

ABSTRAK

Meningkatnya jumlah perguruan tinggi dan program studi di Indonesia membuat calon mahasiswa menghadapi tantangan dalam memilih jurusan yang sesuai minat dan prospek karier. Untuk menjawab tantangan tersebut, dilakukan penelitian yang bertujuan mengidentifikasi tingkat kepopuleran semua prodi di Indonesia berdasarkan data dari Sistem Informasi Daya Tampung (SIDATA) PTN tahun 2020 hingga 2024. Metode yang digunakan adalah K-Medoids Clustering, salah satu algoritma clustering non-hierarkis yang efektif dalam mengelompokkan data berdasarkan kemiripan. Data diperoleh melalui proses web scraping dari situs SIDATA PTN menggunakan Python, kemudian diolah untuk menghasilkan rasio antara jumlah peminat dan daya tampung. Penentuan jumlah kluster optimum dilakukan dengan pendekatan metrik evaluasi Silhouette Score, Dunn Index, dan Connectivity, yang seluruhnya menunjukkan bahwa $K = 2$ merupakan pilihan terbaik, dengan skor Silhouette Score sebesar 0.62613, Dunn Index sebesar 0.00351, dan Connectivity sebesar 274.03455. Hasil pengelompokan menghasilkan dua kategori utama, yaitu prodi Populer dan Kurang Populer. Secara keseluruhan, terdapat 865 program studi yang tergolong Populer, dengan rata-rata peminat yang jauh lebih tinggi dibandingkan daya tampung, serta 2166 program studi yang tergolong Kurang Populer di masing-masing kampus. Berdasarkan hasil clustering secara nasional, ditemukan 50 program studi yang tergolong Populer dengan daya tarik yang kuat, dan 387 program studi yang tergolong Kurang Populer. Penelitian ini diharapkan dapat membantu calon mahasiswa dalam mengambil keputusan akademik, serta memberikan informasi strategis bagi institusi pendidikan dan pembuat kebijakan dalam merancang pengembangan prodi ke depan.

Kata Kunci: *Data Mining*, K-Medoids Clustering, SIDATA PTN, Kepopuleran Prodi, *Clustering*

ABSTRACT

The increasing number of universities and study programs in Indonesia poses a challenge for prospective students in selecting a major that aligns with their interests and career prospects. To address this challenge, this study aims to identify the popularity level of all study programs in Indonesia based on data from the Sistem Informasi Daya Tampung (SIDATA) of public universities from 2020 to 2024. The method employed is K-Medoids Clustering, a non-hierarchical clustering algorithm effective in grouping data based on similarity. The data were obtained through a web scraping process from the SIDATA PTN website using Python, then processed to generate the ratio between the number of applicants and the available capacity. The optimal number of clusters was determined using the evaluation metrics Silhouette Score, Dunn Index, and Connectivity, all indicating that $K = 2$ is the best choice, with a Silhouette Score of 0.62613, Dunn Index of 0.00351, and Connectivity of 274.03455. The clustering results produced two main categories: Popular and Less Popular study programs. Overall, 865 study programs were classified as Popular, with the average number of applicants far exceeding the available capacity, while 2166 study programs were classified as Less Popular within each university. On a national scale, the clustering revealed 50 study programs categorized as Popular with strong appeal, and 387 study programs categorized as Less Popular. This research is expected to assist prospective students in making academic decisions and provide strategic insights for educational institutions and policymakers in planning the future development of study programs.

Keywords: *Data Mining, K-Medoids Clustering, SIDATA PTN, Study Program Popularity, Clustering*