

Implementasi algoritma expectation maximization untuk menangani missing value pada model regresi = A study of imputation methods for missing value with regression model

Dini Rahayu, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20494647&lokasi=lokal>

Abstrak

Masalah yang sering terjadi dalam penelitian adalah adanya missing value padahal data yang lengkap diperlukan untuk mendapatkan hasil analisis yang menggambarkan populasi. Dalam pengolahan data, missing value sering terjadi pada analisis regresi. Analisis regresi merupakan suatu model prediksi dengan melihat hubungan antara variabel respon dan variabel prediktor. Missing value dalam analisis regresi dapat ditemukan baik pada variabel respon maupun variabel prediktor. Penelitian ini membahas imputasi missing value yang terjadi pada kedua variabel tersebut dengan menggunakan imputasi regresi. Algoritma Expectation Maximization (EM) merupakan metode penaksiran parameter regresi dengan menggunakan metode Maximum Likelihood Estimation (MLE) pada data yang memiliki missing value. Untuk menyeimbangkan hasil taksiran parameter model regresi untuk setiap variabel, dilakukan proses penyeimbangan (balance process) untuk mendapatkan hasil taksiran parameter yang konvergen. Simulasi taksiran nilai variabel respon dan prediktor yang hilang dilakukan pada berbagai variasi persentase missingness. Metode penaksiran parameter regresi dengan menggunakan algoritma EM, dapat menghasilkan model yang menjelaskan data sebesar 87% hingga terjadi missing sebanyak 60%.

<hr>

The problem that often occurs in research is the existence of missing values even though complete data is needed to obtain the results of analysis that describe the population. In processing data, missing values often occur in regression analysis. Regression analysis is a prediction model by looking at the relationship between response variables and predictor variables. Missing values in regression analysis can be found in both the response variable and predictor variable. This study discusses the imputation of missing values that occur in both variables using regression imputation. The Expectation Maximization (EM) algorithm is a method of estimating regression parameters using the Maximum Likelihood Estimation (MLE) method on data that has missing value. To balance the estimated parameters of the regression model for each variable, a balance process is performed to obtain the results of the convergent parameter estimates. The estimated simulation of the value of the response variable and missing predictor was carried out in various variations in the percentage of missingness. The method of estimating regression parameters using the EM algorithm, can produce a model that explains the data by 87% until there is missing as much as 60%.