

INTISARI

Missing value merupakan salah satu permasalahan yang sering dijumpai dalam penelitian. Penanganan *missing value* dapat dilakukan menggunakan metode *multiple imputation*. Prinsip metode *multiple imputation* adalah mengisi lebih dari satu nilai estimasi pada *missing value*. Adanya kendala pada *multiple imputation* berupa *imputed dataset* yang memuat *outlier* dapat menyebabkan keakuratan data menjadi kurang optimal. Pada beberapa penelitian, peneliti menggunakan seleksi atribut untuk membuang *outlier* agar akurasi data menjadi optimal.

Tujuan penelitian ini adalah mengkombinasikan algoritme *multiple imputation* dan seleksi atribut. Algoritme yang digunakan pada penelitian ini adalah *markov chain monte carlo* (MCMC) dan *fully conditional specification* (FCS). Sedangkan seleksi atribut dilakukan menggunakan *principal component analysis* (PCA). Metode yang diusulkan menawarkan dua kombinasi metode yang berbeda yaitu *impute first then select* dan *select first then impute*. Perbedaan dari kedua metode adalah tahapan penggunaan seleksi atribut pada penelitian. Pada metode pertama, seleksi atribut dilakukan setelah data melalui proses imputasi. Sedangkan metode kedua, seleksi atribut dilakukan sebelum data dilakukan imputasi.

Hasil akurasi terbaik adalah kombinasi FCS-PCA pada metode *impute first then select* dengan tingkat akurasi mencapai 98,45%. Sedangkan hasil klasifikasi sebelum dilakukan kombinasi metode sebesar 72%. Berdasarkan hasil penelitian, kombinasi *multiple imputation* dan seleksi atribut dapat meningkatkan keakuratan data. Hal ini dapat dilihat dari hasil keakuratan data sebelum dan sesudah dilakukan kombinasi metode.

Kata kunci -- *multiple imputation, markov chain monte carlo, fully conditional specification, principal component analysis.*

ABSTRACT

Missing value is one of the problems that is often encountered in research. Handling missing value can be done using multiple imputation method. The principle of multiple imputation method is to fill in more than one estimate value on missing value. The existence of constraints on multiple imputation in the form of imputed dataset containing outlier can cause the accuracy of data becomes less optimal. In some studies, researchers used attribute selection to remove outliers for optimal data accuracy.

The purpose of this research is to combine multiple imputation algorithm and attribute selection. The algorithm used in this research is markov chain monte carlo (MCMC) and fully conditional specification (FCS). While the selection of attributes performed using principal component analysis (PCA). The proposed method offers two different combinations of methods: impute first then select and select first then impute. The difference between the two methods is the use of attribute selection in the study. In the first method, attribute selection is performed after the data through the imputation process. while the second method, attribute selection done before the data is done imputation.

The best accuracy result is the combination of FCS-PCA on impute first then select method with accuracy level reach 98.45%. While the results of classification before done a combination of methods by 72%. Based on the research results, the combination of multiple imputation and attribute selection can improve the accuracy of the data. This can be seen from the results of data accuracy before and after a combination of methods.

Keywords : *multiple imputation, markov chain monte carlo, fully conditional specification, principal component analysis.*

INTISARI

Missing value merupakan salah satu permasalahan yang sering dijumpai dalam penelitian. Penanganan *missing value* dapat dilakukan menggunakan metode *multiple imputation*. Prinsip metode *multiple imputation* adalah mengisi lebih dari satu nilai estimasi pada *missing value*. Adanya kendala pada *multiple imputation* berupa *imputed dataset* yang memuat *outlier* dapat menyebabkan keakuratan data menjadi kurang optimal. Pada beberapa penelitian, peneliti menggunakan seleksi atribut untuk membuang *outlier* agar akurasi data menjadi optimal.

Tujuan penelitian ini adalah mengkombinasikan algoritme *multiple imputation* dan seleksi atribut. Algoritme yang digunakan pada penelitian ini adalah *markov chain monte carlo* (MCMC) dan *fully conditional specification* (FCS). Sedangkan seleksi atribut dilakukan menggunakan *principal component analysis* (PCA). Metode yang diusulkan menawarkan dua kombinasi metode yang berbeda yaitu *impute first then select* dan *select first then impute*. Perbedaan dari kedua metode adalah tahapan penggunaan seleksi atribut pada penelitian. Pada metode pertama, seleksi atribut dilakukan setelah data melalui proses imputasi. Sedangkan metode kedua, seleksi atribut dilakukan sebelum data dilakukan imputasi.

Hasil akurasi terbaik adalah kombinasi FCS-PCA pada metode *impute first then select* dengan tingkat akurasi mencapai 98,45%. Sedangkan hasil klasifikasi sebelum dilakukan kombinasi metode sebesar 72%. Berdasarkan hasil penelitian, kombinasi *multiple imputation* dan seleksi atribut dapat meningkatkan keakuratan data. Hal ini dapat dilihat dari hasil keakuratan data sebelum dan sesudah dilakukan kombinasi metode.

Kata kunci -- *multiple imputation, markov chain monte carlo, fully conditional specification, principal component analysis.*