ABSTRACT

Information technology that has developed rapidly allows many companies to collect and store a very large amounts of data. Data mining is a process of extracting and analyzing this large amounts of data to obtain the interesting patterns of them, thus we can get knowledges from the data. One of the data mining functionalities is classification, which in this process is to find a description model that can classify the data, for the use of predicting objects that the class lable is still unknown.

This final project implement the MIND algorithm as one of classification algorithm using decision tree for a very large amounts of data, and make an analysis to the performance of the algorithm implementation, which are the classification accuracy, processing time (classification time), and the scalability, where the results are compared to SPRINT algorithm. The implementation itself is done using SQL Server 2000 and Delphi 6. Data used in performance testing are varied from 57 to 3,163 records, and from 10 to 28 attributes, with numerical and categorical attribute types and also the combination of both, and synthetical data to figure out its scalability.

The result we get that MIND algorithm implemented here is proven to have higher accuracy level than SPRINT for being able to give 100% accurate while SPRINT can only give highest accuracy at 99.08%. The decision tree result is sufficiently good and comprehensible. However, MIND algorithm implemented here has slower processing time than SPRINT on the average 19.27 times, for data with many attributes (20 attributes and higher). For large data (100,000 records and higher), the implementation using MIND algorithm made here is not capable to result data classification, therefore this implementation is not scalable yet as a data classification tool.

Keywords: data mining, MIND algorithm, classification, scalable classifier