

Analisis K-Means Kelulusan Mahasiswa TI SI Universitas Katolik Santo Thomas Medan 2019-2020

Yosep Dominicke Telaumbanua¹, Muvly Hendrik Paskah Siregar², Rizki Mo'arota Duha³, Sardo Pardingotan Sipayung⁴

Teknik Informatika, Fakultas Ilmu Komputer, Universitas Katolik Santo Thomas Medan
Email: yosepdominicke01@gmail.com; muvlysiregar@gmail.com; rizkiduha0@gmail.com

Published: Januari 2026

ABSTRACT

Kelulusan mahasiswa tepat waktu merupakan salah satu indikator penting dalam penilaian akreditasi program studi dan kualitas pendidikan di perguruan tinggi. Penelitian ini bertujuan untuk mengelompokkan pola kelulusan mahasiswa angkatan 2019 dan 2020 pada Program Studi Teknik Informatika dan Sistem Informasi di Universitas Katolik Santo Thomas Medan guna mengidentifikasi karakteristik akademik mahasiswa. Metode yang digunakan adalah *Data Mining* dengan algoritma *K-Means Clustering* yang memproses 249 data mahasiswa berdasarkan atribut Lama Studi dan Indeks Prestasi Kumulatif (IPK). Proses iterasi *K-Means* menghasilkan tiga *cluster* optimal yang konvergen pada iterasi keempat. Hasil penelitian menunjukkan terbentuknya *Cluster 1* sebanyak 44 mahasiswa dengan karakteristik masa studi lama (± 1834 hari) dan IPK cukup (3,38), *Cluster 2* sebagai kelompok mayoritas sebanyak 201 mahasiswa dengan karakteristik lulus tepat waktu (± 1502 hari) dan IPK tinggi (3,58), serta *Cluster 3* sebanyak 4 mahasiswa yang memiliki masa studi singkat karena putus studi atau meninggal dunia (± 977 hari) namun memiliki IPK baik. Evaluasi kualitas *cluster* menggunakan *Davies-Bouldin Index* (DBI) menghasilkan nilai 0,367, yang menunjukkan bahwa pengelompokan yang terbentuk sangat baik dan valid untuk memetakan performa akademik mahasiswa.

Kata Kunci: *Data Mining, K-Means Clustering, Kelulusan Mahasiswa, Davies-Bouldin Index.*

ABSTRACT

On-time student graduation is an important indicator in assessing study program accreditation and the quality of education in higher education. This study aims to group the student admission patterns of the 2019 and 2020 intakes in the Informatics Engineering and Information Systems Study Program at Santo Thomas Catholic University Medan in order to identify students' academic characteristics. The method used is Data Mining with the K-Means Clustering algorithm which processes 249 student data based on the attributes of Length of Study and Grade Point Average (GPA). The K-Means iteration process produces three optimal clusters that converge in the fourth iteration. The results show the formation of Cluster 1 of 44 students with characteristics of a long study period (± 1834 days) and a sufficient GPA (3.38), Cluster 2 as the majority group of 201 students with characteristics of graduating on time (± 1502 days) and a high GPA (3.58), and Cluster 3 of 4 students who have a short study period due to dropping out or passing away (± 977 days) but have a good GPA. Evaluation of cluster quality using the Davies-Bouldin Index (DBI) produced a value of 0.367, which indicates that the formed grouping is very good and valid for mapping student academic performance.

Keyword: *Data Mining, K-Means Clustering, Student Graduation, Davies-Bouldin Index.*

© The Author(s). 2021 Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0>

PENDAHULUAN

Kelulusan mahasiswa tepat waktu merupakan salah satu indikator strategis dalam penilaian akreditasi program studi serta penentu kualitas mutu Pendidikan di perguruan tinggi. Tingginya persentase mahasiswa yang lulus tepat waktu mencerminkan efektivitas proses pembelajaran, sedangkan banyaknya mahasiswa yang memiliki masa studi panjang atau *drop-out* dapat mempengaruhi

penilaian kinerja institusi secara keseluruhan [1]. Oleh karena itu , pemantauan terhadap pola akademik mahasiswa menjadi krusial agar pihak program studi dapat melakukan intervensi dini terhadap mahasiswa yang berpotensi mengalami kendala akademik.

Di Universitas Katolik Santo Thomas Medan, khususnya pada Program Studi Teknik Informatika dan Sistem Informasi, data akademik mahasiswa angkatan 2019 dan 2020 menunjukkan variasi status kelulusan yang beragam, mulai dari lulus tepat waktu , lulus terlambat , hingga mengundurkan diri. Namun, data tersebut seringkali hanya tersimpan sebagai arsip administratif dan belum diolah secara mendalam untuk menemukan pola karakteristik mahasiswa . Penelitian sebelumnya telah banyak menerapkan teknik *Data Mining* untuk memprediksi kelulusan, di mana algoritma *K-Means Clustering* terbukti efektif dalam mengelompokkan data akademik yang besar berdasarkan kesamaan karakteristik (homogenitas) tanpa memerlukan label kelas sebelumnya [2] [3].

Penelitian ini bertujuan untuk mengimplementasikan algoritma *K-Means Clustering* guna mengelompokkan status kelulusan mahasiswa Rumpun Ilmu Komputer angkatan 2019-2020 berdasarkan atribut Lama Studi dan Indeks Prestasi Kumulatif (IPK). Kebaruan (*novelty*) dari penelitian ini terletak pada pemetaan karakteristik spesifik mahasiswa di lingkungan Universitas Katolik Santo Thomas Medan yang menghasilkan profil klaster akademik (studi cepat, tepat waktu, dan lambat) [4]. Hasil pengelompokan ini diharapkan dapat menjadi landasan bagi pemangku kebijakan di program studi untuk menyusun strategi pendampingan akademik yang lebih tepat sasaran guna meningkatkan angka kelulusan tepat waktu di masa mendatang.

METODE PENELITIAN

Penelitian ini menggunakan pendekatan kuantitatif dengan metode *Knowledge Discovery in Databases* (KDD) untuk menggali pola kelulusan mahasiswa. Teknik utama yang diterapkan adalah *clustering* menggunakan algoritma *K-Means*, yang bertujuan untuk mengelompokkan data mahasiswa berdasarkan kemiripan karakteristik akademik tanpa adanya label kelas sebelumnya (*unsupervised learning*). Metode ini dipilih karena kemampuannya dalam menangani dataset numerik dengan efisiensi komputasi yang baik [9].

2.1. Sumber dan Karakteristik Data

Data penelitian diperoleh dari Biro Administrasi Akademik Universitas Katolik Santo Thomas Medan. Dataset yang digunakan merupakan data akademik mahasiswa Program Studi Teknik Informatika (TI) dan Sistem Informasi (SI) angkatan 2019 dan 2020. Total populasi data yang dikumpulkan sebanyak 249 rekaman data (*record*).

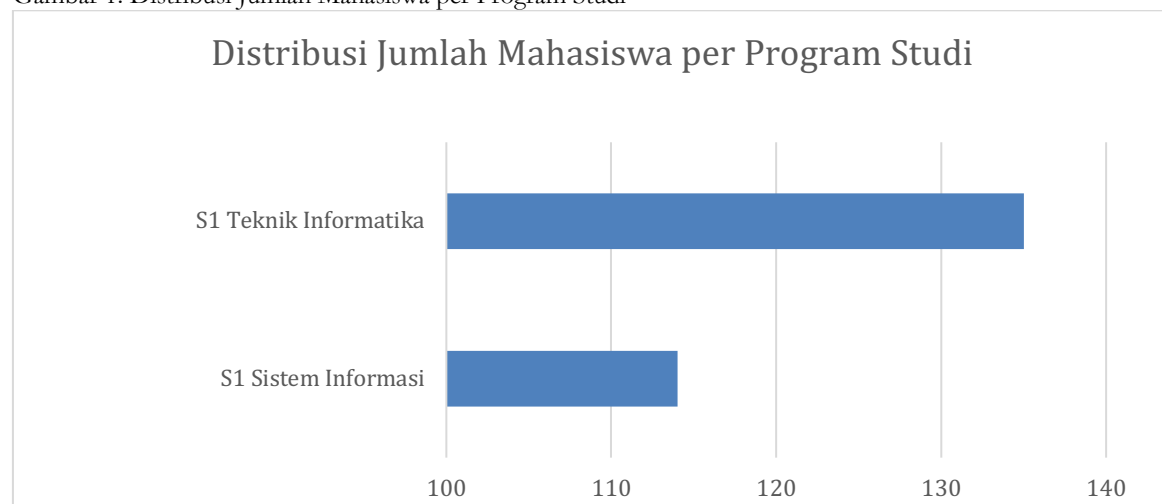
Data tersebut terdiri dari atribut-atribut utama seperti Nomor Induk Mahasiswa (NIM), Nama, Program Studi, Tanggal Masuk, Tanggal Keluar (Lulus/Yudisium), dan Indeks Prestasi Kumulatif (IPK). Sampel data awal sebelum dilakukan pemrosesan dapat dilihat pada Tabel 1.

Tabel 1. Sampel Data Mentah Mahasiswa

Nim	Nama	Prodi	Lama Studi Hari	IPK	Status
190810001	Cahaya Pria Dison Sinurat	S1 Sistem Informasi	1527	3,35	Lulus
190810003	Depi Orpha Harahap	S1 Sistem Informasi	1508	3,71	Lulus
190810005	Efsanjoi Siringoringo	S1 Sistem Informasi	1523	3,74	Lulus
190810006	Desi Yolanda Br Sembiring	S1 Sistem Informasi	1527	3,24	Lulus
190810007	Estermey Giawa	S1 Sistem Informasi	1117	3,44	Lulus
.....					

Distribusi data menunjukkan bahwa sampel didominasi oleh mahasiswa dari Program Studi Teknik Informatika sebanyak 135 orang (54,2%), sedangkan Program Studi Sistem Informasi sebanyak 114 orang (45,8%). Visualisasi sebaran data awal ini ditunjukkan pada Gambar 1.

Gambar 1. Distribusi Jumlah Mahasiswa per Program Studi



2.2. Pra-pemrosesan Data (Preprocessing)

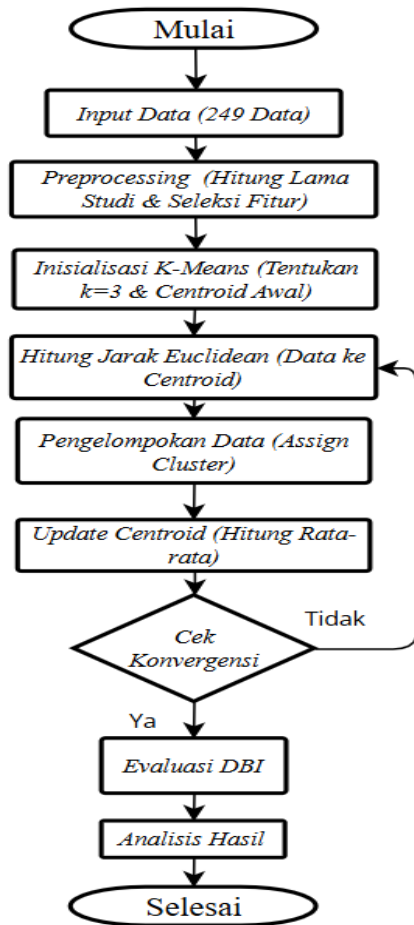
Sebelum dilakukan proses *clustering*, data mentah melalui tahap pra-pemrosesan untuk menjamin kualitas data. Tahapan ini meliputi :

1. Data Cleaning: Memeriksa konsistensi data, menghapus data duplikat, dan memastikan tidak ada nilai kosong (*missing values*) pada atribut penting.
2. Data Transformation: Melakukan konversi tipe data untuk memungkinkan perhitungan matematis. Atribut “Tanggal Masuk” dan “Tanggal Keluar” dikonversi menjadi atribut baru yaitu “Lama Studi (Hari)” dengan menghitung selisih hari antara kedua tanggal tersebut.
3. Feature Selection: Memilih atribut yang relevan untuk dijadikan dimensi dalam *clustering*. Dalam penelitian ini, atribut yang digunakan sebagai input algoritma adalah Lama Studi (Hari) dan IPK. Atribut lain seperti Nama dan Jenis Kelamin diabaikan karena tidak mempengaruhi pola pengelompokan secara langsung.

2.3. Alur Tahapan Penelitian

Untuk memastikan penelitian berjalan sistematis, disusun sebuah kerangka kerja yang menggambarkan urutan proses dari awal hingga akhir. Diagram alir (*flowchart*) tahapan penelitian disajikan pada Gambar 2.

Gambar 2. Metode Penelitian



2.4. Penerapan Algoritma K-Means

Algoritma *K-Means* mempartisi data ke dalam k cluster sedemikian rupa sehingga variansi intra-cluster (jarak data ke pusat cluster) menjadi minimum. Proses penerapan algoritma ini dilakukan dengan langkah-langkah sebagai berikut [5]:

- Penentuan Jumlah Cluster (k)
Penelitian ini menetapkan jumlah cluster sebanyak 3 ($k = 3$), yang merepresentasikan tiga kategori performa studi: Lulus Cepat/Putus Studi, Lulus Tepat Waktu, dan Lulus Terlambat.
- Inisialisasi Centroid Awal
Penentuan titik pusat awal (*initial centroid*) dilakukan dengan metode pemilihan sampel data (*seeds*) yang merepresentasikan karakteristik tiap kategori, bukan secara acak murni. Hal ini bertujuan untuk mempercepat konvergensi dan menghasilkan cluster yang lebih interpretative. Centroid awal diambil dari data ke-183 (Cluster 1), data ke-11 (Cluster 2), dan data ke-159 (Cluster 3).
- Perhitungan Jarak (*Euclidean Distance*)
Jarak antara setiap titik data mahasiswa (x_i) dengan pusat cluster (c_i) dihitung menggunakan rumus *Euclidean Distance* sebagaimana persamaan (1) [7]:

$$d(x, c) = \sqrt{\sum_{i=1}^n (x_i - c_i)^2} \quad (1)$$

Dimana d adalah jarak, x adalah data objek, c adalah centroid, dan n adalah jumlah atribut (dimensi).

d) Iterasi dan Pembaruan Centroid

Setelah setiap data dikelompokkan ke *centroid* terdekat, posisi *centroid* diperbarui dengan menghitung nilai rata-rata (*mean*) dari seluruh anggota *cluster* tersebut. Proses ini diulang terus-menerus hingga posisi centroid tidak lagi berubah (konvergen).

2.5. Pra-pemrosesan Data (Preprocessing)

Untuk memvalidasi kualitas hasil pengelompokan, digunakan metode *Davies-Bouldin Index* (DBI). DBI mengukur rasio antara penyebaran data di dalam cluster (*intra-cluster*) dengan jarak antar cluster (*inter-cluster*). Nilai DBI yang semakin kecil (mendekati 0) menunjukkan struktur *cluster* yang semakin baik (data mengelompok rapat dan terpisah jelas antara kelompok) [6].

PEMBAHASAN

Bagian ini menguraikan tahapan pengolahan data serta analisis hasil pengelompokan (*clustering*) mahasiswa berdasarkan lama studi dan IPK.

3.1. Deskripsi dan Pra-pemrosesan Data (Data Pre-processing)

Berdasarkan tahap pengumpulan data, diperoleh total 249 data mahasiswa dari angkatan 2019 dan 2020. Data tersebut terdiri dari mahasiswa Program Studi Teknik Informatika (54,2%) dan Program Studi Sistem Informasi (45,8%).

Sebelum dilakukan proses *clustering*, dilakukan tahapan pembersihan data (*data cleaning*) untuk memastikan tidak ada data yang kosong (*missing values*) pada atribut utama yang akan diolah, yaitu:

1. Lama Studi (Hari): Durasi total dari tanggal masuk hingga lulus.
2. IPK: Indeks Prestasi Kumulatif terakhir.

3.2. Implementasi K-Means Clustering

Proses pengelompokan dilakukan dengan menetapkan jumlah cluster $K = 3$. Algoritma dimulai dengan penentuan titik pusat awal (*initial centroid*) secara acak, kemudian dilanjutkan dengan perhitungan jarak menggunakan rumus *Euclidean Distance* pada persamaan (2):

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2} \quad (2)$$

Proses iterasi dilakukan untuk memperbarui posisi *centroid* hingga mencapai konvergensi (posisi stabil). Rekapitulasi perubahan nilai *centroid* dari kondisi awal hingga iterasi terakhir disajikan pada Tabel 2.

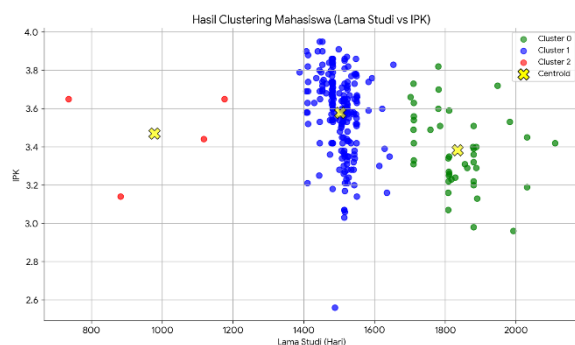
Tabel 2. Perubahan Nilai Centroid (Awal vs Akhir)

Iterasi	Centroid	Lama Studi (Hari)	IPK	Keterangan
Awal	Cluster 1 (Biru)	1653	3,83	Random Initialization
(Iterasi 0)	Cluster 0 (Hijau)	1488	2,56	
	Cluster 2 (Merah)	882	3,14	
Akhir	Cluster 1 (Biru)	1502,43	3,58	Konvergen (Stabil)
(Iterasi 3)	Cluster 0 (Hijau)	1834,43	3,38	
	Cluster 2 (Merah)	977,50	3,47	

3.3. Analisis Hasil Clustering

Berdasarkan hasil iterasi akhir, terbentuk pola pengelompokan mahasiswa yang divisualisasikan dalam bentuk *Scatter Plot* pada Gambar 3.

Gambar 3. Scatter Plot Clustering



Berdasarkan Gambar 3 dan Tabel hasil akhir , karakteristik setiap kelompok dapat dipetakan sebagai berikut Tabel 3:

Tabel 3. Hasil Akhir Pengelompokan Mahasiswa

Kategori Cluster	Jumlah Data	Persentase	Average Lama Studi	Average IPK	Karakteristik
Cluster 1 (Biru)	201	80,7%	± 1502 Hari (4,1 Thn)	3,58	Tepat Waktu (Dominan)
Cluster 0 (Hijau)	44	17,7%	± 1834 Hari (5,0 Thn)	3,38	Terlambat
Cluster 2 (Merah)	4	1,6%	± 977 Hari (2,7 Thn)	3,47	Anomali/ Cepat
Total	249	100%			

3.4. Pembahasan

Bagian ini memberikan interpretasi terhadap hasil pengelompokan untuk menjawab tujuan penelitian.

1. Analisis Kelompok “Tepat Waktu” (Cluster 1 - Biru) kelompok ini mendominasi populasi sebanyak 80,7%. Fakta bahwa rata-rata lama studi hanya 4,1 tahun dengan IPK 3,58 menunjukkan bahwa proses akademik di Program Studi TI dan SI berjalan efektif. mayoritas mahasiswa mampu menyeimbangkan beban studi dengan target waktu kelulusan.
2. Analisis Kelompok “Terlambat” (Cluster 0-Hijau) kelompok ini berisi 44 mahasiswa (17,7%). Temuan menarik di sini adalah rata-rata IPK mereka yang cukup tinggi (3,38). Hal ini mengindikasikan bahwa keterlambatan kelulusan hingga 5 tahun bukan disebabkan oleh ketidakmampuan akademik dalam mata kuliah teori [8]. Hambatan kemungkinan besar terjadi pada fase pengerjaan Skripsi atau Tugas Akhir.
 - Distribusi Prodi: Berdasarkan penelusuran data lebih lanjut, ditemukan bahwa mahasiswa Teknik Informatika (TI) mendominasi kelompok ini (36 orang) dibandingkan Sistem Informasi (8 orang).
 - Rekomendasi: Pengelola Prodi Teknik Informatika perlu mengevaluasi proses bimbingan skripsi, karena mahasiswanya memiliki potensi akademik baik namun terhambat di fase akhir studi.
3. Analisis Kelompok “Anomali” (Cluster 2- Merah) kelompok ini (4 orang) memiliki durasi studi di bawah 3 tahun. Data ini kemungkinan besar merupakan mahasiswa transfer atau pindahan, sehingga dapat dikesampingkan dalam evaluasi kurikulum reguler.

3.5. Evaluasi Validitas Cluster

Untuk memastikan kualitas hasil pengelompokan, dilakukan pengujian menggunakan *Davies-Bouldin Index (DBI)*. Nilai DBI yang kecil (mendekati 0) menunjukkan cluster yang baik. Berdasarkan perhitungan hasil persamaan (3):

$$DBI = \frac{(0,328+0,386+0,386)}{3} = 0,367 \quad (3)$$

Nilai DBI sebesar 0,367 menunjukkan bahwa hasil *clustering* ini valid, dengan pemisahan antar kelompok yang jelas dan tingkat kemiripan anggota dalam satu kelompok yang tinggi.

KESIMPULAN

Penerapan algoritma *K-Means Clustering* pada penelitian ini berhasil memetakan pola kelulusan mahasiswa Angkatan 2019 dan 2020 ke dalam tiga karakteristik utama berdasarkan lama studi dan IPK. Hasil pengelompokan menunjukkan bahwa mayoritas mahasiswa, yaitu sebanyak 201 orang atau 80,7% dari total populasi berada pada *Cluster 1* dengan karakteristik lulus tepat waktu (rata-rata 4,1 tahun) dan memiliki prestasi akademik yang tinggi (rata-rata IPK 3,58). Hal ini mengindikasikan bahwa kurikulum dan proses pembelajaran yang berjalan di program studi saat ini sudah cukup efektif dalam mengantarkan sebagian besar mahasiswa lulus sesuai target waktu yang ditetapkan.

Meskipun demikian, penelitian ini menemukan adanya kelompok mahasiswa pada *Cluster 0* sebanyak 44 orang (17,7%) yang mengalami keterlambatan kelulusan hingga rata-rata 5 tahun. Fakta menarik yang ditemukan adalah kelompok ini memiliki rata-rata IPK yang cukup baik yaitu 3,38, yang menunjukkan bahwa hambatan utama bukan terletak pada kemampuan akademik dalam perkuliahan teori, melainkan pada fase penyelesaian tugas akhir atau skripsi. Berdasarkan sebaran program studi, mahasiswa Teknik Informatika lebih mendominasi kelompok ini dibandingkan Sistem Informasi. Oleh karena itu, diperlukan evaluasi dan pendampingan yang lebih intensif khususnya pada proses bimbingan skripsi bagi mahasiswa Teknik Informatika yang telah memasuki tahun keempat agar tidak mengalami perpanjangan masa studi yang signifikan.

Sebagai saran pengembangan penelitian selanjutnya, disarankan untuk menambahkan atribut lain yang lebih spesifik seperti durasi pengerjaan skripsi, status pekerjaan mahasiswa, atau jalur masuk untuk mendapatkan hasil analisis yang lebih komprehensif [10].

DAFTAR PUSTAKA

- [1] A.P.Windarto, L.S.Dewi, and D.Hartama, "Implementation of K-Means Clustering in Mapping Student Data based on Comorbidity," *Int. J. Inf. Syst. Technol.*, vol. 3, no. 2, pp. 1–10, 2020, doi: 10.30645/ijstech.v4i1.
- [2] J. S. P. Sipayung, "Penerapan Algoritma K-Means Clustering Dalam Pengelompokan Data Mahasiswa Untuk Menentukan Jurusan," *J. Media Inform. Budidarma*, vol. 4, no. 1, pp. 15–22, 2020, doi: 10.30865/mib.v4i1.1557.
- [3] M. R. Lubis and S. P. Sipayung, "Penerapan Data Mining Untuk Mengelompokkan Mahasiswa Berdasarkan Nilai IPK Dengan Metode K-Means Clustering," *J. Tek. Inform. Kaputama (JTik)*, vol. 5, no. 2, pp. 187–194, 2021, doi: 10.59697/jtik.v5i2.
- [4] N. P. A. Indrayani and N. W. W. N. B. W. P. P. A. A. G. S. A. Diatmika, "Analisis Clustering Tingkat Kepuasan Mahasiswa Terhadap Layanan Akademik Menggunakan Algoritma K-Means," *J. Resist. (Rekayasa Sist. Komput.)*, vol. 3, no. 1, pp. 32–43, 2020, doi: 10.31598/jurnalresistor.v3i1.483.

- [5] D. A. O. S. W. N. S. H. S. A. S. R. M. N. H. A. R. T. S. P. Sipayung, “Penerapan Algoritma K-Means Untuk Mengelompokkan Tingkat Kelulusan Mahasiswa,” *J. Sist. Inf. dan Ilmu Komput. Prima (JUSIKOM PRIMA)*, vol. 4, no. 1, pp. 13–19, 2020, doi: 10.34012/jusikom.v4i1.
- [6] E. T. L. Noor, “Penerapan Data Mining Untuk Analisis Pola Masa Studi Mahasiswa Menggunakan Algoritma K-Means,” *J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 3, pp. 543–552, 2021, doi: 10.28932/jutisi.v7i3.3853.
- [7] A. S. Y. H. S. A. S. R. M. N. H. A. R. Nasution, “Penerapan Algoritma K-Means Clustering Untuk Mengelompokkan Mahasiswa Berdasarkan Nilai Akademik,” *J. Sist. Inf. Triguna Dharma (JSI)*, vol. 5, no. 2, pp. 100–108, 2021, doi: 10.53513/jsi.v5i2.
- [8] M. Hardiyanti, R. C. Wihandika, and T. Afrianto, “Penerapan Algoritma K-Means Untuk Clustering Data Mahasiswa Berdasarkan Prestasi Akademik,” *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 3, no. 4, pp. 3363–3368, 2020.
- [9] B. S. Ginting, “Penerapan Data Mining Dalam Pengelompokan Penerima Beasiswa Dengan Metode K-Means,” *J. Sci. Soc. Res.*, vol. 4, no. 2, pp. 112–117, 2021, doi: 10.54314/jssr.v4i2.592.
- [10] S. K. T. Simbolon, “Implementasi Data Mining Dalam Mengelompokkan Data Alumni Menggunakan K-Means Clustering,” *J. Pelita Inform.*, vol. 10, no. 1, pp. 14–19, 2021, doi: 10.31294/jpi.v10i1.