

Implementasi hybrid clustering menggunakan algoritma fuzzy c-means dan algoritma divisive untuk menganalisis kekerabatan dna human papillomavirus penyebab kanker serviks = The implementation of hybrid clustering using fuzzy c means algorithm and divisive algorithm for analysing dna human papillomavirus cause of cervical cancer

Diyah Septi Andryani, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20446401&lokasi=lokal>

Abstrak

Clustering bertujuan untuk mengklasifikasikan pola yang berbeda ke dalam kelompok yang disebut cluster. Analisis gen dengan menggunakan metode clustering dinilai lebih akurat dibandingkan analisis nukleotida menggunakan penyejajaran DNA. Hybrid clustering pada tesis ini mengkombinasikan algoritma fuzzy c-means dan algoritma divisive mampu meningkatkan keakurasian jika dibandingkan pendekatan pengelompokan partitional tradisional. Algoritma divisive akan dijalankan pada step kedua setelah hasil clustering yang diperoleh dari pengelompokan partisi fuzzy c-means.

Penentuan jumlah cluster terbaik ditentukan dari nilai Indeks Davies Bouldin yang paling minimum. Sebanyak 1252 barisan DNA HPV Human papillomavirus diperoleh dari Genbank NCBI dengan proses melakukan ekstraksi ciri DNA, selanjutnya dilakukan normalisasi. Proses ekstraksi ciri, normalisasi, dan penerapan algoritma partisi fuzzy c-means dan divisive dalam metode hybrid clustering menggunakan bantuan program open source.

Pada hasil hybrid clustering level awal diperoleh jumlah cluster optimum sebanyak 3 cluster dengan nilai Indeks Davies Bouldin paling minimum adalah 0.9715919. Pada level ke-2 clustering didapatkan cluster ke-1 terbagi atas 9 sub cluster dengan nilai IDB minimum adalah 0.8909797. Cluster ke-2 terbagi atas 2 sub cluster dengan nilai IDB minimum adalah 0.7650508. Cluster 3 terbagi atas 2 sub cluster dengan nilai IDB minimum adalah 0.9112528. Nilai IDB pada level kedua selalu lebih kecil dibanding nilai IDB pada level 1. Hal ini mengindikasikan bahwa hybrid clustering memberikan hasil yang lebih baik terhadap hasil clustering.

.....

Clustering aims to classify the different patterns into groups called clusters. Analysis gene by using clustering method is considered more accurate than analysis of nucleotide using DNA alignment. In this thesis, hybrid clustering algorithm which combines fuzzy c means and algorithm divisive will be improve accuracy when compared to partitional clustering. Divisive algorithms will applied on second level after clustering partition using fuzzy c means.

To find the best number of clusters is determined using the minimum value of Davies Bouldin Index DBI of the cluster results. The data is 1252 sequences of HPV DNA sequences obtained from Gen Bank Database in the National Centre for Biotechnology Information NCBI at <http://www.ncbi.nlm.nih.gov> in FASTA format. The data is converted into numerical form through feature extraction using n mers frequency. The results on first level hybrid clustering obtained the optimum cluster divided into three clusters with the value of the minimum Davies Bouldin Index is 0.9715919. Moreover, DBI values after implementing the second step of clustering are always producing smaller IDB values compare to the results of using first step clustering only. This condition indicates that the hybrid approach in this study produce better performance

of the cluster results, in term its DBI values.