

Abstract

Process Classification with algorithm machines learning aims to get the best accurate class target. However, in the real world several problem classifications appear when one class has a much lower probability in the training dataset. That problem called imbalanced dataset problem.

One popular approach to solving the imbalanced dataset problem is rebalancing dataset with technique sampling. Generally, technique samplings work to balancing the imbalance data. In this final exam, analysis will use three-technique sampling: *oversampling*, *undersampling* and *Combine*. After rebalance, for measuring the data it will be use the value of precision, recall, Fmeasure and HammLoss as of a result in classification data using *SVM*, *kNN*, and *NaiveBayes*.

In this final exam, analysis will learn to compare the result data value on classification process before and after change data with method sampling. Beside that, process analysis will be trying to find the best technique sampling for rebalance data.

Keyword : imbalance dataset problem, sampling, classification, precision, recall, and Fmeasure.