Abstract

Imbalance class is one of existing problems in data mining. This problem is caused by number of data in which is uneven in the different class, where the one of the class own the number of a lot of data (majority), while the other classes own a little the number of data (minority). In the ordinary classification, the minority class cannot classified. This is caused by when the number of data in the very small class, the class will be wrong classified as majority class.

One of classification method which commonly use is decision tree method. This method will form model like flow chart with structure tree. In this final project is comparing three algorithm of decision tree that is C4.5, CART, and HDDT. Every algorithm own splitting criterion or different attribute selection measure with level of different sensitivities also to data skewnes.

The analysis in final project is knowing how performance from the C4.5, CART, and HDDT algorithm owning different splitting criterion in problems of imbalance class and also how comparison of performance from third of the algorithm decision tree based on value of precision, recall, and f-measure.

The result from this research present that HDDT algorithm own better performance at low level of imbalance compared to CART and C4.5 algorithm. This is caused by algorithm of HDDT own level of low sensitivities to skewness of data becoming internal issue of imbalance class.

Keyword: imbalance class, algorithm C4.5, CART, HDDT, performance.