

Penerapan TWO-level hybrid clustering menggunakan algoritma K-means dan algoritma diana pada barisan DNA human papillomavirus (HPV) = The implementation of two level hybrid clustering using k means algorithm and diana algorithm on DNA sequence of human papillomavirus (HPV)

Khoirul Umam, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20446495&lokasi=lokal>

Abstrak

DNA adalah salah satu pembawa informasi genetik pada makhluk hidup. Sequencing dan clustering barisan DNA telah menjadi pekerjaan utama dan rutin dalam dunia biologi molekuler, khususnya dalam bidang terapan bioinformatika. Secara umum metode clustering dapat dibedakan menjadi dua, yaitu hirarki clustering dan partisi clustering. Penelitian ini menggabungkan dua metode clustering yaitu K-Means partisi clustering pada Level 1 dan DIANA hirarki clustering pada Level 2, oleh karena itu disebut Two-Level Hybrid Clustering. Proses awal dimulai dengan mengumpulkan barisan DNA HPV yang diperoleh dari NCBI National Centre for Biotechnology Information, Ekstraksi Ciri, dan Normalisasi. Kemudian melakukan proses clustering menggunakan algoritma K-Means pada Level 1 dan algoritma DIANA pada Level 2. Untuk menghitung jarak genetik antar barisan DNA HPV digunakan persamaan Euclidian Distance. Dan validitas klaster yang digunakan untuk menentukan banyaknya klaster yang optimum adalah Indeks Davies-Bouldin IDB. Hasil penerapan Two-Level Hybrid Clustering pada 1252 barisan DNA HPV adalah data dikelompokkan menjadi 4 klaster dengan nilai IDB yaitu 0.859154564. Semua perhitungan dan proses clustering menggunakan software R.

.....

DNA is one of the carrier of genetic information in living organisms. Sequencing and clustering DNA sequences has become the key and routine activitis in the molecular biology, in particular on bioinformatics applications. There are two type of clustering, hierarchical clustering and partitioning clustering. In this paper, we combine two type clustering processes including K Means partitioning clustering on Level 1 and DIANA hierarchical clustering on Level 2, therefore it called Two Level Hybrid clustering. The beginning of process is started with collecting DNA sequences of HPV from NCBI National Centre for Biotechnology Information, Characteristics Extraction, and Normalization. The next step is clustering by implementation K Means algorithm on Level 1 and DIANA algorithm on Level 2. To calculate the genetic distance we use Euclidian Distance. Moreover, in validating cluster results in order to get optimum number of clusters, we use Davies Bouldin Index DBI. The result of implementation of Two Level Hybrid Clustering on 1252 sequences of HPV is the data clustered into 4 clusters with minimal IDB value is 0.859154564. All calculating and clustering process in this paper using software R.