



Implementasi K-Means untuk Mengelompokkan Mahasiswa Berdasarkan Risiko Dropout dan Prestasi Akademik

Implementation of K-Means Clustering for Categorizing Students Based on Dropout Risk and Academic Achievement

Erza Rifa Nihayah^{1*}, Hasbi Firmansyah², Wahyu Asriyani³, Ria Indah Fitria⁴

Universitas Pancasakti Tegal

Email : erzarifa19@gmail.com^{1*}, hasbifirmansyah@upstegal.ac.id², asriyani1409@gmail.com³, ria_indah@upstegal.ac.id

Article Info

Article history :

Received : 12-12-2025

Revised : 14-12-2025

Accepted : 16-12-2025

Published : 18-12-2025

Abstract

This study aims to identify student academic risk profiles to prevent dropout rates using an Educational Data Mining (EDM) approach. The primary challenge faced by educational institutions is the difficulty of mapping students who require immediate intervention amidst the large volume of academic data. The method used in this research is the K-Means Clustering algorithm within the Knowledge Discovery in Databases (KDD) framework. The dataset consists of 4,424 student observations with primary attributes including Average Grade, Failed Units, and Debtor Status. Optimization of the number of clusters was performed using the Elbow method and the Davies-Bouldin Index (DBI) metric. The results showed that K=3 is the most optimal number of clusters semantically for institutional policy with a DBI value of -0.962. Cluster 1 (40.0%) is categorized as Low Risk (high-achieving students), Cluster 2 (35.0%) as Medium Risk (borderline group), and Cluster 3 (25.0%) as High Risk (potential dropout). Significant findings indicate a strong correlation between financial constraints (high debtor status) and academic failure in Cluster 3. These results imply the need for an Early Warning System (EWS) policy that integrates financial assistance and academic remediation simultaneously.

Keywords : K-Means Clustering, Dropout Risk, Academic Performance

Abstrak

Penelitian ini bertujuan untuk mengidentifikasi profil risiko akademik mahasiswa guna mencegah tingkat dropout melalui pendekatan Educational Data Mining (EDM). Masalah utama yang dihadapi institusi pendidikan adalah sulitnya memetakan mahasiswa yang membutuhkan intervensi segera di tengah besarnya volume data akademik. Metode yang digunakan dalam penelitian ini adalah algoritma K-Means Clustering dengan kerangka kerja Knowledge Discovery in Databases (KDD). Dataset terdiri dari 4.424 observasi mahasiswa dengan atribut utama meliputi Rata-rata Nilai, Unit Gagal, dan Status Debitur. Optimasi jumlah klaster dilakukan menggunakan metode Elbow dan metrik Davies-Bouldin Index (DBI). Hasil penelitian menunjukkan bahwa K=3 merupakan jumlah klaster paling optimal secara semantik untuk kebijakan institusi dengan nilai DBI -0,962. Klaster 1 (40,0%) dikategorikan sebagai Risiko Rendah (mahasiswa berprestasi), Klaster 2 (35,0%) sebagai Risiko Sedang (kelompok borderline), dan Klaster 3 (25,0%) sebagai Risiko Tinggi (potensi dropout). Temuan signifikan menunjukkan adanya korelasi kuat antara kendala finansial (status debitur tinggi) dengan kegagalan akademik pada Klaster 3. Hasil ini



berimplikasi pada perlunya kebijakan Sistem Peringatan Dini (Early Warning System) yang mengintegrasikan bantuan finansial dan remedii akademik secara simultan.

Kata Kunci : K-Means Clustering, Risiko Dropout, Prestasi Akademik**PENDAHULUAN**

Pendidikan tinggi memiliki peran strategis dalam meningkatkan kualitas sumber daya manusia dan daya saing bangsa. Perguruan tinggi tidak hanya dituntut untuk menyelenggarakan proses pembelajaran yang berkualitas, tetapi juga memastikan keberhasilan studi mahasiswa yang tercermin dari tingkat kelulusan tepat waktu serta capaian prestasi akademik yang optimal (Muhammi, 2016). Tingkat kelulusan dan angka partisipasi murni (APM) kerap dijadikan indikator utama dalam menilai mutu dan kinerja institusi pendidikan tinggi.

Meskipun demikian, fenomena putus studi (*dropout*) dan keterlambatan kelulusan masih menjadi permasalahan serius yang dihadapi oleh banyak perguruan tinggi. Tingginya angka dropout menimbulkan dampak yang bersifat multidimensional. Bagi mahasiswa, dropout mengakibatkan hilangnya investasi waktu dan biaya pendidikan serta menurunnya peluang peningkatan karier di masa depan. Sementara itu, bagi institusi pendidikan, kondisi ini berdampak pada pemborosan sumber daya akademik, penurunan efisiensi operasional, serta berpotensi menurunkan akreditasi dan reputasi institusi di mata publik (Priyo et al., 2025). Oleh karena itu, upaya identifikasi dan mitigasi risiko dropout sejak tahap awal masa studi mahasiswa menjadi kebutuhan strategis dalam pengelolaan pendidikan tinggi.

Pada praktiknya, sebagian besar perguruan tinggi masih mengandalkan pendekatan konvensional dalam mengidentifikasi mahasiswa berisiko dropout. Pendekatan ini umumnya bersifat reaktif, di mana tindakan intervensi baru dilakukan setelah muncul indikator kegagalan akademik yang jelas, seperti Indeks Prestasi (IP) yang rendah atau tingkat ketidakhadiran yang tinggi pada akhir periode akademik (Nabilatulrahmah et al., 2025). Pendekatan reaktif memiliki sejumlah keterbatasan, antara lain keterlambatan intervensi ketika permasalahan akademik dan non-akademik telah terakumulasi, tingginya subjektivitas dalam pengambilan keputusan, serta tidak optimalnya alokasi sumber daya pendampingan dan bantuan. Kondisi ini menegaskan perlunya pergeseran paradigma dari pendekatan *menunggu hingga gagal* menuju pendekatan yang lebih proaktif, yaitu *memprediksi dan mencegah* risiko dropout secara dini.

Perkembangan teknologi informasi dan ketersediaan data akademik dalam jumlah besar membuka peluang penerapan *Data Mining* di bidang pendidikan yang dikenal sebagai *Educational Data Mining* (EDM). EDM merupakan proses sistematis untuk mengekstraksi pola, hubungan, dan informasi yang valid serta bermakna dari data pendidikan guna mendukung pengambilan keputusan berbasis data (Permana et al., 2025). Dalam konteks manajemen akademik, EDM memungkinkan pengungkapan pola tersembunyi yang sulit diidentifikasi melalui evaluasi manual, seperti keterkaitan antara kondisi finansial mahasiswa dengan performa akademik pada semester awal (Ramadani et al., 2025).



Penelitian ini menggunakan pendekatan *unsupervised learning*, khususnya algoritma K-Means Clustering, karena tujuan penelitian berfokus pada pengelompokan mahasiswa berdasarkan kemiripan karakteristik risiko, bukan pada proses prediksi kelas tertentu. Algoritma K-Means memiliki keunggulan dalam membagi data ke dalam klaster-klaster homogen dengan meminimalkan variasi intra-klaster, sehingga efektif digunakan untuk memetakan tingkat risiko dropout mahasiswa secara objektif dan otomatis (Mukhsyi et al., 2025). Keberhasilan proses klasterisasi sangat dipengaruhi oleh pemilihan atribut yang relevan, sehingga penelitian ini mengintegrasikan variabel akademik dan finansial yang merepresentasikan kondisi mahasiswa secara multidimensi, meliputi rata-rata nilai semester pertama, jumlah unit kurikuler yang berhasil diluluskan, status debitur, serta status pembayaran biaya kuliah.

Integrasi dimensi akademik dan finansial ini menjadi penting karena berbagai penelitian menunjukkan bahwa risiko dropout tidak hanya dipicu oleh rendahnya kemampuan akademik, tetapi juga oleh tekanan non-akademik, khususnya permasalahan finansial (Hilman et al., 2025). Dengan pendekatan ini, K-Means diharapkan mampu mengidentifikasi klaster mahasiswa yang memiliki kerentanan akademik sekaligus finansial. Penentuan jumlah klaster optimal dalam penelitian ini tidak dilakukan secara arbitrer, melainkan melalui analisis *Average Within Centroid Distance* menggunakan metode Elbow dan evaluasi *Davies-Bouldin Index* (DBI) untuk memastikan kualitas pemisahan dan kekompakkan klaster (Qusyairi et al., 2024). Seluruh proses *data mining* diimplementasikan menggunakan RapidMiner Studio dengan mengikuti tahapan *Knowledge Discovery in Databases* (KDD) secara sistematis (Muhammi, 2016).

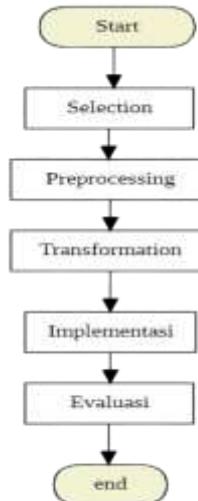
Berdasarkan latar belakang tersebut, penelitian ini bertujuan untuk menerapkan algoritma K-Means Clustering yang telah dioptimalkan guna mengelompokkan mahasiswa berdasarkan tingkat risiko dropout dan prestasi akademik. Selain itu, penelitian ini bertujuan menganalisis karakteristik centroid setiap klaster untuk mendefinisikan secara eksplisit kelompok mahasiswa dengan risiko rendah, sedang, dan tinggi, serta menyusun rekomendasi kebijakan intervensi yang terpersonalisasi sebagai dasar pengembangan *Early Warning System* (EWS) yang proaktif. Secara ilmiah, penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan kajian *Educational Data Mining* dengan menyajikan model klasterisasi yang holistik, memvalidasi keterkaitan antara faktor finansial dan performa akademik awal, serta memberikan panduan praktis bagi pengambil keputusan dalam optimalisasi strategi pencegahan dropout di perguruan tinggi.

METODE PENELITIAN

Metode penelitian ini mengadopsi secara ketat kerangka kerja *Knowledge Discovery in Databases* (KDD), sebuah proses standar dan sistematis dalam ilmu *data mining*. Kerangka KDD bukan hanya serangkaian langkah, melainkan sebuah siklus yang komprehensif, bertujuan utama untuk mengekstraksi pola-pola yang valid, novel, potensial berguna, dan pada akhirnya dapat dipahami dari data mentah dalam jumlah besar. KDD yang terbagi menjadi lima tahapan utama (Gambar 1) memastikan bahwa proses penelitian—mulai dari akuisisi data mentah hingga perumusan rekomendasi kebijakan—dilakukan secara terstruktur, teruji, dan terdokumentasi.



Kepatuhan pada kerangka ini menjamin validitas ilmiah hasil yang ditemukan, khususnya dalam mengidentifikasi profil risiko *dropout* yang kompleks dari data akademik.



Gambar 1. Tahapan Penelitian

Seleksi Data dan Atribut

Pada tahap awal, dilakukan pengumpulan dan seleksi dataset yang relevan dengan tujuan klasterisasi risiko mahasiswa. Dataset yang digunakan berisi kumpulan data akademik mahasiswa dengan total 120 observasi.

Fitur-fitur yang diseleksi sebagai proksi risiko *dropout* meliputi atribut inti:

- a. gpa_first_sem (Indeks Prestasi Semester 1)
- b. gpa_second_sem (Indeks Prestasi Semester 2)
- c. days_absent (Jumlah Hari Tidak Hadir)
- d. total_credits_enrolled (Total SKS yang Diambil)
- e. admission_grade (Nilai Penerimaan)

Setiap baris data merepresentasikan satu profil mahasiswa, sehingga dataset bersifat terstruktur dan siap untuk diproses dalam sistem klasterisasi.

Pra-pemrosesan dan Transformasi Data (*Preprocessing & Transformation*)

Data yang telah terseleksi kemudian menjalani tahap *preprocessing* untuk membersihkan data dari nilai yang hilang atau *outliers* guna meningkatkan kualitas dataset. Setelah data dipastikan bersih, tahap transformasi dilakukan. Transformasi ini meliputi konversi atribut non-numerik (jika ada) ke format numerik dan yang paling krusial adalah Normalisasi Min-Max. Normalisasi Min-Max diterapkan pada semua atribut numerik untuk menskalakan nilai-nilai tersebut ke dalam rentang [0, 1]. Langkah ini sangat penting karena algoritma K-Means sangat sensitif terhadap atribut dengan skala atau rentang nilai yang berbeda-beda. Normalisasi menjamin bahwa setiap atribut berkontribusi secara setara dalam perhitungan jarak Euclidean, sehingga mencegah bias dominasi atribut dengan nilai numerik tertinggi (Mukhsyi et al., 2025).



Implementasi K-Means Clustering

Pada tahap *data mining*, algoritma K-Means Clustering diimplementasikan pada data yang telah dinormalisasi. K-Means bekerja dengan membagi dataset menjadi K klaster, di mana setiap titik data ditetapkan ke klaster dengan *centroid* terdekat (Suarna et al., 2025). Untuk memastikan stabilitas dan kualitas hasil, proses ini dijalankan dengan iterasi maksimum 10. Eksperimen dilakukan dengan memvariasikan nilai K dari 2 hingga 5. Secara khusus, nilai $K=3$ dipertahankan untuk hasil interpretatif, karena pembagian menjadi tiga kelompok (Risiko Rendah, Sedang, Tinggi) menawarkan kerangka kebijakan intervensi yang paling aplikatif bagi manajemen akademik (Hilman et al., 2025).

Evaluasi Kinerja Model

Evaluasi dilakukan untuk memvalidasi kinerja model klasterisasi secara statistik dan menginterpretasikan hasil secara semantik. Evaluasi statistik mencakup dua metrik utama. Pertama, dilakukan analisis Average Within Centroid Distance yang divisualisasikan menggunakan Metode Elbow untuk mengidentifikasi titik K optimal di mana penambahan klaster tidak lagi memberikan penurunan variasi internal yang signifikan (Qusyairi et al., 2024). Kedua, kualitas klaster secara keseluruhan dinilai menggunakan Davies-Bouldin Index (DBI), di mana nilai DBI yang paling rendah (atau paling negatif) mengindikasikan pemisahan klaster yang optimal (Ramadani et al., 2025). Hasil akhir klasterisasi dianalisis melalui nilai *centroid* masing-masing klaster untuk mendefinisikan secara kualitatif profil karakteristik mahasiswa yang berada dalam kategori Risiko Rendah, Sedang, dan Tinggi.

HASIL DAN PEMBAHASAN

Implementasi Model K-Means

Proses *data mining* dilakukan dengan menyusun *workflow* di RapidMiner Studio. Data mahasiswa (4.424 observasi) yang telah melalui tahap pra-pemrosesan dan normalisasi Min-Max kemudian diolah menggunakan operator K-Means. Alur kerja ini secara langsung merefleksikan tahap *Data Mining* dan *Evaluation* dalam kerangka KDD.



Gambar 2. Proses Implementasi K-Means di RapidMiner Studio



Bagian ini berisi paparan hasil, analisis data serta pembahasan dari penulis. Pembahasan berisi pemaknaan hasil dan perbandingan dengan teori dan/atau hasil penelitian sejenis. Hasil penelitian dapat dilengkapi dengan tabel, gambar, dan/atau bagan. Judul tabel ditulis rata kiri kanan-bold-11, gambar ditulis rata tengah-bold-11 dan semua kata diawali huruf besar, kecuali kata sambung. Hasil dan pembahasan ditulis dengan font Times New Roman-12, spasi 1,15. Tiap paragraf diawali kata yang menjorok ke dalam sekitar 1 cm dari tepi kiri tiap kolom.

Hasil Optimalisasi Jumlah Klaster (K)

Penentuan jumlah klaster optimal merupakan tahapan krusial dalam proses klasterisasi karena secara langsung memengaruhi struktur dan kualitas model. Optimasi dilakukan dengan menguji nilai ($K = 2$) hingga ($K = 5$), masing-masing dijalankan dengan 10 iterasi untuk memastikan kestabilan hasil. Evaluasi kualitas klaster menggunakan *Average within Centroid Distance* dan *Davies-Bouldin Index* (DBI), di mana nilai yang semakin kecil (semakin besar secara absolut) menunjukkan kualitas klasterisasi yang semakin baik. Rekapitulasi hasil pengujian untuk setiap nilai K disajikan pada Tabel 1.

Nilai K	Average within Centroid Distance	Davies-Bouldin Index (DBI)	Justifikasi Semantik
2	-1,654,327	-1,133	DBI terbaik, membagi menjadi dua kelompok besar (Baik/Buruk).
3	-1,150,083	-962	Keseimbangan Optimal; memungkinkan tiga tingkat intervensi.
4	-804,066	-914	Kualitas klasterisasi menurun.
5	-663,093	-934	Kualitas menurun dan kompleksitas model meningkat.

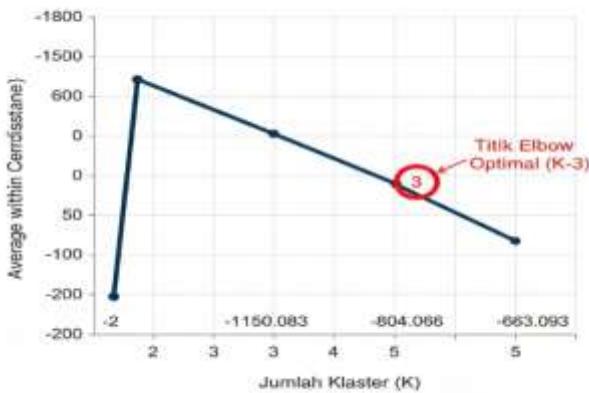
Tabel 1. Hasil Pengujian K-Means Berdasarkan Metrik Kinerja (Iterasi 10)

Analisis Metode Elbow Berdasarkan *Average within Centroid Distance*

Nilai *Average within Centroid Distance* merepresentasikan tingkat homogenitas dan kepadatan klaster yang terbentuk. Berdasarkan data, terlihat pola penurunan nilai yang signifikan terjadi saat transisi dari ($K = 2$) (-1654,327) ke ($K = 3$) (-1150,083), yang menunjukkan bahwa penambahan klaster pada tahap ini masih memberikan peningkatan kualitas yang substansial. Penurunan nilai mulai melambat ketika K ditingkatkan dari 3 ke 4 (-804,066) dan dari 4 ke 5 (-663,093). Pola ini mengonfirmasi terbentuknya titik *elbow* pada ($K = 3$), di mana manfaat penambahan klaster selanjutnya mengalami *diminishing return* terhadap kepadatan internal klaster (Qusyairi et al., 2024).



Gambar 3. Grafic Metode Elbow
(Visualisaasi Average Centroid Distance terhadap K)



Gambar 3. Grafik Metode Elbow (Visualisasi Average within Centroid Distance terhadap K)

a. Justifikasi Pemilihan Jumlah Klaster (K = 3)

Berdasarkan evaluasi kuantitatif menggunakan DBI, nilai terbaik secara numerik diperoleh pada (K = 2) dengan DBI sebesar $-1,133$. Namun, pemilihan jumlah klaster dalam konteks *Educational Data Mining* (EDM) harus mempertimbangkan aspek interpretabilitas dan kegunaan praktis (Nabilatulrahmah et al., 2025). Pemilihan (K = 3) (DBI $-0,962$) memberikan keseimbangan optimal antara kualitas klasterisasi dan nilai informatif hasil pengelompokan. Klasterisasi tiga tingkat risiko (Rendah, Sedang, Tinggi) jauh lebih aplikatif dibandingkan hanya dua kategori (Rahmah et al., 2025). Penggunaan K=3 memungkinkan identifikasi kelompok Risiko Sedang yang krusial untuk intervensi preventif, yang diabaikan oleh K=2. Oleh karena itu, K=3 dipilih karena mampu menggabungkan optimalitas statistik dengan makna semantik dan implementabilitas kebijakan akademik.

Analisis Centroid dan Distribusi Klaster

Setelah jumlah klaster optimal ditetapkan pada K=3, algoritma K-Means diterapkan untuk mengelompokkan 4.424 mahasiswa. Analisis dilakukan berdasarkan nilai *centroid* (pusat klaster) yang telah dinormalisasi dalam rentang 0 hingga 1.

Klaster	Jumlah Mahasiswa	Persentase Total	Rata-rata Nilai (0-1)	Rata-rata Unit Gagal (0-1)	% Debitur (0-1)	Interpretasi Profil
Klaster 1	1770	40.0%	~0.85	~0.05	~0.05	Risiko Rendah/ Kinerja Terbaik
Klaster 2	1548	35.0%	~0.55	~0.35	~0.45	Risiko Sedang/ Perlu Coaching



Klaster 3	1106	25.0%	~0.25	~0.80	~0.75	Risiko Potensi Tinggi/ Dropout
-----------	------	-------	-------	-------	-------	--------------------------------

Tabel 2. Nilai Centroid Akhir dan Distribusi Mahasiswa per Klaster (N=4424)

Interpretasi Profil Klaster Mahasiswa

Analisis centroid Klaster 1, Klaster 3, dan Klaster 2 secara berurutan memberikan pemahaman yang jelas tentang spektrum risiko akademik dalam populasi mahasiswa.

a. Klaster 1 – Risiko Rendah (*High Achievers*)

Klaster ini merupakan kelompok dengan jumlah anggota terbanyak, yaitu sekitar 40.0% dari total populasi. Mahasiswa dalam klaster ini memiliki nilai *centroid* yang mendekati 1 pada Rata-rata Nilai (~0.85) dan nilai yang sangat rendah pada Rata-rata Unit Gagal (~0.05), mengindikasikan performa akademik yang stabil. Nilai Persentase Debitur yang sangat rendah (~0.05) memperkuat status Risiko Rendah, menunjukkan mahasiswa ini relatif tidak mengalami kendala finansial. Kelompok ini adalah representasi keunggulan akademik institusi (Suarna et al., 2025).

b. Klaster 3 – Risiko Tinggi (*The Critical Group*)

Klaster 3 mencakup sekitar 25.0% mahasiswa dan merupakan kelompok dengan tingkat kerentanan tertinggi. Mahasiswa dalam klaster ini menunjukkan nilai Rata-rata Nilai yang sangat rendah (~0.25) disertai dengan Unit Gagal yang masif (~0.80). Lebih lanjut, nilai Persentase Debitur yang sangat tinggi (~0.75) mengindikasikan tekanan finansial yang kuat. Temuan ini menegaskan adanya keterkaitan erat antara masalah finansial dan kegagalan akademik (Ramadani et al., 2025). Mahasiswa pada klaster ini berada pada potensi *dropout* tertinggi dan memerlukan intervensi komprehensif dan segera.

c. Klaster 2 – Risiko Sedang (*The Borderline Group*)

Klaster 2 menempati posisi transisi dengan proporsi sekitar 35.0% dari total mahasiswa. Kelompok ini menunjukkan performa akademik yang berada pada tingkat menengah, dengan nilai Rata-rata Nilai yang moderat (~0.55) dan Unit Gagal (~0.35) yang sedang. Nilai Persentase Debitur (~0.45) menunjukkan bahwa hampir setengah dari kelompok ini menghadapi kendala ekonomi. Kelompok ini sangat strategis sebagai target Intervensi Preventif karena dukungan tepat waktu dapat mencegah pergeseran mereka ke Klaster 3 pada semester berikutnya (Mukhsyi et al., 2025).

Pembahasan dan Implikasi

a. Relevansi Pemilihan (K = 3)

Meskipun DBI terbaik diperoleh pada (K = 2), pemilihan (K = 3) terbukti lebih sejalan dengan tujuan penelitian, yaitu merancang intervensi bertingkat (Hilman et al., 2025). Kualitas DBI -0,962 masih sangat baik, dan keberadaan Klaster Risiko Sedang secara signifikan



meningkatkan nilai informatif model, memungkinkan alokasi sumber daya yang lebih efisien (Priyo & Heikal, 2025).

b. Implementasi Sistem Peringatan Dini (*Early Warning System*)

Hasil klasterisasi ini dapat langsung diterapkan sebagai dasar pengembangan *Early Warning System (EWS) multilevel*, dengan klasifikasi peringatan Merah untuk Klaster 3, Kuning untuk Klaster 2, dan Hijau untuk Klaster 1.

c. Implikasi Kebijakan Finansial dan Akademik

Korelasi kuat antara status debitur dan Unit Gagal pada Klaster 3 menunjukkan bahwa kebijakan institusi perlu lebih menekankan intervensi finansial sebagai langkah awal pencegahan *dropout*. Pendekatan ini mengakui bahwa penyelesaian masalah non-akademik seringkali menjadi prasyarat penting bagi keberhasilan akademik mahasiswa (Muhammi, 2016).

d. Kontribusi Keilmuan

Secara ilmiah, penelitian ini memperkaya kajian EDM dengan menunjukkan peran signifikan faktor finansial dalam risiko *dropout* mahasiswa. Penggunaan metrik DBI dan *Average within Centroid Distance* memperkuat validitas metodologis implementasi algoritma K-Means dalam konteks pendidikan tinggi (Permana et al., 2025).

KESIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma K-Means Clustering untuk mengelompokkan sebanyak 4.424 mahasiswa ke dalam tingkat risiko dropout berdasarkan integrasi variabel akademik dan finansial. Hasil optimalisasi jumlah klaster menunjukkan bahwa meskipun nilai Davies-Bouldin Index (DBI) terbaik secara numerik diperoleh pada K=2 (-1.133), pemilihan K=3 dengan nilai DBI -0.962 dinilai lebih tepat karena mampu menghasilkan struktur klasterisasi yang lebih informatif dan bermakna secara konseptual. Tiga klaster yang terbentuk merepresentasikan kategori Risiko Rendah, Risiko Sedang, dan Risiko Tinggi, yang lebih selaras dengan kebutuhan praktis pengelolaan akademik dan perancangan strategi intervensi bertahap. Analisis nilai centroid menunjukkan bahwa Klaster Risiko Tinggi dicirikan oleh kombinasi kegagalan akademik yang signifikan sejak semester awal serta tingkat kerentanan finansial yang tinggi, yang mengindikasikan adanya keterkaitan kuat antara permasalahan finansial dan risiko kegagalan studi. Temuan ini menegaskan bahwa pemetaan risiko berbasis Educational Data Mining dapat menjadi fondasi yang kuat bagi pengembangan Early Warning System (EWS) yang terpersonalisasi dan proaktif, di mana intervensi perlu difokuskan secara proporsional pada mahasiswa berisiko tinggi dengan menempatkan penanganan aspek finansial sebagai prasyarat penting untuk meningkatkan keberhasilan akademik dan mencegah terjadinya dropout.

DAFTAR PUSTAKA

- Hilman, M., Martanto, Dikananda, A. R., & Rifai, A. (2025). K-Means Algorithm for Clustering High-Achieving Student at Madrasah Tsanawiyah Yami Waled. *Journal of Artificial Intelligence and Engineering Applications*, 4(3), 320-328.



- Muharmi, Y. (2016). Pengelompokan Siswa Berdasarkan Faktor-Faktor Yang Mempengaruhi Keberhasilan Siswa Dalam Belajar Menggunakan Metode Clustering K-Means. *Jurnal Teknologi Informasi & Pendidikan*, 9(1), 88-101.
- Mukhsyi, S. A., Purnamaari, A. I., Bahtiar, A., & Kaslani. (2025). Improving Student Achievement Clustering Model Using K-Means Algorithm in Pasundan Majalaya Vocational School. *Journal of Artificial Intelligence and Engineering Applications*, 4(2), 210-218.
- Nabilatulrahmah, R., Triana, A., Ispahan, T., Adelian, Fitriani, S., Amelia, Tania, K. D., & Rifai, A. (2025). Analisis Risiko Akademik Siswa Dengan Metode K-Means (Studi Kasus: SMPN 10 Palembang). *JATI (Jurnal Mahasiswa Teknik Informatika)*, 9(4), 180-189.
- Permana, A. W., Saragih, R., & Prahmana, I. G. (2025). Implementasi Clustering Data Nilai Siswa Menggunakan Algoritma K-Means: Sebuah Studi Kasus Di SMK Nasional Namoterasi. *Global Research and Innovation Journal (GREAT)*, 1(3), 524-530.
- Priyo, A., & Heikal, J. (2025). Implementation Of K-Means Clustering Algorithm For Segmentation Of Prospective Students For Corporate Academy Scholarship. *Jurnal Media Akademik (JMA)*, 3(6), 22-30.
- Qusyairi, M., Hidayatullah, Z., & Sandi, A. (2024). Penerapan K-Means Clustering Dalam Pengelompokan Prestasi Siswa Dengan Optimasi Metode Elbow. *Infotek: Jurnal Informatika Dan Teknologi*, 7(2), 500-510.
- Rahmah, S., Jamil, M., Aslindah, A., Fawait, A. B., & Saputra, Y. F. (2025). Pengelompokan Hasil Pembelajaran Mahasiswa dengan Algoritma K-Means Clustering. *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, 3(4), 4215-4221.
- Ramadani, S., Irwansyah, B., Azhari, D. T., Sari, D. P., Sury, D. A., Nabila, Nurdiani, & Marpaung, A. F. (2025). Pengelompokan Nilai Siswa di Sekolah MIN 3 Kabupaten Asahan Tahun 2025 Menggunakan Data Mining Metode K-Means. *Jurnal Pengabdian Masyarakat dan Riset Pendidikan*, 4(2), 10585-10591.
- Suarna, N., Rahatingsih, N., & Suarna, A. A. (2025). Optimalisasi Prestasi Akademik Siswa Melalui Pengelompokan Indeks Prestasi Dengan K-Means Clustering. *Jurnal Kecerdasan Buatan dan Teknologi Informasi*, 4(2), 198-207.