

ABSTRAK

Grace Turnip. 231442105. 2024. “Penerapan *Data Mining* Menggunakan Algoritma Pembelajaran Tidak Terawasi (*Unsupervised Learning*) Pada Penerimaan Mahasiswa Baru Universitas Sanata Dharma”. Tesis. Program Studi Magister Pendidikan Matematika, Jurusan Pendidikan Matematika dan Ilmu Pengetahuan Alam, Fakultas Keguruan dan Ilmu Pendidikan, Universitas Sanata Dharma, Yogyakarta.

Penelitian ini bertujuan untuk: (1) menganalisis data penerimaan mahasiswa baru Universitas Sanata Dharma periode 2015–2023 melalui visualisasi data, (2) mengidentifikasi variabel yang paling memengaruhi kelulusan mahasiswa baru, dan (3) mengelompokkan data kelulusan berdasarkan klaster. Metode penelitian yang digunakan adalah deskriptif-kuantitatif dengan objek berupa data penerimaan mahasiswa baru tahun 2015–2023. Penelitian mencakup pengumpulan literatur, pengumpulan data demografis, pembersihan data, serta visualisasi yang dikelompokkan menjadi tiga kategori: keseluruhan, sebelum Covid-19, dan masa Covid-19. Analisis dilakukan menggunakan PCA (Principal Component Analysis) dan klasterisasi berdasarkan nilai rapor dan nilai tes dengan dukungan aplikasi Microsoft Excel, R Studio, dan Jupyter Notebook.

Hasil visualisasi menunjukkan bahwa profil pendaftar periode 2015–2023 tidak menunjukkan perbedaan signifikan antara sebelum dan selama pandemi Covid-19. Sebagian besar pendaftar berasal dari Yogyakarta dan Jawa Tengah, berjenis kelamin perempuan, beragama Katolik, dan lulusan SMA jurusan IPA. Faktor seleksi utama kelulusan, seperti nilai rapor, nilai tes, jalur penerimaan, serta variabel asal mahasiswa (tempat lahir dan asal sekolah), memiliki peran penting dalam menentukan hasil kelulusan.

Klasterisasi data menghasilkan tiga kelompok berdasarkan nilai (rendah, sedang, tinggi). Klaster nilai rendah didominasi oleh mahasiswa laki-laki, sedangkan klaster nilai tinggi didominasi oleh mahasiswa dari luar Pulau Jawa. Sebagian besar mahasiswa dalam semua klaster berasal dari Pulau Jawa, terutama Yogyakarta dan Jawa Tengah, dengan dominasi perempuan pada klaster nilai menengah dan tinggi.

Kata Kunci : Profil mahasiswa, visualisasi data, analisis komponen utama, analisis klaster, dan Covid-19

ABSTRACT

Grace Turnip. 231442105. 2024. “Application of Data Mining Using Unsupervised Learning Algorithm on New Student Admission of Sanata Dharma University”. Thesis. Master of Mathematics Education Study Program, Department of Mathematics and Natural Sciences Education, Faculty of Teacher Training and Education, Sanata Dharma University, Yogyakarta.

This study aims to: (1) analyze the admission data of new students at Universitas Sanata Dharma for the 2015–2023 period through data visualization, (2) identify the variables most influencing new student graduation, and (3) classify graduation data into clusters. The research employs a descriptive-quantitative method with the object of study being new student admission data from 2015 to 2023. The research process includes literature review, demographic data collection, data cleaning, and visualization grouped into three categories: overall, pre-Covid-19, and during Covid-19. Analyses were conducted using Principal Component Analysis (PCA) and clustering based on grade point averages and test scores, supported by Microsoft Excel, R Studio, and Jupyter Notebook.

The data visualization results indicate that the profile of applicants for the 2015–2023 period shows no significant differences between the pre-pandemic and pandemic periods. Most applicants come from Yogyakarta and Central Java, are female, Catholic, and high school graduates, particularly from the science stream. Key selection factors for graduation, such as school grades, test scores, admission pathways, and origin variables (place of birth and school origin), play a significant role in determining graduation outcomes.

Cluster analysis groups the data into three categories based on scores (low, medium, high). The low-score cluster is predominantly male, while the high-score cluster is dominated by students from outside Java. Most students across all clusters come from Java, particularly Yogyakarta and Central Java, with females dominating the medium- and high-score clusters.

Keywords: Student profile, data visualization, principal component analysis, cluster analysis, and Covid-19.