

Implementasi Metode Sequential Biclustering Berbasis Skor Mean Square Residue dan Jarak Euclidean (SBi-MSREimpute) untuk Imputasi Missing Values pada Data Ekspresi Gen Pasien Kanker Payudara = Implementation of Sequential Biclustering Based on Mean Squared Residue and Euclidean Distance Method (SBi-MSREimpute) for Missing Values Imputation on Gene Expression Data of Patients with Breast Cancer

Adinda Dwi Putri, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20525198&lokasi=lokal>

Abstrak

Bioinformatika kerap digunakan oleh para peneliti untuk mempelajari berbagai penyakit yang ada pada tubuh manusia, salah satunya yaitu kanker payudara. Penelitian terhadap kanker payudara tersebut dilakukan dengan tujuan untuk menemukan jenis pengobatan terbaik bagi para pasien penderita kanker payudara. Data ekspresi gen merupakan salah satu komponen utama dalam penelitian mengenai pengobatan kanker payudara dan data tersebut dapat diperoleh dengan menggunakan alat dan teknologi microarray. Akan tetapi, seringkali ditemukan beberapa nilai yang hilang (missing values) pada data ekspresi gen yang dapat disebabkan oleh kesalahan teknis seperti kerusakan pada chip dan gambar. Adanya missing values juga dapat mengakibatkan masalah ketika proses analisis data selanjutnya, dimana terdapat metode analisis data yang memerlukan data lengkap seperti klasifikasi dan clustering. Oleh sebab itu, perlu dilakukan proses imputasi terhadap missing values agar hasil analisis data yang diperoleh lebih akurat. Pada penelitian ini, metode imputasi missing values yang digunakan yaitu SBi-MSREimpute. SBi-MSREimpute adalah metode imputasi berbasis biclustering dimana bicluster dibentuk berdasarkan suatu kriteria yang melibatkan skor Mean Squared Residue dan jarak Euclidean. Metode SBi-MSREimpute diimplementasikan pada data ekspresi gen pasien penderita kanker payudara stadium awal yang telah diberikan jenis obat MK-2206. Kinerja metode SBi-MSREimpute dilihat dengan membandingkan hasil imputasi metode SBi-MSREimpute dengan metode imputasi lain yaitu metode imputasi menggunakan weighted average berdasarkan skor Normalized Root-Mean-Square-Error (NRMSE). Hasil evaluasi dengan skor NRMSE tersebut menunjukkan bahwa kinerja metode SBi-MSREimpute dapat dipengaruhi oleh penentuan nilai k yang ada pada metode SBi-MSREimpute.

.....Bioinformatics is often used by researchers to study various diseases that exist in the human body, one of which is breast cancer. The research on breast cancer was conducted with the aim of finding the best type of treatment for breast cancer patients. Gene expression data is one of the main components in research on breast cancer treatment and this data can be obtained using microarray tools and technology. However, there are often missing values found in gene expression data that can be caused by technical errors such as damage to chips and images. The existence of missing values can also cause problems during the data analysis process, where there are data analysis methods that require complete data such as classification and clustering. Therefore, it is necessary to carry out an imputation process for missing values so that the data analysis results obtained are more accurate. In this study, the missing values imputation method used was SBi-MSREimpute. SBi-MSREimpute is a biclustering-based imputation method where the bicluster is formed based on a criterion involving Mean Squared Residue and Euclidean Distance. In this study, the SBi-

MSREimpute method was applied to the gene expression data of patients with early stage breast cancer who had been given the MK-2206 type of drug. The performance of the SBi-MSREimpute method is assessed by comparing the results of the imputation using SBi-MSREimpute method with other imputation methods, namely the imputation method using weighted average, based on the Normalized Root-Mean-Square-Error score (NRMSE). The results of the evaluation with NRMSE score showed that the performance of the SBi-MSREimpute method can be affected by the determination of k value in the SBi-MSREimpute method.