

ABSTRAK

Prodi Informatika Universitas Sahid Surakarta selama periode tahun 2004-2021 telah meluluskan 329 alumni. Dengan topik Skripsi dan Tugas Akhir yang beragam. Selama ini Skripsi mahasiswa ditata berdasarkan periode wisuda tanpa memperhatikan topik dari Skripsi dan Tugas Akhir, sehingga mahasiswa sering kesulitan menentukan topik karena keterbatasan pustaka dan dalam mencari daftar pustaka dari penelitian sebelumnya mahasiswa harus memilih satu per satu. Hal ini juga berpengaruh pada data yang digunakan dalam penelitian ini, karena keterbatasan pustaka maka data yang digunakan dalam penelitian ini terbatas pada Skripsi dan Tugas Akhir Prodi Informatika Universitas Sahid Surakarta dengan periode wisuda ke-21 sampai wisuda ke-28 dengan jumlah 143 data.

Penelitian ini bertujuan untuk mengelompokkan Tugas Akhir dan Skripsi Prodi Informatika Universitas Sahid Surakarta. Penelitian ini menggunakan metode CRISP-DM (*Cross-Industry Standard Process Model for Data Mining*) sedangkan Pengelompokan data pada penelitian ini menggunakan gabungan *Hierarchical clustering* dan *K-Means clustering*. Pusat klaster dan jumlah klaster akan ditentukan melalui metode *Hierarchical clustering* kemudian *K-Means clustering* akan mengoptimalkan posisi *centroid* dengan melakukan hitungan berulang pada *centroid* dari tiap klaster. Penghitungan ini akan terus berlangsung hingga nilai *centroid* stabil atau batas iterasi tercapai. Setelah *k-means* mencapai *centroid* yang stabil, maka nilai *centroid* dianggap sudah akurat. Kombinasi keduanya dilakukan agar mendapatkan hasil *clustering* yang lebih baik.

Berdasarkan hasil *clustering* yang dilakukan, terbentuk 10 klaster. 9 klaster yang memiliki topik spesifik dan 1 klaster tidak spesifik. Dari 9 *cluster* dengan topik spesifik ini terdapat topik yang sama yaitu *cluster 0* dan *cluster 8* dengan topik aplikasi android. Klaster ke-4 merupakan klaster yang tidak spesifik. Hasil klaster dihitung kualitasnya menggunakan *intra cluster similarity* dan *inter cluster similarity* dengan rata-rata nilai *inter cluster similarity* 0.210 dan rata-rata nilai *intra cluster similarity* 0.716. Nilai klaster tersebut masuk dalam kategori yang baik, sebagai acuan nilai terbaik dari *inter cluster similarity* adalah 0 dan nilai terbaik dari *intra cluster similarity* adalah 1.

Kata kunci : *Clustering, Hierarchical, K-Means, CRISP-DM, Centroid*

ABSTRACT

The Informatics study program of Sahid Surakarta University has graduated 329 alumnae with various thesis and final project topics during the 2004-2021 period. Students' undergraduate thesis is organized based on the graduation period without considering the undergraduate thesis and final project topic, so students often have difficulty determining the topic due to limited literature. In addition, in searching for a bibliography of previous research, students must select one by one. It also affects the data used in this study. Limited literature limits the data used in this study to the undergraduate thesis and final assignment of the Informatics Study Program of Sahid Surakarta University, with the 21st to 28th graduation period totaling 143 data.

This study will classify the undergraduate thesis and final assignment of the Informatics Study Program at Sahid Surakarta University. This study used the CRISP-DM (Cross-Industry Standard Process Model for Data Mining) method while grouping data combines hierarchical and K-Means clustering. The hierarchical clustering method will determine the cluster center and the number of clusters. K-Means clustering will optimize the position of the centroid by performing repeated calculations on the centroid of each cluster. This calculation will continue until the centroid value is stable or the iteration limit is reached. After k-means reaches a stable centroid, the centroid value is considered accurate. The combination of the two is done to get better clustering results.

Based on the clustering results, 10 clusters were formed. 9 clusters have specific topics and 1 is an unspecific cluster. Based on the 9 clusters with this specific topic, there are the same topics of android applications in cluster 0 and cluster 8. The 4th cluster is a non-specific cluster. Cluster results were calculated for quality using intra-cluster similarity and inter-cluster similarity with an average inter-cluster similarity value of 0.210 and an average intra-cluster similarity value of 0.716. The cluster value is included in the good category. As a reference, the best value of inter-cluster similarity is 0 and intra-cluster similarity is 1.

Keywords: Clustering, Hierarchical, K-Means, CRISP-DM, Centroid

