Abstract

Problems that occurred in the data is the amount that is too big. One of the problem is the number of attributes that exist in the data. To handle the problem, we need to make data reduction on dimension of attribute. This technique is called Feature Selection. This technique is one of the techniques performed on the data preprocessing. The purpose of doing feature selection, in addition to reducing the number of attributes, will be able to give a better performance of classification when we compared to data that without attribute selection.

In this research, the author implements a feature selection algorithm called ReliefF. This algorithm is attribute selection algorithm that based in instance. The selection attributes is done by calculating differences weight for each instance with selected by randomly as near hit (nearest neighbor was elected in the same class) and near miss (nearest neighbor was elected in the different class).

Calculation of performance will be based on new data that contains data with attribute that are selected from feature selection for the classification. From the classification process, difference of performance will be calculated the data that has not been done with the selection of features that have made the process of selection of features. Results are captured the value of precision and recall.

The results of the implementation, testing, and analysis on this Final Project shows that the performance of the algorithm is highly dependent on the number of iterations performed, the number of nearest neighbors, election of threshold and also the quality of the randomly selected instant when the algorithm is run. Dataset that has undergone a process of selection of features has been able to improve the performance of the classification results.

Keyword : feature selection, ReliefF, classification, near hit, near miss.