

Implementasi hybrid clustering berdasarkan algoritma partitioning around medoids (PAM) dan divisive analysis (Diana) untuk menganalisis kekerabatan DNA-HPV = Implementation of hybrid clustering based on partitioning around medoids (PAM) and divisive analysis (Diana) in human papillomavirus phylogenetic analysis

Mentari Dian Arimbi, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20446399&lokasi=lokal>

Abstrak

Kanker serviks merupakan penyakit kanker yang menyerang daerah kelamin. Kanker serviks ini disebabkan oleh infeksi Human Papillomavirus HPV. Terdapat lebih dari 100 tipe HPV yang sudah teridentifikasi, namun tidak semua tipe memiliki tingkat keganasan yang sama. Clustering dapat dilakukan dengan metode partisi atau metode hirarki. Kedua metode tersebut dapat dikombinasikan dengan metode partisi pada level pertama dan hirarki pada level kedua.

Metode ini dikenal dengan nama hybrid clustering. Fase partisi dapat dilakukan menggunakan PAM, K-means, atau Fuzzy c-means. Kami memilih PAM pada level pertama dan Divisive Analysis DIANA pada level kedua untuk memperoleh hasil cluster yang lebih spesifik. Kami memilih hasil clustering yang meminimalkan nilai IDB. Pada penelitian ini dilakukan clustering pada 1252 barisan DNA HPV yang diperoleh dari GenBank. Hasil dari hybrid clustering diperoleh 15 clusters, dan hasil clustering menunjukkan data dapat dikelompokkan berdasarkan genusnya.

.....

Cervical cancer is a type of cancer that attacks the genital areas which may cause death. Cervical cancer is caused by Human Papillomavirus HPV which attacks uterus cells More than 100 types of HPV have been identified, but not all types have the same malignancy. One way to determine which types of HPV high risk lead to cervical cancer is by clustering or grouping data. Clustering can be achieved through partition or hierarchical method. Both clustering methods can be combined by processing partition algorithm in the first level and hierarchical in the second level.

This method is known as a hybrid clustering. The partition phase can be done by using PAM, K means, or Fuzzy c means methods. We selected Divisive Analysis DIANA algorithm for the second level in order to get more accurate clustering. We choose the clustering results which minimize the DBI value. In this work, we conduct the clustering on 1252 HPV DNA sequences data from GenBank. Results of hybrid clustering obtained 15 clusters, and the result shows that the data can be identified by genus.