

# Imputasi Biclustering Berbasis Shifting-and-Scaling Similarity (SSSim) Menggunakan Regresi Kuantil Ke-t (NCBI-SSSim-tQR) pada Data Ekspresi Gen = Shifting-and-Scaling Similarity (SSSim)-Based Biclustering Imputation Using t-th Quantile Regression (NCBI-SSSim-tQR) on Gene Expression Data

Rasendriya Maheswari, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20529025&lokasi=lokal>

---

## Abstrak

Teknologi microarray merupakan alat terapan ilmu bioinformatika dalam bidang biologi molekuler yang dapat menghasilkan suatu data ekspresi gen. Namun, pada data ekspresi gen sering ditemukan missing value yang dapat mengganggu proses analisis data. Untuk mengatasi masalah tersebut, missing value dapat diestimasi menggunakan teknik imputasi. Penelitian ini menggunakan metode NCBI-SSSim-QR yang merupakan metode imputasi biclustering berbasis Shifting-and-Scaling Similarity (SSSim) dan menggunakan regresi kuantil dalam mengestimasi missing values. Metode ini merupakan modifikasi dari metode NCBI-SSSim-LS yang menggunakan regresi least square dalam proses imputasinya. Regresi kuantil pada metode NCBI-SSSim-QR digunakan untuk menangani keberadaan outlier pada data karena regresi kuantil dapat menyesuaikan kemiringan distribusi data pada titik-titik kuantil tertentu. Pada penelitian ini, metode NCBI-SSSim-QR diimplementasikan pada data ekspresi gen kanker serviks yang mengandung outlier dan diujikan pada missing rate 5%, 10%, 15%, 20%, 25%, 30%, dan 35% dengan mekanisme MCAR (Missing Completely at Random) dalam pembentukan missing values. Terdapat beberapa titik kuantil yang diuji yaitu kuantil ke = 0,3, 0,4, 0,5, 0,6, 0,7 yang kemudian performanya dibandingkan dengan metode NCBI-SSSim-LS yang berbasis mean. Penelitian menunjukkan bahwa proses estimasi missing values pada data ekspresi gen kanker serviks paling baik dilakukan oleh metode NCBI-SSSim-QR dengan kuantil 0,5 berdasarkan nilai Root Mean Square Error (RMSE).

.....Microarray technology is a bioinformatics tools in molecular biology that capable for quantitating hundreds or thousands of gene transcripts called gene expression data. However, the presence of missing values is often found in gene expression data which can interfere the further analysis process. Therefore, imputation technique used for estimating missing values. This research used NCBI-SSSim-QR method which is a Shifting-and-Scaling Similarity (SSSim)-based biclustering imputation method using quantile regression to estimate the missing values. This method is a modification of the NCBI-SSSim-LS method that using the least square regression for imputation process. Quantile regression in the NCBI-SSSim-QR method is used to handle the presence of outliers in the data because quantile regression can follow the skewness of the data distribution at certain quantile points. In this research, the NCBI-SSSim-QR method will be implemented on cervical cancer gene expression data containing outliers and tested on missing rate of 5%, 10%, 15%, 20%, 25%, 30%, and 35% with MCAR (Missing Completely at Random) mechanism in generating missing values. There are several quantile points to be tested, = 0,3, 0,4, 0,5, 0,6, 0,7 which will then be compared with the NCBI-SSSim-LS method which is mean-based. The research shows that the estimation process for missing values in cervical cancer gene expression data is better estimated by NCBI-SSSim-QR method with a quantile of 0,5 based on the Root Mean Square Error (RMSE) value.