

ANALISIS KARAKTERISTIK AKADEMIK MAHASISWA MENGUNAKAN K-MEANS CLUSTERING BERDASARKAN TREN NILAI DAN HAMBATAN UJIAN

Husnul Khair*¹, Tengku Didi Ferdillah²

STMIK Kaputama, Binjai

e-mail: ¹husnul.khair@gmail.com, ²tengkididiferdillah23@gmail.com

Abstract: *This study categorizes students from the 2021 cohort based on academic performance dynamics to overcome the limitations of traditional, static GPA-based evaluations. This approach integrates grade growth trend (slope), remedial exam frequency, and thesis topic classification within the CRISP-DM framework. Utilizing the K-Means algorithm, students are segmented into three distinct clusters: High Achiever, At-Risk, and Critical Failure. The empirical results reveal that the group experiencing the sharpest performance decline (slope -0.720) exhibits the highest remedial exam frequency, averaging 3.69 times, and has not yet proposed a research topic. Conversely, students with positive grade trends display mature research readiness, predominantly focusing on Data Mining. These findings confirm a strong correlation between academic schedule interruptions and delayed graduation risks. The resulting taxonomy successfully maps holistic student risk profiles, serving as an effective early warning system for department policymakers to implement personalized and targeted interventions.*

Keyword: *K-Means; CRISP-DM; academic; text mining.*

Abstrak: Penelitian ini mengelompokkan mahasiswa angkatan 2021 berdasarkan dinamika performa akademik guna mengatasi keterbatasan evaluasi tradisional berbasis IPK statis. Pendekatan ini mengintegrasikan variabel tren pertumbuhan nilai (slope), frekuensi ujian susulan, dan klasifikasi topik skripsi dalam kerangka kerja *CRISP-DM*. Menggunakan algoritma *K-Means*, mahasiswa dikelompokkan ke dalam tiga kluster: High Achiever, At-Risk, dan Critical Failure. Hasil analisis menunjukkan bahwa kelompok mahasiswa yang mengalami penurunan performa paling tajam (slope -0,720) memiliki rata-rata ujian susulan tertinggi, mencapai 3,69 kali, serta belum mengajukan topik tugas akhir. Sebaliknya, mahasiswa dengan tren nilai positif menunjukkan kesiapan riset yang matang, didominasi oleh topik Data Mining. Temuan ini membuktikan adanya korelasi kuat antara interupsi jadwal akademik dengan risiko keterlambatan kelulusan. Model taksonomi yang dihasilkan sukses memetakan profil risiko mahasiswa secara holistik, sekaligus berfungsi sebagai sistem deteksi dini bagi pengambil kebijakan program studi untuk melakukan intervensi yang lebih personal dan tepat sasaran.

Kata kunci: K-Means; CRISP-DM; akademik; text mining.

PENDAHULUAN

Evaluasi performa akademik mahasiswa merupakan instrumen krusial bagi institusi pendidikan tinggi untuk menjamin kualitas lulusan dan menekan angka keterlambatan studi. Namun, pendekatan konvensional dalam evaluasi ini sering kali terjebak pada penggunaan

metrik statis seperti Indeks Prestasi Kumulatif (IPK). Meskipun IPK memberikan gambaran umum mengenai kompetensi kognitif, metrik ini memiliki keterbatasan dalam merepresentasikan dinamika perkembangan mahasiswa dari waktu ke waktu. Kegagalan dalam memantau tren performa jangka panjang menghambat kemampuan institusi untuk

melakukan tindakan preventif secara dini (Asif *et al.*, 2017). Akibatnya, institusi sering kali gagal mendeteksi mahasiswa yang mengalami degradasi performa secara mendadak di tengah masa studi, sebuah fenomena yang jika dibiarkan akan berujung pada kegagalan akademik.

Dalam domain *Educational Data Mining* (EDM), pengolahan data historis mahasiswa kini beralih dari sekadar pelaporan deskriptif menuju analisis prediktif dan preskriptif untuk memahami perilaku belajar yang kompleks (Romero and Ventura, 2020). Penggunaan teknik *clustering* terbukti efektif dalam membedah heterogenitas profil akademik mahasiswa, di mana variabel non-kognitif sering kali memberikan kontribusi signifikan terhadap klasifikasi tingkat keberhasilan studi (Hussain *et al.*, 2018). Walaupun metode pengelompokan telah banyak digunakan, penggabungan antara dinamika tren performa, seperti *slope*, dengan indikator hambatan administratif yang lebih spesifik misalnya frekuensi ujian susulan masih jarang diintegrasikan dalam satu kerangka analisis yang komprehensif. Hambatan-hambatan kecil yang tidak termitigasi secara sistematis berpotensi kuat memicu keputusan mahasiswa untuk menghentikan studi secara prematur (Shafiq *et al.*, 2022).

Penelitian ini mengusulkan sebuah pendekatan dengan menerapkan algoritma *clustering* K-Means yang diperkaya dengan variabel *slope* (kemiringan) dari regresi linear. Penggunaan *slope* memungkinkan sistem untuk menangkap dinamika nilai mahasiswa semester demi semester secara longitudinal, sehingga mahasiswa dengan tren penurunan tajam dapat diidentifikasi meskipun IPK mereka saat ini masih dalam kategori aman.

Selain variabel performa, penelitian ini mengintegrasikan dua faktor eksternal kritical yang sering kali diabaikan diantaranya frekuensi ujian susulan dan progres pengajuan judul skripsi. Frekuensi ujian susulan diposisikan sebagai proksi dari hambatan eksternal (masalah kesehatan atau kendala personal), sementara kesiapan judul

skripsi mencerminkan kematangan orientasi tugas akhir mahasiswa. Dengan menggabungkan data tren kuantitatif dan analisis teks judul skripsi melalui kerangka kerja CRISP-DM, penelitian ini bertujuan untuk memberikan taksonomi karakteristik mahasiswa yang lebih holistik. Kontribusi utama dari studi ini adalah penyediaan model deteksi dini bagi pengambil kebijakan di program studi untuk melakukan intervensi yang lebih personal dan tepat sasaran bagi mahasiswa dalam kategori *At-Risk*.

METODE

Penelitian ini menerapkan pendekatan kuantitatif dengan mengikuti kerangka kerja *Cross-Industry Standard Process for Data Mining* (CRISP-DM). Metodologi ini dipilih karena sifatnya yang iteratif dan terstruktur dalam mentransformasi data akademik mentah menjadi wawasan strategis bagi institusi (Schröer, Kruse and Gómez, 2021).

Business & Data Understanding Tahap awal difokuskan pada identifikasi variabel kunci yang tersedia pada pangkalan data akademik mahasiswa angkatan 2021. Data diekstraksi dari tiga entitas utama: riwayat Kartu Rencana Studi (KRS), log ujian susulan, dan pendaftaran tugas akhir. Penggunaan data historis secara longitudinal ini krusial untuk menangkap dinamika belajar mahasiswa dari waktu ke waktu (Govindasamy and Thambusamy, 2018).

Data Preparation & Feature Engineering Data mentah dibersihkan dari anomali seperti data duplikat dan nilai null. Selanjutnya, dilakukan feature engineering untuk membentuk variabel turunan sebagai berikut:

1. Mean IP: Rata-rata Indeks Prestasi semester untuk mengukur level kognitif umum.
2. Laju Pertumbuhan (Slope): Dihitung menggunakan koefisien regresi linear sederhana (β) untuk merepresentasikan tren performa akademik[3].

$$\gamma = \alpha + \beta x + \varepsilon$$

3. Hambatan Administratif: Frekuensi ujian susulan dihitung per individu sebagai proksi hambatan eksternal.
4. Topik Skripsi: Dikategorikan menggunakan teknik keyword-based extraction untuk memetakan kesiapan tugas akhir mahasiswa.

Modeling (K-Means Clustering) Sebelum proses pemodelan, seluruh variabel numerik diproses menggunakan *StandardScaler* untuk menormalisasi skala data. Algoritma *K-Means* kemudian diterapkan untuk mengelompokkan mahasiswa berdasarkan kedekatan jarak Euclidean (Romero and Ventura, 2020). Penggunaan *K-Means* terbukti andal dalam memetakan kelompok performa mahasiswa yang heterogen menjadi taksonomi tingkat kelulusan yang lebih terukur (Danu *et al.*, 2025).

$$j = \sum_{i=1}^k \sum_{t=1}^n ||x_i^{(j)} - c_j||^2$$

Penelitian ini menetapkan jumlah kelompok ($k=3$) untuk mengclusterkan mahasiswa ke dalam kategori sesuai karakteristik dari masing-masing *cluster*.

Evaluation, Tahap evaluasi dilakukan dengan menganalisis karakteristik setiap *cluster* melalui tabulasi silang (*crosstab*) antara label kelompok dengan variabel ujian susulan dan topik skripsi. Hal ini bertujuan untuk memvalidasi pola perilaku akademik yang tersembunyi (*hidden patterns*) pada setiap kelompok mahasiswa (M., F. and A., 2018).

HASIL DAN PEMBAHASAN

Analisis Karakteristik Cluster Mahasiswa Berdasarkan implementasi algoritma K-Means terhadap variabel performa akademik (*Mean IP* dan *Slope*), diperoleh tiga kelompok karakteristik yang berbeda secara signifikan. Hasil statistik deskriptif untuk setiap cluster disajikan pada Tabel 1.

Tabel 1 Karakteristik Hasil Clustering Mahasiswa

Cluster	Mean IP	Slope (Tren)	Avg Hadir	Jml Susulan	Jumlah Mhs	Topik Dominan	Interpretasi
0	3.61	0.088	13.85	0.44	177	Data Mining	High Achiever
1	3.15	-0.302	12.62	0.95	41	Belum Ada	At-Risk
2	1.19	-0.720	8.73	3.69	13	Belum Ada	Critical Failure

Interpretasi mendalam dari hasil pengelompokan di atas adalah:

1. **Cluster 0 (High Achiever):** Kelompok mayoritas (177 mahasiswa) yang menunjukkan performa sangat baik. Memiliki rata-rata IP tertinggi (3.61) dengan tren

pertumbuhan positif (*slope* 0.088). Kelompok ini juga paling aktif dalam riset tugas akhir dengan dominasi topik *Data Mining*.

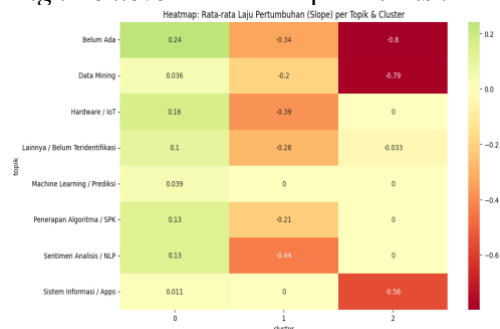
2. **Cluster 1 (At-Risk):** Kelompok menengah (41 mahasiswa) dengan IP yang cukup (3.15) namun memiliki

tren penurunan yang signifikan (*slope* -0.302). Kelompok ini menunjukkan gejala stagnasi akademik dengan indikator utama berupa belum adanya pengajuan judul skripsi.

3. **Cluster 2 (Critical Failure):** Kelompok terkecil (13 mahasiswa) namun paling kritis. Mahasiswa pada kelompok ini memiliki IP yang sangat rendah (1.19) dengan penurunan nilai yang paling ekstrem (*slope* -0.720).

Korelasi Antara Interupsi Akademik dan Penurunan Performa, Data menunjukkan korelasi yang sangat kuat antara Ujian Susulan dan Kehadiran terhadap penurunan performa. Mahasiswa pada *Cluster 2* memiliki rata-rata ujian susulan tertinggi (3.69 kali), hampir 9 kali lipat dibandingkan *Cluster 0*. Hal ini membuktikan secara empiris bahwa interupsi jadwal ujian utama (yang memaksa mahasiswa mengambil ujian susulan) merupakan prediktor kuat terhadap rendahnya *Mean* IP dan tajamnya penurunan *slope*. Rendahnya rata-rata kehadiran (8.73) pada *Cluster 2* juga mempertegas adanya kendala partisipasi aktif dalam proses belajar-mengajar.

Analisis Kesiapan Tugas Akhir, Untuk mendalami hubungan antara performa akademik dan progres tugas akhir, dilakukan analisis distribusi topik skripsi pada setiap *cluster*. Hasil visualisasi *heatmap* menunjukkan adanya polarisasi yang tajam antara kelompok *High Achiever* dan kelompok *At-Risk*.



Gambar 1 Heatmap Topik Tugas Akhir dan Cluster

Interpretasi Hasil Heatmap:

1. **Dominasi Riset pada Cluster 0:** Mahasiswa dalam kelompok ini memiliki kepadatan data yang tinggi pada topik-topik spesifik, terutama *Data Mining*. Hal ini mengindikasikan bahwa stabilitas nilai dan tren pertumbuhan positif memberikan ruang bagi mahasiswa untuk menentukan fokus riset lebih awal dan lebih terarah.
2. **Vakum Kesiapan pada Cluster 1 dan 2:** Terdapat konsentrasi warna yang sangat kontras pada kategori "Belum Ada/Tidak mengajukan judul" untuk mahasiswa di *Cluster 1* dan *Cluster 2*. Hal ini menunjukkan bahwa mahasiswa yang mengalami penurunan performa akademik (*slope* negatif) cenderung mengalami hambatan dalam memulai penelitian tugas akhir mereka.
3. **Sinyal Keterlambatan Kelulusan:** Kelompok *Cluster 2* yang memiliki tren penurunan nilai paling ekstrem (-0.720) secara keseluruhan belum memiliki topik penelitian. Secara statistik, ini mengonfirmasi bahwa penurunan nilai merupakan indikator awal (*early warning*) dari risiko keterlambatan kelulusan yang disebabkan oleh ketidaksiapan tugas akhir.

SIMPULAN

Penelitian ini berhasil mengimplementasikan algoritma *K-Means* dengan pendekatan *CRISP-DM* untuk memetakan karakteristik akademik mahasiswa angkatan 2021 secara lebih komprehensif. Berdasarkan analisis data, dapat disimpulkan bahwa:

Penggunaan variabel *slope* (laju pertumbuhan) terbukti efektif dalam mendeteksi anomali performa akademik yang tidak tertangkap oleh metrik IPK statis. *Cluster 1* dan *Cluster 2* menunjukkan tren penurunan nilai yang signifikan meskipun beberapa mahasiswa masih memiliki rata-rata IP di atas

ambang batas minimal.

Terdapat korelasi linear yang kuat antara hambatan administratif (ujian susulan) dengan penurunan performa. Mahasiswa pada *Cluster* 2 (kelompok paling kritis) memiliki rata-rata ujian susulan tertinggi (3,69 kali) dan kehadiran terendah (8,73), yang menegaskan bahwa interupsi pada jadwal akademik rutin merupakan prediktor utama kegagalan performa.

Analisis topik tugas akhir melalui heatmap mengungkapkan bahwa penurunan performa akademik berbanding lurus dengan ketidaksiapan riset. Mahasiswa dengan tren nilai negatif (*Cluster* 1 dan 2) secara dominan belum memiliki topik penelitian, sementara mahasiswa dengan tren positif (*Cluster* 0) telah mantap pada domain riset tertentu seperti *Data Mining*.

UCAPAN TERIMA KASIH

Penulis mengucapkan puji dan syukur kepada Tuhan Yang Maha Esa atas segala rahmat, karunia, dan penyertaan-Nya sehingga penelitian ini dapat diselesaikan dengan baik. Penulis juga menyampaikan terima kasih yang sebesar-besarnya kepada STMIK Kaputama, khususnya Program Studi Teknik Informatika, yang telah memberikan dukungan, fasilitas, serta lingkungan akademik yang kondusif selama proses penelitian berlangsung.

Ucapan terima kasih juga disampaikan kepada seluruh pihak yang telah membantu dalam penyediaan data akademik, proses pengumpulan dan pengolahan data, serta memberikan berbagai masukan dan saran yang konstruktif selama pelaksanaan penelitian. Apresiasi yang tinggi diberikan kepada rekan-rekan dosen dan tenaga kependidikan yang telah memberikan dukungan, diskusi ilmiah, serta motivasi sehingga penelitian ini dapat diselesaikan sesuai dengan tujuan yang diharapkan.

Penulis berharap hasil penelitian ini dapat memberikan manfaat bagi Program

Studi Teknik Informatika sebagai bahan pertimbangan dalam mengembangkan sistem pemantauan performa akademik mahasiswa berbasis data. Selain itu, penelitian ini diharapkan dapat menjadi referensi bagi peneliti selanjutnya dalam mengembangkan metode Educational Data Mining dan teknik clustering untuk mendukung pengambilan keputusan yang lebih efektif di lingkungan perguruan tinggi.

DAFTAR PUSTAKA

- Asif, R. *et al.* (2017) 'Analyzing undergraduate students' performance using educational data mining', *Computers & Education*, 113, pp. 177–194. Available at: <https://doi.org/https://doi.org/10.1016/j.compedu.2017.05.007>.
- Danu, R. *et al.* (2025) 'Pengelompokkan Data Nilai Mahasiswa Menggunakan Metode K-Means', *Jurnal Algoritme*, 5(3), pp. 350–361. Available at: <https://doi.org/10.35957/algoritme.v6i1.11313>.
- Govindasamy, K. and Thambusamy, V. (2018) 'Analysis of student academic performance using clustering techniques', *International Journal of Pure and Applied Mathematics*, 119, pp. 309–322.
- Hussain, S. *et al.* (2018) 'Educational data mining and analysis of students' academic performance using WEKA', *Indonesian Journal of Electrical Engineering and Computer Science*, 9(2), pp. 447–459. Available at: <https://doi.org/10.11591/ijeecs.v9.i2.pp447-459>.
- M., R., F., N. and A., A. (2018) 'Predicting and Analysis of Students' Academic Performance using Data Mining Techniques', *International Journal of Computer Applications*, 182(32), pp. 1–6. Available at: <https://doi.org/10.5120/ijca2018918250>.
- Romero, C. and Ventura, S. (2020)

-
- ‘Educational data mining and learning analytics: An updated survey’, *WIREs Data Mining and Knowledge Discovery*, 10(3), p. e1355. Available at: <https://doi.org/https://doi.org/10.1002/widm.1355>.
- Schröer, C., Kruse, F. and Gómez, J.M. (2021) ‘A Systematic Literature Review on Applying CRISP-DM Process Model’, *Procedia Computer Science*, 181, pp. 526–534. Available <https://doi.org/https://doi.org/10.1016/j.procs.2021.01.199>.
- Shafiq, D.A. *et al.* (2022) ‘Student Retention Using Educational Data Mining and Predictive Analytics: A Systematic Literature Review’, *IEEE Access*, 10(July), pp. 72480–72503. Available at: <https://doi.org/10.1109/ACCESS.2022.3188767>.

