

ABSTRACT

The rapid growth of the accumulation of the data has created a condition which is often called as rich in data but poor in information. Data preprocessing is one the way that can be done to resolve this problem. One part in the data preprocessing is data reduction, which is a technique that can be applied to obtain the reduction of the number of datasets with a smaller total volume of data. The method used to apply the data reduction is the method of Density-Based Multiscale Data Condensation, aimed to reduce the data at different scales. With these different scales, is expected to obtain information that can represent the whole of the data. The use of different scales are applied to determine the radius of data reduction which depends on the proximity distance between data. The result of the performance of data preprocessing, which is the value of precision and recall increase and model building time become faster.

Keywords: *data preprocessing, data reduction, Density-Based Multiscale Data Condensation, scale, precision, recall, time.*