

Imputasi Missing Values pada Microarray Data Ekspresi Gen Menggunakan Fuzzy Biclustering Means Particle Swarm Optimization = Imputation of Missing Values in Gene Expression Microarray Data Using Fuzzy Biclustering Means Particle Swarm Optimization

Sianipar, Nadia Sukei, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=9999920577796&lokasi=lokal>

Abstrak

Microarray data ekspresi gen merupakan data yang berisi hasil proses transformasi genetika dari DNA atau RNA menjadi protein yang disajikan dalam bentuk matriks. Data ekspresi gen yaitu data yang dihasilkan dari eksperimen microarray yang sering kali mengalami missing values akibat keterbatasan teknis dalam proses pengukuran. Permasalahan missing values pada data ekspresi gen memerlukan penanganan yang tepat melalui penerapan metode imputasi guna memperoleh estimasi nilai yang akurat. Pada penelitian ini digunakan metode imputasi dengan penerapan konsep biclustering pada metode Fuzzy Clustering Means (FCM) dengan bahasa pemrograman Python. Hasil imputasi data yang diperoleh menggunakan metode Fuzzy Biclustering Means (FuBiCM) masih memiliki potensi untuk ditingkatkan akurasi. Oleh karena itu, diusulkan algoritma Particle Swarm Optimization (PSO) untuk memperoleh estimasi nilai yang lebih akurat pada titik-titik data yang hilang. Penggabungan metode ini diusulkan sebagai metode Fuzzy Biclustering Means Particle Swarm Optimization (FuBiCMPSO). Percobaan imputasi data diterapkan pada missing rate yang berbeda, yaitu 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, dan 50% serta pada jumlah cluster yang berbeda, yaitu cluster berjumlah 2, 3, 4, dan 5. Selanjutnya, dibandingkan performa metode FCM, FuBiCM, dan FuBiCMPSO dalam mengimputasi missing values berdasarkan nilai NRMSE (Normalized Root Means Square Error) dan Mean Absolute Error (MAE). Berdasarkan hasil dari nilai NRMSE dan MAE, didapatkan bahwa rata-rata dari metode FuBiCMPSO menghasilkan nilai yang lebih rendah dibandingkan dengan metode FCM dan FuBiCM. Oleh karena itu, dapat dikatakan bahwa metode FuBiCMPSO menghasilkan performa yang lebih baik dibandingkan metode FCM dan FuBiCM dalam mengimputasi missing values.

.....Microarray gene expression data represent the results of the genetic transformation process from DNA or RNA into proteins, presented in the form of a matrix. Gene expression data, which are derived from microarray experiments, often contain missing values due to technical limitations during the measurement process. The issue of missing values in gene expression data requires appropriate handling through imputation methods to obtain accurate value estimates. This study employs an imputation approach by applying a biclustering concept to the Fuzzy Clustering Means (FCM) method, implemented using the Python programming language. The imputed results obtained from the Fuzzy Biclustering Means (FuBiCM) method still have room for improvement in terms of accuracy. Therefore, this study proposes the use of the Particle Swarm Optimization (PSO) algorithm to obtain more accurate estimations of the missing data points. This integrated method is proposed as the Fuzzy Biclustering Means Particle Swarm Optimization (FuBiCMPSO) method. Data imputation experiments were conducted under varying missing rates 1%, 5%, 10%, 15%, 20%, 25%, 30%, 35%, 40%, 45%, and 50% and different numbers of clusters 2, 3, 4, and 5. Subsequently, the performance of the FCM, FuBiCM, and FuBiCMPSO methods in imputing missing values was compared based on the Normalized Root Mean Square Error (NRMSE) and Mean Absolute

Error (MAE). Based on the NRMSE and MAE results, the FuBiCMPSO method achieved lower average error values compared to FCM and FuBiCM. Therefore, it can be concluded that the FuBiCMPSO method outperforms both FCM and FuBiCM in imputing missing values.