

Sensitivity of missing values in classification tree for large sample

ABSTRACT

Missing values either in predictor or in response variables are a very common problem in statistics and data mining. Cases with missing values are often ignored which results in loss of information and possible bias. The objectives of our research were to investigate the sensitivity of missing data in classification tree model for large sample. Data were obtained from one of the high level educational institutions in Malaysia. Students' background data were randomly eliminated and classification tree was used to predict students degree classification. The results showed that for large sample, the structure of the classification tree was sensitive to missing values especially for sample contains more than ten percent missing values.

Keyword: Classification tree; Missing data