

Analisis Pengelompokan Risiko Keterlambatan Penyelesaian Studi Mahasiswa Menggunakan Algoritma K-Means

Balqis Cipta Cinta Ramadhani¹, Manda Rohandi², Muthia³, Lillyan Hadjaratie⁴, Sri Ayu Ashari⁵, Ihsanulfu'ad Suwandi⁶

^{1,2,3,4,5,6}Pendidikan Teknologi Informasi, Universitas Negeri Gorontalo
email: balqisramadhanicrt59@gmail.com

Abstract

Delayed study completion is one of the issues that can affect the quality of graduates and the effectiveness of higher education institutions. Academic data from students of the Information Technology Education and Information Systems Study Programs at Universitas Negeri Gorontalo indicate that there are still students who have not completed their studies within the expected time frame. Therefore, an analysis of academic data is needed to identify groups of students who are at risk of delayed study completion. This study aims to classify students based on their risk level of delayed study completion using the K-Means Clustering algorithm. The research employed a data mining approach using the Cross Industry Standard Process for Data Mining (CRISP-DM) framework. The dataset consisted of academic records of 1,012 students obtained from the Integrated Academic Information System (SIAT) of Universitas Negeri Gorontalo. The variables used in this study were Grade Point Average (GPA), Semester Grade Point Average (SGPA), and Study Duration. The clustering process was performed using the K-Means algorithm and evaluated using the Davies-Bouldin Index (DBI). The results showed that the optimal number of clusters was obtained at $k=3$ with a DBI value of 0.713. The clustering process produced three groups of students categorized as low, moderate, and high risk. The findings demonstrate that the K-Means algorithm is capable of identifying students' academic patterns and can be utilized as a basis for academic monitoring and support for students at risk of delayed study completion.

Keywords : Clustering; CRISP-DM; Data Mining; K-Means; Study Delay Risk

Abstrak

Keterlambatan penyelesaian studi mahasiswa merupakan salah satu permasalahan yang dapat memengaruhi kualitas lulusan dan efektivitas penyelenggaraan pendidikan tinggi. Data akademik mahasiswa Program Studi Pendidikan Teknologi Informasi dan Sistem Informasi Universitas Negeri Gorontalo menunjukkan masih terdapat mahasiswa yang belum menyelesaikan studi sesuai waktu yang ditetapkan. Oleh karena itu, diperlukan analisis data akademik untuk mengidentifikasi kelompok mahasiswa yang berisiko mengalami keterlambatan penyelesaian studi. Penelitian ini bertujuan mengelompokkan mahasiswa berdasarkan tingkat risiko keterlambatan penyelesaian studi menggunakan algoritma K-Means Clustering. Metode penelitian menggunakan pendekatan data mining dengan kerangka kerja Cross Industry Standard Process for Data Mining (CRISP-DM). Data yang digunakan merupakan data akademik sebanyak 1.012 mahasiswa yang diperoleh dari Sistem Informasi Akademik Terpadu (SIAT) Universitas Negeri Gorontalo dengan variabel Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS), dan Lama Studi. Proses clustering dilakukan menggunakan algoritma K-Means dan dievaluasi menggunakan Davies-Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa jumlah cluster optimal diperoleh pada $k=3$ dengan nilai DBI sebesar 0,713. Hasil clustering menghasilkan tiga kelompok mahasiswa dengan kategori risiko rendah, sedang, dan tinggi. Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengidentifikasi pola akademik mahasiswa sehingga dapat dimanfaatkan sebagai dasar monitoring akademik dan pendampingan bagi mahasiswa yang berisiko mengalami keterlambatan penyelesaian studi.

Kata Kunci : Clustering; CRISP-DM; Data Mining; K-Means; Risiko Keterlambatan Studi

PENDAHULUAN

Pendidikan tinggi memiliki peranan penting dalam menghasilkan sumber daya manusia yang berkualitas. Salah satu indikator keberhasilan perguruan tinggi dapat dilihat dari tingkat ketepatan waktu mahasiswa dalam menyelesaikan studi. Berdasarkan Peraturan Menteri Riset, Teknologi, dan Pendidikan Tinggi Nomor 44 Tahun 2015, masa studi maksimal program sarjana adalah tujuh tahun, sedangkan mahasiswa dikategorikan lulus tepat waktu apabila mampu menyelesaikan studi dalam waktu empat tahun.

Data akademik mahasiswa Program Studi Pendidikan Teknologi Informasi dan Sistem Informasi Fakultas Teknik Universitas Negeri Gorontalo yang bersumber dari Sistem Informasi Akademik (SIAT) tahun 2018–2024 menunjukkan jumlah mahasiswa sebanyak 1.012 mahasiswa. Dari jumlah tersebut, sebanyak 835 mahasiswa (82,5%) masih berstatus aktif, 97 mahasiswa (9,58%) berstatus nonaktif, 66 mahasiswa (6,52%) telah lulus, 11 mahasiswa (1,09%) berstatus cuti, dan 3 mahasiswa (0,30%) berstatus keluar. Selain itu, masih terdapat mahasiswa dari angkatan sebelumnya yang berstatus aktif maupun nonaktif. Kondisi tersebut menunjukkan adanya potensi keterlambatan penyelesaian studi yang dapat berdampak pada kualitas lulusan dan efektivitas proses akademik program studi.

Selain status akademik, capaian prestasi mahasiswa juga menjadi faktor yang memengaruhi Lama Studi. Berdasarkan data yang diperoleh, terdapat 123 mahasiswa (12,15%) yang memiliki Indeks Prestasi Kumulatif (IPK) kurang dari atau sama dengan 2,75 dan 211 mahasiswa (20,85%) yang memiliki Indeks Prestasi Semester (IPS) kurang dari atau sama dengan 2,75. Capaian akademik yang belum optimal tersebut berpotensi memperpanjang masa studi mahasiswa. Hasil wawancara dengan Ketua Jurusan Teknik Informatika Fakultas Teknik Universitas Negeri Gorontalo menunjukkan bahwa keterlambatan penyelesaian studi tidak hanya dipengaruhi oleh rendahnya capaian akademik, tetapi juga oleh keterlambatan mengikuti perkuliahan, pengulangan mata kuliah, serta kendala dalam penyelesaian tugas akhir.

Penelitian Atti dkk. (2021) menunjukkan bahwa semakin tinggi IPK mahasiswa maka semakin singkat Lama Studi yang ditempuh. Berdasarkan temuan tersebut, penelitian ini menggunakan variabel Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS), dan Lama Studi karena ketiga variabel tersebut berkaitan langsung dengan tujuan penelitian, yaitu mengelompokkan mahasiswa berdasarkan tingkat risiko keterlambatan penyelesaian studi. IPK digunakan untuk menggambarkan capaian akademik mahasiswa secara keseluruhan, IPS merepresentasikan performa akademik mahasiswa pada setiap semester, sedangkan Lama Studi menunjukkan durasi penyelesaian studi yang menjadi indikator keterlambatan penyelesaian studi. Meskipun dataset awal memuat atribut lain seperti Nama, NIM, dan Program Studi, pada tahap *Select Attributes* hanya dipilih atribut IPK, IPS, dan Lama Studi karena ketiga atribut tersebut merupakan data numerik yang relevan dalam proses perhitungan jarak pada algoritma K-Means, sedangkan atribut lainnya hanya berfungsi sebagai identitas data sehingga tidak digunakan dalam proses *clustering*. Oleh karena itu, diperlukan analisis terhadap data akademik mahasiswa untuk membantu mengidentifikasi kelompok mahasiswa yang berisiko mengalami keterlambatan penyelesaian studi. Salah satu pendekatan yang dapat digunakan adalah *data mining*. Menurut Rahayu dkk. (2024), *data mining* merupakan proses menemukan pola atau informasi penting dari sekumpulan data berukuran besar menggunakan teknik analisis tertentu. Siregar dan Puspabhuana (2017) menyatakan bahwa *data mining* digunakan untuk menelusuri dan menganalisis data dalam jumlah besar guna menemukan pola tersembunyi yang dapat mendukung pengambilan keputusan, sedangkan Arhami dan Nasir (2020) menjelaskan bahwa *data mining* merupakan proses pencarian pola dan informasi

bermakna dari data berukuran besar yang dikenal sebagai *knowledge discovery*.

Salah satu teknik dalam data mining adalah clustering. Menurut Kurniawan (2025), clustering digunakan untuk mengidentifikasi pola atau struktur alami dari sekumpulan data tanpa memerlukan label tertentu, sedangkan Rahmadden dkk. (2026) menyatakan bahwa clustering bertujuan mengelompokkan objek yang memiliki karakteristik serupa ke dalam satu cluster. Algoritma K-Means merupakan metode clustering non-hierarki yang digunakan untuk membagi data ke dalam beberapa kelompok berdasarkan kedekatan jarak terhadap centroid cluster (Zuhal, 2022). Menurut Rahayu dkk. (2024), K-Means termasuk metode partitioning clustering yang digunakan untuk membantu memahami distribusi data dan karakteristik cluster yang terbentuk.

Berdasarkan permasalahan tersebut, penelitian ini dilakukan untuk mengelompokkan mahasiswa berdasarkan tingkat risiko keterlambatan penyelesaian studi menggunakan algoritma K-Means dengan pendekatan CRISP-DM. Hasil penelitian diharapkan dapat membantu program studi dalam melakukan monitoring akademik, mengidentifikasi mahasiswa yang berisiko mengalami keterlambatan penyelesaian studi, serta memberikan pendampingan akademik yang lebih tepat sasaran.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining menggunakan algoritma K-Means Clustering. Tahapan penelitian menggunakan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*). Menurut Pratama dkk. (2023), CRISP-DM merupakan kerangka kerja data mining yang terdiri dari enam tahapan, yaitu *business understanding*, *data understanding*, *data preparation*, *modeling*, *evaluation*, dan *deployment*. Menurut Suradiradja dan Farizy (2024), metode data mining terdiri atas beberapa pendekatan seperti klasifikasi, clustering, asosiasi, dan prediksi. Pada penelitian ini digunakan pendekatan clustering karena bertujuan mengelompokkan mahasiswa berdasarkan kemiripan karakteristik akademik.

Data penelitian merupakan data sekunder yang diperoleh dari Program Studi Pendidikan Teknologi Informasi dan Program Studi Sistem Informasi Universitas Negeri Gorontalo melalui Sistem Informasi Akademik Terpadu (SIAT). Dataset yang digunakan berjumlah 1.012 data mahasiswa angkatan 2018–2024. Variabel yang digunakan meliputi Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS), dan Lama Studi (Jumlah semester yang ditempuh).

Tahap persiapan data dilakukan melalui proses seleksi atribut dan normalisasi data menggunakan metode *Min-Max Normalization*. Normalisasi dilakukan karena algoritma K-Means sensitif terhadap perbedaan skala data. Oleh karena itu, proses normalisasi diperlukan agar setiap atribut memiliki rentang nilai yang seimbang sehingga proses pengelompokan dapat berjalan secara optimal.

Proses clustering dilakukan menggunakan algoritma K-Means dengan pengujian beberapa nilai jumlah cluster, yaitu $k=2$, $k=3$, dan $k=4$. Menurut Sulistiyawati dan Supriyanto (2021), algoritma K-Means bekerja dengan menentukan centroid awal, menghitung jarak antar data menggunakan Euclidean Distance, kemudian mengelompokkan data berdasarkan centroid terdekat dan memperbarui centroid hingga mencapai kondisi konvergen. Pada penelitian ini, implementasi K-Means dilakukan menggunakan RapidMiner dengan konfigurasi parameter yang sama pada setiap pengujian. Parameter yang digunakan meliputi Maximum Runs sebanyak 10 kali, Euclidean Distance sebagai metode pengukuran jarak, Maximum Optimization Steps sebanyak 100 iterasi, serta inisialisasi centroid menggunakan fitur

Determine Good Start Values dengan mengaktifkan Use Local Random Seed pada nilai 450 untuk menjaga konsistensi hasil clustering.

Evaluasi hasil clustering dilakukan menggunakan Davies-Bouldin Index (DBI). Menurut Sholeh dan Aeni (2023), nilai DBI yang lebih kecil menunjukkan kualitas cluster yang lebih baik karena memiliki tingkat pemisahan cluster yang lebih optimal. Pada penelitian ini, DBI digunakan sebagai metode evaluasi internal dengan membandingkan nilai DBI pada tiga skenario jumlah cluster, yaitu $k=2$, $k=3$, dan $k=4$. Jumlah cluster dengan nilai DBI terendah dipilih sebagai model clustering terbaik.

HASIL DAN PEMBAHASAN

Hasil

Data yang digunakan dalam penelitian ini terdiri dari variabel IPK, IPS, dan Lama Studi (Jumlah Semester yang ditempuh) dengan jumlah data sebanyak 1.012 mahasiswa. Tahap *data preparation* dilakukan menggunakan *RapidMiner* melalui proses *Import Data*, seleksi atribut, dan normalisasi data.

Tabel 1. Hasil Evaluasi Nilai *Davies-Bouldin Index* (DBI)

Jumlah Klaster (k)	Nilai DBI	Keterangan
$k=2$	1,063	-
$k=3$	0,713	Optimal
$k=4$	0,791	-

Berdasarkan hasil evaluasi pada Tabel 1, pengujian dilakukan menggunakan tiga alternatif jumlah cluster yaitu $k=2$, $k=3$, $k=4$. Hasil menunjukkan bahwa nilai *Davies-Bouldin Index* (DBI) terendah diperoleh pada $k=3$ dengan nilai sebesar 0,713. Nilai DBI yang lebih kecil menunjukkan kualitas cluster yang lebih baik karena memiliki tingkat homogenitas yang

tinggi dalam cluster dan tingkat perbedaan yang lebih jelas antar cluster. Oleh karena itu, jumlah cluster sebanyak tiga dipilih sebagai model terbaik dalam penelitian ini karena menghasilkan kualitas pengelompokan yang paling optimal dibandingkan alternatif lainnya.

Tabel 2. Distribusi Anggota Cluster Mahasiswa

Cluster	Profil Risiko	Jumlah Mahasiswa	Persentase
Cluster 1	Risiko Rendah	477	47,1%
Cluster 2	Risiko Sedang	362	35,8%
Cluster 0	Risiko Tinggi	173	17,1%
Total		1.012	100%

Berdasarkan Tabel 2, hasil pengelompokan menghasilkan tiga kelompok mahasiswa dengan jumlah anggota yang berbeda. Cluster 1 memiliki jumlah anggota terbanyak yaitu 477 mahasiswa atau 47,1% dari total data dan dikategorikan sebagai kelompok risiko rendah. Cluster 2 terdiri dari 362 mahasiswa atau 35,8% yang termasuk kelompok risiko sedang.

Sementara itu, Cluster 0 memiliki jumlah anggota paling sedikit yaitu 173 mahasiswa atau 17,1% dan dikategorikan sebagai kelompok risiko tinggi. Distribusi tersebut menunjukkan

bahwa sebagian besar mahasiswa berada pada kelompok dengan risiko keterlambatan studi yang relatif rendah.

Tabel 3. Karakteristik Centroid Final dan Profil Risiko

Cluster	Jumlah Mahasiswa	Karakteristik	Risiko
Cluster 1	477	IPK dan IPS tinggi serta Lama Studi rendah	Risiko Rendah
Cluster 2	362	IPK dan IPS tinggi namun Lama Studi tinggi	Risiko Sedang
Cluster 0	173	IPK dan IPS lebih rendah	Risiko Tinggi

Karakteristik setiap cluster ditentukan berdasarkan nilai centroid akhir yang diperoleh dari proses K-Means Clustering. Cluster 1 memiliki karakteristik berupa nilai IPK dan IPS yang tinggi serta lama studi yang rendah sehingga dikategorikan sebagai kelompok risiko rendah. Cluster 2 memiliki nilai IPK dan IPS yang tinggi, namun menunjukkan lama studi yang lebih tinggi dibandingkan cluster lainnya sehingga dikategorikan sebagai kelompok risiko sedang. Adapun Cluster 0 memiliki nilai IPK dan IPS yang lebih rendah dibandingkan cluster lain sehingga dikategorikan sebagai kelompok risiko tinggi terhadap keterlambatan penyelesaian studi.

Tabel 4. Nilai Centroid

Cluster	Lama Studi	IPK	IPS
Cluster 1	0,129	0,869	0,875
Cluster 2	0,659	0,873	0,851
Cluster 0	0,381	0,624	0,293

Nilai centroid pada setiap cluster menunjukkan karakteristik rata-rata anggota kelompok setelah proses normalisasi data. Pada Cluster 1 diperoleh nilai centroid Lama Studi sebesar 0,129, IPK sebesar 0,869, dan IPS sebesar 0,875. Kombinasi nilai tersebut menunjukkan bahwa mahasiswa dalam cluster ini memiliki capaian akademik yang baik serta lama studi yang relatif rendah. Oleh karena itu, cluster ini dikategorikan sebagai kelompok dengan risiko keterlambatan penyelesaian studi yang rendah.

Cluster 2 memiliki nilai centroid Lama Studi sebesar 0,659, IPK sebesar 0,873, dan IPS sebesar 0,851. Meskipun memiliki nilai IPK dan IPS yang tinggi, cluster ini menunjukkan nilai lama studi yang lebih tinggi dibandingkan cluster lainnya. Karakteristik tersebut menyebabkan cluster ini dikategorikan sebagai kelompok dengan risiko sedang terhadap keterlambatan penyelesaian studi.

Sementara itu, Cluster 0 memiliki nilai centroid Lama Studi sebesar 0,381, IPK sebesar 0,624, dan IPS sebesar 0,293. Nilai IPK dan IPS yang lebih rendah dibandingkan cluster lainnya menunjukkan adanya kecenderungan performa akademik yang kurang optimal. Karakteristik tersebut menyebabkan cluster ini dikategorikan sebagai kelompok mahasiswa dengan risiko keterlambatan penyelesaian studi yang lebih tinggi.

Pembahasan

Hasil penelitian menunjukkan bahwa algoritma K-Means mampu mengelompokkan mahasiswa berdasarkan karakteristik akademik yang dimiliki. Pengelompokan dilakukan menggunakan variabel IPK, IPS, dan Lama Studi (Jumlah Semester yang ditempuh) sehingga

menghasilkan tiga kelompok mahasiswa dengan tingkat risiko keterlambatan studi yang berbeda.

Berdasarkan hasil evaluasi *Davies-Bouldin Index* (DBI), nilai terbaik diperoleh pada $k=3$ dengan nilai sebesar 0,713. Menurut Sholeh dan Aeni (2023), nilai DBI yang semakin kecil menunjukkan kualitas cluster yang semakin baik karena memiliki tingkat pemisahan cluster yang lebih optimal. Oleh karena itu, penggunaan tiga cluster dinilai paling sesuai dalam penelitian ini.

Cluster 1 merupakan kelompok dengan jumlah anggota terbanyak, yaitu 477 mahasiswa atau 47,1% dari total data. Cluster ini memiliki nilai IPK dan IPS yang tinggi serta lama studi yang rendah. Karakteristik tersebut menunjukkan bahwa mahasiswa dalam kelompok ini memiliki performa akademik yang baik dan berpotensi menyelesaikan studi tepat waktu sehingga dikategorikan sebagai kelompok risiko rendah. Temuan ini sejalan dengan penelitian Atti dkk. (2021) yang menyatakan bahwa capaian akademik memiliki hubungan dengan lama studi mahasiswa.

Cluster 2 terdiri dari 362 mahasiswa atau 35,8% dari total data. Cluster ini memiliki nilai IPK dan IPS yang tinggi, tetapi menunjukkan nilai lama studi yang lebih tinggi dibandingkan cluster lainnya. Temuan ini menunjukkan bahwa mahasiswa dengan capaian akademik yang baik belum tentu menyelesaikan studi lebih cepat. Hasil ini mengindikasikan bahwa terdapat faktor lain di luar prestasi akademik yang dapat memengaruhi lama studi mahasiswa. Temuan tersebut mendukung hasil penelitian Tatimu dkk. (2025) yang menunjukkan bahwa keterlambatan penyelesaian studi tidak selalu berkaitan dengan rendahnya prestasi akademik mahasiswa.

Cluster 0 merupakan kelompok dengan jumlah anggota paling sedikit yaitu 173 mahasiswa atau 17,1% dari total data. Cluster ini memiliki nilai IPK dan IPS yang lebih rendah dibandingkan cluster lainnya sehingga menunjukkan kecenderungan risiko keterlambatan studi yang lebih tinggi. Oleh karena itu, kelompok ini memerlukan perhatian dan pendampingan akademik yang lebih intensif agar mampu meningkatkan performa akademik dan menyelesaikan studi tepat waktu.

Menurut Jollyta dkk. (2020), *clustering* bertujuan untuk membentuk kelompok data yang memiliki tingkat kemiripan tinggi dalam satu cluster dan perbedaan yang jelas antar cluster. Hal tersebut terlihat pada hasil penelitian ini, di mana setiap cluster menunjukkan karakteristik yang berbeda berdasarkan nilai IPK, IPS, dan Lama Studi. Perbedaan karakteristik tersebut menunjukkan bahwa algoritma K-Means mampu mengidentifikasi pola akademik mahasiswa dan mengelompokkannya sesuai tingkat risiko keterlambatan studi.

Menurut Rachbini dan Zulkifli (2025), hasil *clustering* dapat dimanfaatkan untuk mendukung pengambilan keputusan dalam institusi pendidikan. Oleh karena itu, hasil penelitian ini diharapkan dapat membantu program studi dalam melakukan monitoring akademik, mengidentifikasi mahasiswa yang berisiko mengalami keterlambatan penyelesaian studi, serta memberikan pendampingan akademik yang lebih tepat sasaran.

SIMPULAN

Penelitian ini menggunakan algoritma K-Means untuk mengelompokkan mahasiswa berdasarkan tingkat risiko keterlambatan penyelesaian studi menggunakan variabel Indeks Prestasi Kumulatif (IPK), Indeks Prestasi Semester (IPS), dan Lama Studi (Jumlah Semester yang ditempuh).

Evaluasi menggunakan Davies-Bouldin Index (DBI) menunjukkan bahwa jumlah cluster

terbaik diperoleh pada $k=3$ dengan nilai DBI sebesar 0,713. Proses Clustering menghasilkan tiga kelompok mahasiswa, yaitu Cluster 1 sebagai kelompok risiko rendah sebanyak 477 mahasiswa atau 47,1%, Cluster 2 sebagai kelompok risiko sedang sebanyak 362 mahasiswa atau 35,8%, dan Cluster 0 sebagai kelompok risiko tinggi sebanyak 173 mahasiswa atau 17,1%. Hasil penelitian menunjukkan bahwa metode K-Means mampu mengidentifikasi pola akademik mahasiswa dan dapat dimanfaatkan sebagai dasar monitoring akademik serta pengambilan keputusan untuk mengurangi risiko keterlambatan penyelesaian studi.

DAFTAR PUSTAKA

- Arhami, M., & Nasir, M. (2020). *Data Mining : Algoritma dan Implementasi*. Penerbit Andi.
- Atti, A., Kleden, M. A., & Lobo, M. (2021). Prediksi Lama Masa Studi Mahasiswa Program Studi Matematika Berdasarkan IPK. In *Jurnal Penelitian Pembelajaran Matematika* (Vol. 14).
- Jollyta, D., Ramdhan, W., & Zarlis, M. (2020). *Konsep Data Mining dan Penerapan*. Deepublish.
- Kurniawan, A. (2025). *Kalkulus Dan Data Mining : Jembatan Teori dan Aplikasi* (Suryaningrat, Ed.). CV Eureka Media Aksara.
- Pratama, S., Sevtian, A., & Putri Anjani, T. (2023). Penerapan Data Mining Untuk Memprediksi Prestasi Akademik Mahasiswa Menggunakan Algoritma C4.5 dengan CRISP-DM. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 7, Number 1). <http://jurnal.polibatam.ac.id/index.php/JAIC>
- Rachbini, W., & Zulkifli. (2025). *Data Science Untuk Pemula dengan RapidMiner*. CV. AA. Rizky.
- Rahayu, N. D., Anshor, A. H., Afriantoro, I., & Halim Anshor, A. (2024). Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means Oleh : Penerapan Data Mining untuk Pemetaan Siswa Berprestasi menggunakan Metode Clustering K-Means. *JUKI : Jurnal Komputer dan Informatika*, 6.
- Rahayu, P. W., Sudipa, I. G. I., Suryani, S., Surachman, A., Ridwan, A., Darmawiguna, I. G. M., Sutoyo, Muh. N., Slamet, I., Harlina, S., & Maysanjaya, I. M. D. (2024). *Buku Ajar Data Mining* (Efitra, Ed.). PT. Sonpedia Publishing Indonesia.
- Rahmadden, Adrianto, S., Muhammad, Hidayatullah, R., & Daulay, S. (2026). *PENGANTAR LENGKAP MACHINE LEARNING : Supervised, Unsupervised, dan Reinforcement Learning Dengan Python* (J. Sumah, Ed.). PT. Serasi Media Teknologi.
- Sholeh, M., & Aeni, K. (2023). *STRING (Satuan Tulisan Riset dan Inovasi Teknologi) Perbandingan Evaluasi Metode Davies Bouldin, Elbow Dan Silhouette Pada Model Clustering Dengan Menggunakan Algoritma K Means*. Diambil <https://archive.ics.uci.edu/ml/datasets/Trave>
- Siregar, A. M., & Puspabhuana, A. (2017). *Data Mining: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV. Kekata Group.
- Sulistiyawati, A., & Supriyanto, E. (2021). *Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan*. 15(2).
- Suradiradja, K. H., & Farizy, S. (2024). *Data Mining: Teori Praktis dan Implementasi Menggunakan WEKA dan Bahasa Pemrograman Java* (U. T. Arsa, Ed.). PT. Adab Indonesia Grup / Penerbit Adab.
- Tatimu, F., Palilingan, V. R., & Rianto, I. (2025). Analisis Klasifikasi Kelulusan Mahasiswa menggunakan Algoritma K-Nearest Neighbor di Fakultas Teknik Universitas Negeri Manado. Dalam *Journal Of Education Method And Technology* (Vol. 3).

Zuhal, N. K. (2022). Study Comparison K-Means Clustering dengan Algoritma Hierarchical Clustering. Dalam *Universitas Nusantara PGRI Kediri. Kediri* (Vol. 1).