

Abstract

One of the common problem that are mostly held in raw data is missing values. This will certainly affect the results of statistical analysis on data and reduce the level of data accuracy. Therefore data preprocessing stage needed for handling missing value. One way is to fill in the handling of the predicted value obtained by applying certain algorithm, which is called imputation method. Missing value imputation method used in this Final Project is *Predictive Mean Matching* (PMM).

PMM is a method of imputation which apply algorithm-based model, where the missing values were filled by the estimated value obtained from using the *Least Square Regression* models are formed from the rows that have no missing value in the data, then the imputed missing values from the closest observation of the model. Imputation test results on the system performance is evaluated through the parameter *Mean Squared Error* (MSE). At this final project is also carried out analysis of the influence of imputation of data for the classification process with parameter testing *Precision*, *Recall*, and *F-Measure*. Based on the results of tests performed, it is evident that the PMM is an imputation algorithm to estimate the missing values are close to the actual value.

Keywords: imputation, missing value, *Predictive Mean Matching* (PMM)