

Abstract

Cost Sensitive Learning is an approach that used in classifier studying process which in classified process, we must also consider misclassification cost. Misclassification cost is a cost that occurs everytime there is a data mislabel. Misclassification cost is very important because when a data is mislabeled it may give different cost. We used AdaCost studying method as cost sensitive learning implementation in classifying imbalance data that using misclassification cost estimation. And as classifier method we used Naïve Bayes, which this method using probabilistic estimation to solve classification problem.

In this Final Project, it has been analysed and implemented imbalance data classification using AdaCost and Naive Bayes. The test was done in different scenarios, such as the test using different number of boosting rounds, the test to compare AdaCost performance with another classification and boosting methods, and the test using data that contains missing value. Classification methods which is used for the comparison are Decision Tree, OneR, Neural Network, and Naive Bayes. While for the boosting methods, we used AdaBoost, AdaBoost M1, MultiBoost AB, and LogitBoost.

From the test result it has been concluded that AdaCost algorithm prioritize searching imbalance class pattern, boosting the performance of the base classifier to classify imbalance class, more than AdaBoost, AdaBoost M1, Multiboost AB or LogitBoost. But, AdaCost's performance depends on its boosting rounds.

Keyword : *cost sensitive learning, misclassification cost, data imbalance, adacost, naïve bayes*