

Abstract

Data mining is the extraction technique or the discovery of patterns, information, knowledge from large amounts of data. In the process of discovery data quality factor is very important. If the data is not quality then the quality of the data analysis becomes less accurate. One of the determinants of its quality factor is missing value of the data used. Missing values will be significant if the amount of missing values is large enough that could affect the accuracy and quality of the conclusions resulting pattern.

In resolving the case of missing value imputation method will be used. Imputation method used is the Expectation maximization algorithm. So in the Final Project will be conducted the analysis with the handling of missing values Expectation maximization algorithm. The purpose of this analysis is to see the performance of these algorithms in dealing with missing values using the parameters Normalized Root Mean Squared Error (NRMSE). Then analyze the data imputation using the classification of KNN in Weka with the parameters used are Precision, Recall and F-Measure. Based on the results of research has conducted, it was found that the algorithm is able to handle missing values properly and data imputation produced has a quality that is almost adjacent to the complete dataset used.

Key words: *Data minining, missing values, Imputation, EM algorithm, NRMSE, Precision, Recall, F-Measure.*