

# KLASTERISASI DATA MAHASISWA UNDIKSHA MENGGUNAKAN ALGORITMA K-MEANS DAN DBSCAN SERTA ANALISIS PELUANG KELULUSAN MAHASISWA PADA SUATU *CLUSTER*

Oleh

Gusti Ayu Gita Mulya Sari, NIM 1915101011

Jurusan Teknik Informatika

## ABSTRAK

Penilaian dan evaluasi kualitas perguruan tinggi atau program studi melibatkan persentase kelulusan tepat waktu sebagai salah satu indikator penting. Namun, dalam kenyataannya, masih banyak mahasiswa yang gagal menyelesaikan studi dalam waktu yang tepat. Upaya awal untuk mengatasi situasi ini adalah dengan melakukan *clustering* untuk mengelompokkan data mahasiswa Undiksha ke dalam beberapa *cluster*, lalu menganalisis peluang kelulusan, serta untuk mengetahui algoritma terbaik untuk klasterisasi. Pada penelitian ini algoritma *clustering* yang digunakan adalah K-Means dan DBSCAN. Penelitian ini mengkombinasikan metode PCA dengan K-Means dan DBSCAN. Metode PCA digunakan untuk mereduksi dimensi *dataset* sebelum data diklasterisasi.  $k=3$  merupakan jumlah *cluster* optimal untuk K-Means berdasarkan metode *Elbow*. Nilai parameter  $\text{MinPts}=4$  dan  $\text{Eps}=0.02$  merupakan nilai yang paling optimal untuk DBSCAN berdasarkan metode KNN 4-dist plot. Setelah *cluster* terbentuk maka dilakukan interpretasi *cluster* dengan melihat nilai *centroid* atau rata-rata anggota *cluster* dan dilanjutkan dengan melakukan analisis probabilitas dengan menggunakan metode perhitungan peluang sederhana dan *Softmax*. Untuk mengukur validitas *cluster* digunakan metode pengukuran *Silhouette Coefficient*. Dimana didapatkan nilai *Silhouette Coefficient* untuk metode K-Means saat  $k=3$  sebesar 0.769, sedangkan nilai *Silhouette Coefficient* DBSCAN saat nilai  $\text{Eps}=0.02$ ,  $\text{MinPts}=4$  yaitu sebesar 0.926. Sehubungan dengan perbandingan nilai *Silhouette Coefficient*, DBSCAN menghasilkan kualitas *cluster* yang lebih unggul jika dibandingkan dengan K-Means.

**Kata Kunci :** *clustering*, K-Means, DBSCAN, PCA, *silhouette coefficient*

**CLASSTERIZATION OF UNDIKSHA STUDENT DATA USING K-MEANS  
AND DBSCAN ALGORITHM AND ANALYSIS OF STUDENT  
GRADUATION OPPORTUNITIES**

**By**

**Gusti Ayu Gita Mulya Sari, NIM 1915101011**

**Informatic Engineering Major**

**ABSTRACT**

*Assessment and evaluation of the quality of a university or study program involves the percentage of on-time graduation as one of the important indicators. However, in reality, there are still many students who fail to complete their studies in a timely manner. The initial effort to overcome this situation is to do clustering to group Undiksha student data into several clusters, then analyze the chances of graduation, and to find out the best algorithm for clustering. In this research, the clustering algorithms used are K-Means and DBSCAN. This research combines the PCA method with K-Means and DBSCAN algorithm. The PCA method is used to reduce the dimension of the dataset before the data is clustered.  $k=3$  is the optimal number of clusters for K-Means based on the Elbow method. The parameter values  $MinPts=4$  and  $Eps=0.02$  are the most optimal values for DBSCAN based on the KNN 4-dist plot method. After the cluster is formed, cluster interpretation is carried out by looking at the centroid value or the average cluster member and continued by conducting probability analysis using the simple probability calculation method and Softmax. To measure cluster validity, the Silhouette Coefficient measurement method is used. Where the Silhouette Coefficient value for the K-Means method when  $k=3$  is 0.769, while the Silhouette Coefficient value of DBSCAN when the  $Eps$  value= $0.02$ ,  $MinPts=4$  is 0.926. In connection with the comparison of Silhouette Coefficient values, DBSCAN produces superior cluster quality when compared to K-Means.*

**Keywords:** *clustering, K-Means, DBSCAN, PCA, silhouette coefficient*