Abstract

Missing value is one of the common problems that occur in the data. *Missing value* is missing information in the data. Handling *missing value* is one of the tasks in data preprocessing, aimed so that the data become more qualified to serve as input for the process of data mining. One way of handling is by *imputation*. This *imputation* is that will automatically replace the *missing value* in the data with a number of plausible values.

This final assignment, the authors implement one method of imputation that is *Robust Least Squares Imputation with Principal Component* (RLSP). Stages of the process of imputation methods include the selection of *k* similar instance, the process of formation of principal components and *imputation* by using quantile regression based on principal component obtained in the previous step. This process produces data *imputation* results and performance within the parameters of *Normalized Root Mean Squared Error* (NRMSE).

Evaluation of the imputation sistem is based on the results NRMSE directly and indirectly through the process of classification based on the parameters precission and recall. Classification process is done by *Naïve Bayes classifier* using WEKA. The purpose of this indirect test is to determine how well the imputation of data quality results when used in the process of data mining.

Keywords: *missing value*, *imputation*, RLSP, NRMSE, *classification*, *Naïve Bayes*