

Imputasi missing values pada data ekspresi gen pasien penderita kanker menggunakan metode Biclustor-based Bayesian Principal Component Analysis (Bi-BPCA) = Imputation of missing values on cancer patient gene expression data using Biclustor-based Bayesian Principal Component Analysis method (Bi-BPCA)

Bulan Firdanisa, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20515532&lokasi=lokal>

Abstrak

Penelitian bioinformatika sering diterapkan untuk mempelajari penyakit dalam tubuh manusia. Penelitian yang sampai saat ini masih aktif dilakukan ialah penelitian terhadap pasien penderita kanker. Tujuan dari berbagai penelitian ini yaitu untuk menemukan pengobatan terbaik bagi pasien penderita kanker. Salah satu pengobatan yang baru ini muncul dikenal sebagai imunoterapi. Imunoterapi memungkinkan sel-sel imun tubuh kita sendiri digunakan untuk melawan sel-sel kanker. Instrumen utama dalam penelitian terhadap efektifitas imunoterapi juga kasus bioinformatika lainnya ialah data ekspresi gen. Namun, pada data ekspresi gen seringkali ditemukan nilai yang hilang atau missing values yang biasanya disebabkan oleh kerusakan gambar atau kesalahan dalam proses hibridisasi. Keberadaan missing values pada data ekspresi gen dapat menyebabkan kesulitan pada analisis lebih lanjut, di mana banyak analisis ekspresi gen memerlukan data yang lengkap seperti klasifikasi dan pengelompokan. Oleh karena itu, perlu dilakukan imputasi terhadap missing values agar analisis yang dilakukan dapat lebih akurat. Pada penelitian ini dilakukan imputasi menggunakan metode Bi-BPCA. Bi-BPCA merupakan metode imputasi dengan mengombinasikan analisis biclustering dan imputasi BPCA. Metode Bi-BPCA diterapkan pada data ekspresi gen di sekitar kanker setelah dilakukan imunoterapi. Setelah itu, performa dari metode Bi-BPCA dilihat dengan membandingkan hasil imputasi metode Bi-BPCA dengan metode imputasi lainnya diantaranya imputasi menggunakan rata-rata baris, rata-rata kolom, dan metode imputasi BPCA melalui nilai NRMSE. Selain itu, koefisien korelasi Pearson digunakan untuk menghitung korelasi antara nilai hasil imputasi metode Bi-BPCA dengan nilai aslinya. Berdasarkan penelitian ini metode Bi-BPCA menghasilkan NRMSE kurang dari 0.6 untuk missing rate 1-30%, lebih rendah dibandingkan NRMSE dari metode imputasi lainnya. Kemudian, metode Bi-BPCA menghasilkan nilai koefisien korelasi Pearson mayoritas di atas 0.9 mendekati 1. Hasil ini menunjukkan bahwa metode Bi-BPCA menghasilkan nilai imputasi yang lebih baik untuk menggantikan missing values dibandingkan dengan metode imputasi BPCA, rata-rata kolom, dan rata-rata baris.

.....Bioinformatics research is often applied to study diseases in the human body. Research that is still actively being carried out is research on cancer patients. The aim of those studies is to find the best treatment for cancer patients. One treatment that has recently emerged is known as immunotherapy. Immunotherapy allows our body's own immune cells to be used to fight cancer cells. The main instrument in research on the effectiveness of immunotherapy as well as other cases of bioinformatics is gene expression data.. However, in gene expression data, it is often found missing values which are usually caused by image defects and errors in the hybridization process. The existence of missing values in gene expression data can cause difficulties in further analysis, where many analysis of gene expression requires complete data such as classification and clustering. Therefore, it is necessary to impute the missing values so that the analysis can be carried out more accurately. In this study, imputation was carried out using the Bi-BPCA method. Bi-

BPCA is an imputation method by combining biclustering analysis and BPCA imputation. The Bi-BPCA method was applied to gene expression data around cancer after immunotherapy. After that, the performance of the Bi-BPCA method was seen by comparing the imputation results of the Bi-BPCA method with other imputation methods including imputation using row averages, column averages, and the BPCA imputation method through the NRMSE value. In addition, the Pearson correlation coefficient was used to calculate the correlation between the imputed value of the Bi-BPCA method and the original value. Based on this study, the Bi-BPCA method produces NRMSE values less than 0.6 for missing rates 1 to 30 percent, which is lower than NRMSE from other imputation methods. In addition, the Bi-BPCA method produces in a majority Pearson correlation coefficient above 0.9. These results indicate that the Bi-BPCA method produces better imputation values to replace the missing values.