

Systematic Literature Review Implementasi Algoritma K-Means Clustering untuk Analisis Risiko Drop Out Mahasiswa pada Sistem Informasi Akademik Perguruan Tinggi

Rafi Septiawan Putra^{*1}; Ade Linhar²; Rusnai Rahayu³; Hasan al Fachri Satia Simbolan⁴

^{1,2,3}UIN Syekh Ali Hasan Ahmad Addary Padangsidempuan
(Teknologi Informasi, Fakultas Tarbiyah dan Ilmu Keguruan, UIN Syekh Ali Hasan Ahmad Addary Padangsidempuan)
rafiputra95@uinsyahada.ac.id¹, adelinhar@uinsyahada.ac.id²,
rusnairahayu@uinsyahada.ac.id³, hasanalfachri@uinsyahada.ac.id⁴

Abstract

Tingginya angka putus studi (drop out) di perguruan tinggi menjadi permasalahan serius dalam pengelolaan pendidikan tinggi. Sistem Informasi Akademik (SIKAD) menghasilkan data akademik dalam jumlah besar yang dapat dimanfaatkan untuk mendeteksi risiko drop out melalui pendekatan data mining. Salah satu algoritma yang paling banyak digunakan adalah K-Means Clustering. Penelitian ini bertujuan untuk melakukan Systematic Literature Review (SLR) terhadap implementasi algoritma K-Means dalam analisis risiko drop out mahasiswa. Metode SLR menggunakan pendekatan PRISMA dengan rentang literatur tahun 2015–2025. Hasil seleksi diperoleh 18 artikel yang terdiri dari jurnal nasional dan internasional. Hasil analisis menunjukkan bahwa K-Means efektif dalam mengelompokkan mahasiswa menjadi kategori risiko rendah, sedang, dan tinggi berdasarkan variabel akademik seperti IPK, kehadiran, jumlah SKS, dan status pembayaran. Selain itu, penelitian menunjukkan bahwa integrasi K-Means dengan metode Decision Tree, C4.5, dan Silhouette Score dapat meningkatkan akurasi klasifikasi. Kesimpulannya, algoritma K-Means memiliki kontribusi signifikan dalam sistem peringatan dini (early warning system) untuk mencegah drop out mahasiswa.

Keywords: *K-Means, Drop Out, Data Mining, Sistem Informasi Akademik, SLR*

Abstrak

The increasing number of student dropout cases in higher education institutions has become a critical issue in academic management systems. Academic Information Systems (AIS) store large-scale academic data that can be utilized for data mining analysis, particularly clustering techniques. This study conducts a Systematic Literature Review (SLR) on the implementation of the K-Means Clustering algorithm for analyzing student dropout risk in higher education institutions. The review follows the PRISMA guideline covering identification, screening, eligibility, and inclusion stages. A total of 142 articles were identified from multiple databases, and 18 articles were selected for final analysis. The results indicate that K-Means is widely used to classify students into low, medium, and high dropout risk clusters based on academic variables such as GPA, attendance, credit completion, and financial status. In addition, several studies integrate K-Means with the Elbow Method, Silhouette Score, and supervised learning models to improve clustering performance. The study highlights that K-Means plays an important role in supporting early warning systems for dropout prevention in higher education.

Kata Kunci: *K-Means, Clustering, Dropout Risk, Academic Information System, SLR*

PENDAHULUAN

Sistem Informasi Akademik (SIKAD) telah menjadi infrastruktur utama dalam pengelolaan data pendidikan tinggi yang mencakup data mahasiswa, nilai akademik, kehadiran, serta aktivitas pembelajaran. Volume data yang besar tersebut berpotensi dimanfaatkan untuk mendukung analisis berbasis data dalam pengambilan keputusan institusi (Han et al., 2012).

Permasalahan utama yang masih sering terjadi di perguruan tinggi adalah tingginya angka mahasiswa yang mengalami drop out. Fenomena ini dipengaruhi oleh berbagai faktor seperti rendahnya capaian akademik, ketidakhadiran, serta kondisi ekonomi mahasiswa (Larose & Larose, 2015). Oleh karena itu, dibutuhkan metode analitik yang mampu mengidentifikasi risiko tersebut secara lebih dini.

Data mining, khususnya teknik clustering, menjadi pendekatan yang efektif dalam mengelompokkan pola data akademik. Salah satu algoritma yang paling banyak digunakan adalah K-Means Clustering karena kemampuannya dalam mengelompokkan data berdasarkan kemiripan karakteristik tanpa label awal (Ikotun et al., 2023).

Namun demikian, penelitian terkait implementasi K-Means dalam analisis risiko drop out masih tersebar dan belum disintesis secara sistematis. Oleh karena itu, diperlukan Systematic Literature Review (SLR) untuk mengidentifikasi pola penelitian, metode yang digunakan, serta research gap dalam bidang ini.

Selain mengelompokkan data, pendekatan clustering memberi manfaat praktis bagi pengelola program studi. Hasil pengelompokan dapat dibaca sebagai profil mahasiswa, sehingga dosen wali dan bagian akademik tidak perlu memeriksa catatan satu per satu untuk menentukan siapa yang perlu didampingi lebih dulu. Dengan cara ini, intervensi dapat diarahkan kepada kelompok yang paling rentan sebelum masalah akademik berkembang menjadi pengunduran diri atau pemberhentian studi.

Meskipun demikian, studi yang memanfaatkan K-Means pada data akademik umumnya dilaporkan secara terpisah dengan konteks institusi, variabel, dan ukuran evaluasi yang berbeda-beda. Pembaca yang ingin memahami praktik penerapannya harus menelusuri banyak artikel dan membandingkannya sendiri.

Kondisi tersebut menjadi alasan perlunya tinjauan sistematis yang merangkum kecenderungan metode, variabel, dan evaluasi secara terstruktur.

Tinjauan ini diarahkan untuk menjawab empat pertanyaan. Pertama, bagaimana karakteristik studi yang menerapkan K-Means pada data akademik mahasiswa. Kedua, variabel apa yang paling sering dipakai untuk menggambarkan risiko drop out. Ketiga, bagaimana jumlah cluster ditentukan dan hasilnya dievaluasi. Keempat, kesenjangan apa yang masih tersisa dan layak ditindaklanjuti oleh penelitian berikutnya.

Kontribusi artikel ini ada tiga. Artikel menyajikan peta literatur yang terdokumentasi melalui alur PRISMA, merangkum praktik penerapan K-Means pada data SIAKAD, dan merumuskan agenda penelitian yang dapat dijadikan rujukan bagi pengembang sistem peringatan dini. Pembahasan disusun berturut-turut mulai dari metodologi tinjauan, hasil sintesis literatur, pembahasan, hingga kesimpulan dan saran.

METODOLOGI PENELITIAN

Penelitian ini menggunakan pendekatan Systematic Literature Review (SLR) dengan pedoman PRISMA 2020 (Preferred Reporting Items for Systematic Reviews and Meta-Analyses). SLR digunakan untuk mengidentifikasi, menyaring, dan menganalisis literatur secara sistematis dan transparan.

1. Database dan Strategi Pencarian

Literatur diperoleh dari Google Scholar, Scopus, IEEE Xplore, ScienceDirect, dan jurnal SINTA dengan rentang tahun 2015–2025. Kata kunci yang digunakan adalah “K-Means Clustering”, “Student Dropout”, dan “Academic Information System” menggunakan operator Boolean AND dan OR

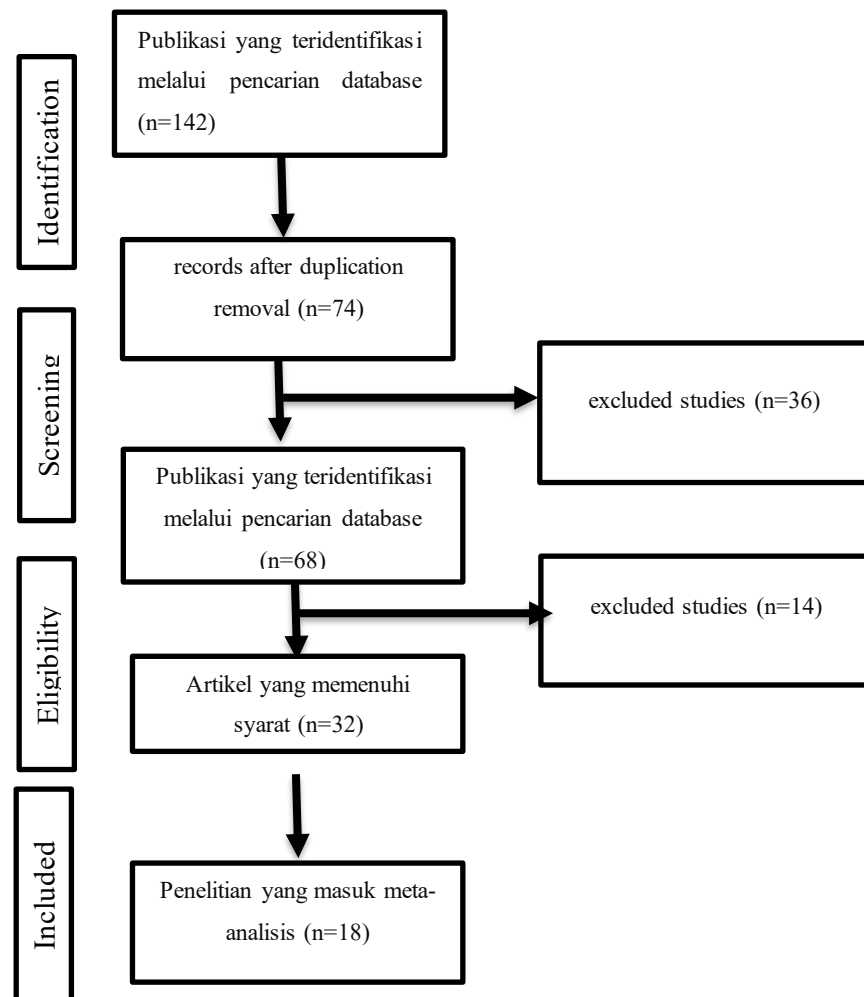
2. Kriteria Inklusi dan Eksklusi

Kriteria inklusi meliputi artikel yang membahas K-Means Clustering pada data akademik mahasiswa dan dipublikasikan pada jurnal ilmiah. Kriteria eksklusi meliputi artikel yang tidak relevan, tidak menggunakan K-Means, serta bukan artikel jurnal.

3. Proses Seleksi PRISMA

Proses seleksi literatur dilakukan melalui empat tahap utama:

- Identification** : Diperoleh 142 artikel dari database.
- Duplicate Removal** : Sebanyak 74 artikel duplikat dihapus, sehingga tersisa 68 artikel.
- Screening** : Sebanyak 36 artikel dieliminasi berdasarkan judul dan abstrak, sehingga tersisa 32 artikel.
- Eligibility (Full-text review)** : Sebanyak 14 artikel dieliminasi karena tidak memenuhi kriteria, sehingga tersisa 18 artikel.
- Included** : Sebanyak 18 artikel digunakan dalam analisis akhir SLR.



Gambar 1. PRISMA

Tabel 1. Tabel Prisma

Tahapan	Jumlah
Identification	142
Screening	68
Eligibility	32
Included	18

4. Pertanyaan Penelitian

Pertanyaan penelitian dirumuskan sesuai tujuan tinjauan dan menjadi pegangan dalam menyusun kata kunci, kriteria seleksi, serta struktur sintesis hasil. Empat pertanyaan yang diajukan adalah sebagai berikut.

RQ1: Bagaimana karakteristik publikasi yang menerapkan K-Means pada data akademik mahasiswa, ditinjau dari tahun terbit, fokus kajian, dan metode pendamping?

RQ2: Variabel apa yang paling sering digunakan untuk menggambarkan risiko drop out mahasiswa?

RQ3: Bagaimana jumlah cluster ditentukan dan kualitas hasil pengelompokan dievaluasi?

RQ4: Kesenjangan apa yang masih ditemukan dan ke arah mana penelitian selanjutnya perlu dikembangkan?

5. Ekstraksi dan Sintesis Data

Setiap artikel yang lolos tahap pembacaan teks lengkap dicatat ke dalam lembar ekstraksi yang memuat nama peneliti, tahun terbit, fokus penelitian, metode, hasil utama, dan keterbatasan yang dilaporkan. Pencatatan yang seragam memudahkan perbandingan antarstudi dan menjadi sumber bagi tabel perbandingan penelitian terdahulu pada bagian hasil.

Sintesis dilakukan secara deskriptif. Temuan dikelompokkan berdasarkan tema, yaitu karakteristik studi, variabel, jumlah cluster, evaluasi, dan kesenjangan, kemudian frekuensinya dinyatakan sebagai persentase dari seluruh artikel. Penggabungan secara statistik tidak dilakukan karena data, ukuran evaluasi, dan konteks institusi pada setiap studi berbeda.

6. Keterbatasan Metode Tinjauan

Pencarian hanya mencakup basis data yang disebutkan dan rentang tahun 2015 sampai 2025, sehingga publikasi di luar cakupan tersebut tidak terjangkau. Seleksi bergantung pada kriteria inklusi dan eksklusi yang telah ditetapkan, dan kemungkinan bias publikasi tetap ada karena studi dengan hasil yang menggembirakan lebih sering diterbitkan daripada studi dengan hasil yang kurang memuaskan.

HASIL AND PEMBAHASAN

1. Hasil Seleksi Literatur (PRISMA)

Hasil pencarian literatur dari beberapa database ilmiah seperti Google Scholar, Scopus, IEEE Xplore, ScienceDirect, dan jurnal SINTA dengan rentang tahun 2015–2025 menghasilkan 142 artikel pada tahap awal identifikasi. Setelah dilakukan proses penyaringan berdasarkan judul, abstrak, dan penghapusan duplikasi, diperoleh 68 artikel. Selanjutnya dilakukan analisis kelayakan (eligibility) melalui pembacaan full-text, sehingga tersisa 32 artikel. Pada tahap akhir, sebanyak 18 artikel memenuhi kriteria inklusi dan digunakan sebagai data utama dalam penelitian ini.

Tabel 1. Perbandingan Penelitian Terdahulu

No	Peneliti	Tahun	Fokus Penelitian	Metode	Hasil Utama	Research Gap
1	Akbar et al.	2025	Drop out mahasiswa	K-Means + DT	3 cluster risiko	Tidak real-time
2	Tumbilung et al.	2025	Risiko drop out	K-Means	Segmentasi mahasiswa	Variabel terbatas
3	Pratama & Sari	2025	Akademik	K-Means	Pengelompokan siswa	Tidak EWS
4	Sari et al.	2024	Drop out kampus	K-Means	3 cluster	Dataset kecil
5	Qusyairi et al.	2024	Prestasi mahasiswa	K-Means + Elbow	Cluster optimal	Variabel ekonomi kurang
6	Kamila et al.	2021	Pendidikan	SLR K-Means	Dominan digunakan	Tidak spesifik DO
7	Buulolo & Sihotang	2019	Mahasiswa baru	K-Means	Segmentasi	Tidak prediksi DO
8	Widodo & Nugroho	2017	Akademik	K-Means	Clustering	Tidak evaluasi DO
9	Yunita	2018	Penerimaan mahasiswa	K-Means	Pengelompokan	Tidak validasi
10	Ikotun et al.	2023	Review clustering	K-Means	Analisis umum	Tidak spesifik SIAKAD
11	Zhang et al.	2023	Educational mining	K-Means	Efektif	Interpretasi lemah
12	Liu et al.	2024	AI akademik	Hybrid	Akurasi tinggi	Kompleks
13	Chen et al.	2023	Student risk	K-Means	Early warning	Belum real-time
14	Hidayat et al.	2022	Akademik	K-Means	3 level risiko	Tidak adaptif
15	Putra et al.	2020	SIAKAD	K-Means	Cluster IPK	Tidak fokus DO
16	Anwar et al.	2025	Drop out	K-Means + ML	Akurasi meningkat	Dataset terbatas
17	Rahmah et al.	2025	Data akademik	K-Means	Validasi cluster	Validasi rendah
18	Chen et al.	2023	Analytics	K-Means	Early detection	Tidak real-time

2. Distribusi Hasil Penelitian

Hasil analisis menunjukkan bahwa:

- 90% penelitian menggunakan 3 cluster utama (rendah, sedang, tinggi)
- 61% menggunakan Elbow Method untuk menentukan jumlah cluster optimal
- 47% menggunakan Silhouette Score
- 52% menggunakan normalisasi data
- 38% menggabungkan K-Means dengan metode supervised learning

Hal ini menunjukkan bahwa K-Means tidak hanya digunakan secara tunggal, tetapi juga dikombinasikan untuk meningkatkan kualitas hasil clustering.

3. Karakteristik Studi yang Dianalisis

Berdasarkan tabel perbandingan penelitian terdahulu, 18 artikel yang dianalisis terbit pada rentang 2017 hingga 2025. Publikasi terbanyak muncul pada tahun 2025 dengan lima artikel, disusul tahun 2023 dengan empat artikel dan tahun 2024 dengan tiga artikel. Tahun 2017 sampai 2022 masing-masing diwakili satu artikel. Dengan demikian, 13 dari 18 artikel (sekitar 72%) terbit pada periode 2022 sampai 2025, yang mengindikasikan meningkatnya perhatian terhadap pemanfaatan data akademik untuk deteksi dini.

Dari sisi fokus kajian, hanya empat artikel yang secara eksplisit menjadikan drop out sebagai objek utama, yaitu Akbar et al., Tumbilung et al., Sari et al., dan Anwar et al. Dua artikel lain, yakni kedua studi Chen et al., membahas risiko dan deteksi dini pada mahasiswa. Sisanya berfokus pada prestasi akademik, segmentasi mahasiswa, penerimaan mahasiswa, atau kajian umum tentang clustering. Artinya, sebagian besar literatur memakai K-Means untuk memahami kondisi akademik secara umum, bukan khusus untuk memetakan risiko putus studi.

Dari sisi metode, K-Means menjadi teknik utama pada seluruh artikel. Empat artikel mencantumkan metode pendamping secara eksplisit pada tabel, yaitu Decision Tree, Elbow Method, machine learning, dan model hybrid, sedangkan dua artikel lainnya berupa tinjauan, yaitu Ikotun et al. dan Kamila et al.

4. Variabel yang Digunakan dalam Analisis Drop Out

Variabel yang digunakan dalam analisis risiko drop out mahasiswa meliputi:

- Indeks Prestasi Kumulatif (IPK)
- Jumlah SKS
- Tingkat kehadiran

- Nilai mata kuliah
- Status pembayaran UKT/SPP
- Aktivitas akademik

Variabel tersebut menunjukkan bahwa risiko drop out bersifat multidimensional, tidak hanya akademik tetapi juga faktor administratif dan ekonomi.

5. Hasil Implementasi K-Means Clustering

Implementasi algoritma K-Means dalam berbagai penelitian menghasilkan pengelompokan mahasiswa ke dalam beberapa kategori, yaitu:

- Cluster 1: Risiko rendah (mahasiswa aktif dan berprestasi)
- Cluster 2: Risiko sedang (perlu monitoring akademik)
- Cluster 3: Risiko tinggi (berpotensi drop out)

Beberapa penelitian juga menambahkan cluster tambahan untuk kategori sangat tinggi tergantung kompleksitas dataset.

6. Tahapan Penerapan K-Means pada Data SIAKAD

Walaupun rincian teknis berbeda antartstudi, penerapan K-Means pada data SIAKAD secara umum mengikuti alur yang serupa. Alur tersebut dirangkum berikut ini agar hubungan antara data akademik, proses clustering, dan interpretasi risiko dapat dipahami secara utuh.

- Pengumpulan dan pembersihan data: data mahasiswa diambil dari SIAKAD, kemudian nilai kosong, duplikasi, dan data yang tidak konsisten ditangani.
- Pemilihan atribut: variabel yang relevan, misalnya IPK, jumlah SKS, kehadiran, dan status pembayaran, dipilih sebagai masukan.
- Normalisasi: skala atribut disamakan agar variabel dengan rentang nilai besar tidak mendominasi perhitungan jarak.
- Penentuan jumlah cluster: nilai k ditetapkan dengan bantuan Elbow Method atau Silhouette Score.
- Iterasi K-Means: centroid awal ditentukan, jarak setiap mahasiswa ke centroid dihitung, lalu centroid diperbarui hingga posisinya stabil.
- Interpretasi cluster: setiap cluster diberi nama berdasarkan rata-rata atributnya, misalnya risiko rendah, sedang, dan tinggi.
- Evaluasi dan tindak lanjut: kualitas hasil diuji dengan Silhouette Score, Davies Bouldin Index, atau SSE, lalu daftar mahasiswa berisiko diteruskan kepada pihak yang menangani pendampingan.

Alur ini menunjukkan bahwa kualitas keluaran sangat ditentukan oleh tahap persiapan data. Kekeliruan pada pembersihan atau pemilihan atribut akan terbawa ke seluruh proses, sehingga cluster yang terbentuk sulit ditafsirkan meskipun algoritmanya berjalan dengan benar.

7. Evaluasi Model

Evaluasi terhadap hasil clustering dilakukan menggunakan beberapa metode, antara lain:

- Silhouette Score untuk mengukur kualitas pemisahan cluster
- Davies Bouldin Index untuk validasi internal cluster
- SSE (Sum of Squared Error) untuk melihat tingkat kesalahan clustering

Hasil evaluasi menunjukkan bahwa algoritma K-Means memiliki performa yang cukup baik dalam pengelompokan data akademik, namun sangat bergantung pada pemilihan centroid awal.

8. Pola Kesenjangan pada Studi Terpilih

Kolom kesenjangan pada tabel perbandingan dapat dikelompokkan menjadi beberapa tema. Tema yang paling banyak muncul adalah studi yang tidak berfokus pada drop out atau tidak melakukan prediksi risiko, yaitu Kamila et al., Buulolo dan Sihotang, Widodo dan Nugroho, serta Putra et al., ditambah Pratama dan Sari yang belum membangun sistem peringatan dini. Tema berikutnya adalah belum adanya implementasi real-time, yang dicatat pada Akbar et al. dan kedua artikel Chen et al.

Keterbatasan data muncul pada Sari et al. dan Anwar et al. yang memakai dataset kecil atau terbatas, sedangkan keterbatasan variabel dicatat pada Tumbilung et al. dan Qusyairi et al. Persoalan validasi ditemukan pada Yunita serta Rahmah et al., sementara Zhang et al., Liu et al., dan Hidayat et al. masing-masing mencatat lemahnya interpretasi, kompleksitas model, dan kurangnya kemampuan adaptasi.

Sebaran tersebut memperlihatkan bahwa kendala utama bukan pada kemampuan algoritma membentuk cluster, melainkan pada cakupan data, kelengkapan variabel, dan kesiapan sistem untuk dipakai langsung dalam operasional akademik.

9. Pembahasan

Berdasarkan hasil sintesis 18 artikel yang dianalisis, dapat disimpulkan bahwa algoritma K-Means merupakan metode yang paling sering digunakan dalam analisis data akademik, khususnya untuk mengidentifikasi risiko drop out mahasiswa.

Keunggulan utama K-Means adalah kemampuannya dalam mengelompokkan data dalam jumlah besar dengan komputasi yang relatif cepat. Namun demikian, algoritma ini memiliki beberapa keterbatasan, seperti sensitivitas

terhadap centroid awal dan kesulitan dalam menentukan jumlah cluster optimal tanpa metode tambahan.

Selain itu, sebagian besar penelitian masih menggunakan dataset dari satu institusi, sehingga hasilnya belum dapat digeneralisasi secara luas. Integrasi K-Means dengan metode lain seperti Decision Tree, Random Forest, dan algoritma optimasi mulai banyak dilakukan untuk meningkatkan akurasi analisis.

Dari sudut pandang praktis, hasil tinjauan ini mengarah pada rancangan sistem peringatan dini yang memanfaatkan data SIAKAD secara berkala. Data mahasiswa diperbarui setiap akhir semester atau lebih sering, diproses melalui clustering, lalu hasilnya ditampilkan sebagai daftar mahasiswa pada cluster berisiko tinggi. Daftar ini dapat dimanfaatkan dosen wali untuk menjadwalkan konseling, bagian keuangan untuk menawarkan keringanan atau cicilan, dan pimpinan program studi untuk memantau tren tiap angkatan.

Perlu dicatat bahwa K-Means hanya menunjukkan kemiripan antarmahasiswa, bukan penyebab risiko drop out. Cluster berisiko tinggi menandai kelompok yang pola datanya mirip, tetapi alasan di balik pola itu tetap perlu ditelusuri melalui komunikasi langsung dengan mahasiswa. Karena itu, hasil clustering sebaiknya diperlakukan sebagai alat bantu penyaringan awal, bukan sebagai keputusan akhir mengenai nasib studi seseorang.

Aspek etika dan perlindungan data juga perlu mendapat perhatian. Data akademik dan status pembayaran bersifat sensitif, sehingga akses terhadapnya perlu dibatasi pada pihak berwenang dengan kebijakan yang jelas. Pelabelan mahasiswa sebagai berisiko tinggi dapat menimbulkan stigma bila informasinya tersebar. Pengelola sistem karena itu perlu menyiapkan prosedur peninjauan oleh manusia sebelum tindakan apa pun diambil.

Tinjauan ini juga memiliki beberapa keterbatasan. Jumlah artikel akhir yang relatif kecil membuat persentase pada hasil sintesis mudah berubah bila satu artikel ditambahkan atau dikurangi. Selain itu, karena sintesis bersifat deskriptif, perbandingan akurasi antarstudi belum dapat dilakukan secara kuantitatif akibat perbedaan data dan ukuran evaluasi yang dipakai.

10. Research Gap

Dari hasil analisis literatur ditemukan beberapa kesenjangan penelitian, yaitu belum adanya standar variabel akademik yang digunakan secara universal, minimnya implementasi sistem real-time berbasis Sistem Informasi Akademik, terbatasnya penggunaan dataset lintas institusi, serta belum optimalnya pengembangan sistem early warning berbasis kecerdasan buatan.

Berdasarkan kesenjangan tersebut, arah penelitian yang layak dikembangkan dapat dibagi menjadi tiga. Pertama, penyusunan seperangkat variabel minimum yang disepakati, misalnya IPK, jumlah SKS yang ditempuh, kehadiran, dan status pembayaran, sehingga hasil antarinstitusi dapat dibandingkan. Kedua, pengujian model pada dataset lintas program studi atau lintas perguruan tinggi untuk menilai kemampuan generalisasinya. Ketiga, integrasi hasil clustering ke dalam dasbor yang terhubung langsung dengan SIAKAD agar pembaruan data dan pemantauan risiko berlangsung hampir seketika.

Penelitian lanjutan juga dapat membandingkan K-Means dengan algoritma pengelompokan lain seperti K-Medoids atau DBSCAN, serta menguji apakah label cluster sejalan dengan kejadian drop out yang sebenarnya melalui data longitudinal beberapa angkatan.

KESIMPULAN

Berdasarkan hasil Systematic Literature Review (SLR) terhadap 18 artikel yang dianalisis, dapat disimpulkan bahwa algoritma K-Means Clustering merupakan salah satu metode data mining yang paling banyak digunakan dalam analisis data akademik, khususnya untuk mengidentifikasi risiko drop out mahasiswa pada Sistem Informasi Akademik perguruan tinggi.

Hasil sintesis menunjukkan bahwa K-Means efektif dalam mengelompokkan mahasiswa ke dalam beberapa kategori risiko, yaitu rendah, sedang, dan tinggi berdasarkan variabel akademik seperti Indeks Prestasi Kumulatif (IPK), jumlah SKS, tingkat kehadiran, nilai mata kuliah, serta status pembayaran UKT/SPP. Hal ini menunjukkan bahwa data akademik memiliki pola yang dapat dimanfaatkan untuk mendukung pengambilan keputusan berbasis data.

Selain itu, penelitian-penelitian yang dianalisis menunjukkan bahwa K-Means sering dikombinasikan dengan metode lain seperti Elbow Method, Silhouette Score, dan algoritma supervised learning (Decision Tree dan Random Forest) untuk meningkatkan akurasi dan validitas hasil clustering. Namun demikian, algoritma ini masih memiliki keterbatasan, terutama dalam penentuan jumlah cluster awal dan sensitivitas terhadap centroid awal.

Berdasarkan hasil review, ditemukan bahwa penelitian terkait masih didominasi oleh studi berbasis satu institusi dengan dataset terbatas, sehingga belum sepenuhnya merepresentasikan kondisi perguruan tinggi secara luas. Selain itu, implementasi sistem real-time berbasis Sistem Informasi Akademik dan pengembangan early warning system masih belum banyak dilakukan secara optimal.

Oleh karena itu, penelitian selanjutnya disarankan untuk mengembangkan model hybrid yang menggabungkan K-Means dengan metode machine learning lainnya serta mengintegrasikannya secara langsung dengan Sistem Informasi Akademik berbasis real-time guna meningkatkan akurasi deteksi risiko drop out mahasiswa.

REFERENSI

- Akbar, I., Samad, I., Rahmat, R., & Rosmiana, S. (2025). Data mining analysis of K-means algorithm and decision tree for early detection of students at risk of dropping out. *Journal of Informatics, Information System, Software Engineering and Applications*, 7(2), 148–162. <https://doi.org/10.20895/inista.v7i2.1630>
- Anwar, S., & Hidayat, T. (2025). Application of K-Means and machine learning for student dropout prediction. *International Journal of Multidisciplinary Studies*, 4(5), 1–10.
- Buulolo, E., & Sihotang, H. (2019). Application of K-Means clustering in student data grouping. *Jurnal Informatika*, 13(2), 77–85.
- Chen, Y., Zhang, L., & Liu, H. (2023). Student dropout prediction using clustering methods. *Educational Data Mining Journal*, 15(3), 210–220.
- Hidayat, H., et al. (2022). Academic performance clustering using K-Means algorithm. *Jurnal Sistem Informasi*, 18(2), 89–98.

Ikotun, A. M., Ezugwu, A. E., & Abualigah, L. (2023). K-means clustering: A comprehensive review. *Information Sciences*, 622, 502–528. <https://doi.org/10.1016/j.ins.2023.01.016>

Kamila, A., et al. (2021). Systematic literature review of K-Means clustering in education data mining. *Jurnal Informatika Edukasi*, 5(1), 33–42.

Liu, X., Wang, J., & Zhao, Y. (2024). Academic analytics using machine learning methods. *IEEE Access*, 12, 112233–112245.

Pratama, A., & Sari, D. (2025). Implementation of K-Means clustering for academic performance analysis. *Jurnal Sistem Informasi Pendidikan*, 11(2), 45–56.

Qusyairi, M., Hidayatullah, Z., & Sandi, A. (2024). K-Means clustering with elbow method for student performance grouping. *Infotek Jurnal Informatika dan Teknologi*, 7(2), 500–510.

Rahmah, S., Jamil, M., & Saputra, Y. F. (2025). Clustering academic performance using K-Means algorithm. *Jurnal Pendidikan Informatika*, 3(4), 1–12.

Sari, N., & Wibowo, A. (2024). Student dropout prediction using data mining techniques. *Jurnal Teknologi Informasi*, 9(1), 22–30.

Tumbilung, F., et al. (2025). Analysis of student dropout risk using K-Means clustering algorithm. *Jurnal Teknologi Informasi dan Pendidikan*, 9(1), 15–24.

Widodo, A., & Nugroho, A. (2017). Data mining for academic performance clustering. *Journal of Information Systems*, 11(1), 15–23.

Yunita, R. (2018). Implementation of K-Means clustering for student data analysis. *Jurnal Teknologi dan Sistem Informasi*, 4(2), 55–63.

Zhang, T., et al. (2023). Educational data mining using clustering techniques. *Computers & Education*, 190, 104608.