesis i right parenthesis log space fraction numerator p left parenthesis i right parenthesis over denominator q left parenthesis i right parenthesis end fraction .

1. Compute the distribution of classes in the training data on the "Priority" label obtained from df in TextClassification.ipynb. Is this distribution close to that of df?
2. Compute the distrubiton of classes in the training data on the "Priority" label obtained from df\_balanced in TextClassfication.ipynb. Is this distribution close to that of df\_new?
3. What conclusions can you draw about the training set selections against the underlying datasets from the label-distribution point of view?