

Abstract

In Data Mining, specifically, the problem of imbalance data-sets has been considered as one of the emergent challenges that must be resolved. Imbalance is a condition in which portions of data for purposes of learning in a class of more (major classes) rather than the other class (minor class). These problems can hinder the performance and accuracy on Machine Learning. One case that has Imbalance issue is Churn prediction. Churn occurs when all the services used by the customer or the company terminated.

A standard classification algorithm has a bias of the major classes that tend to missclassification in minor-class. GP-COACH (Genetic Programming-Based Evolutionary Algorithm for Learning Compact and accurate the FRBCS) can be used to handle Imbalance class with a rule solution. The begin process will be done preprocessing data (outside the system) using Weka tools, such as: Feature selection, Resample, and SMOTE. It is intended to minimize the amount of data that are too large and through SMOTE to improve churn information by duplicating the synthetic data. Furthermore, the data can be processed by the method of GP-COACH through the system.

The test results show that the performance of the GP-COACH is able to handle Churn prediction with the evaluation of prediction accuracy is good value and generate some simple solutions rule. But this method is not better than C5.0, both in terms of accuracy and duration of the execution time. However, the number of rules generated from C5.0 is far more than the GP-COACH. Thus the method of GP-COACH is still able to handle Churn prediction rule through several fewer solutions than the rule that is used C5.0 decision tree in classifying a class.

Keywords: *Churn prediction, Imbalance, Machine Learning, FRBCS, GP-COACH, C5.0*