

Abstract

Missing value is the condition on the data where the data has a value attribute that is incomplete or missing. Missing values in the data can be handled in different ways, one of which is the imputation of missing values so obtained are suitable for filling the missing value. CMI (Cluster-based Missing Value Imputation) is one of the missing value imputation methods of numerical data based on clusters. The data will be clustered by CMI use K-means clustering and then will fill in missing values with the help of kernel functions based on the values that exist in clusters in which the missing value is located. Kernel function in CMI aims to address non-parametric imputation where the data is not considered to be regressed in general. Tests performed were divided into two parts, namely the direct testing and indirect testing. Direct testing performed by calculating the RMSE of the estimates used to fill in missing values to actual values. Indirect testing is done by calculating F-measure of the classification results of test data on the actual class of data prior to testing. CMI can handle missing values with a relatively good level of accuracy seen from the results of RMSE and F-measure is obtained.

Keywords : missing value, cluster, numerical , CMI, non-parametric