Journal of Innovative and Creativity
Journal Homepage: https://joecy.org/index.php/joecy
e-mail: journaljoecy@gmail.com

Faculty of Education
University of Pahlawan Tuanku Tambusai
ISSN 2775-771X (Online)
ISSN 2962-570X **(Printed)**
5 (3) 2025, Page 14651-14663

Penerapan Teknik Clustering untuk Identifikasi Pola Perilaku Belajar Siswa Menggunakan Dataset Publik Learning Analytics

Ninik Tri Hartanti[1], Dina Maulina[2], Erni Seniwati[3], Irma Rofni Wulandari[4]
[1,2,3,4] Universitas Amikom Yogyakarta
Corresponding Author: Ninik Tri Hartanti, e-mail: ninik.t@amikom.ac.id

Published: Desember, 2025

ABSTRAK

Kemajuan teknologi informasi mendorong berkembangnya bidang *Educational Data Mining* (EDM) yang berfokus pada pemanfaatan data pendidikan untuk menganalisis perilaku belajar peserta didik. Penggunaan dataset publik *Learning Analytics* dalam EDM, jika dikombinasikan dengan metode klasterisasi, dapat membantu proses pengelompokan karakteristik belajar siswa. Penelitian ini bertujuan mengungkap pola perilaku belajar siswa melalui penerapan teknik clustering pada dataset publik *learning analytics*. Pendekatan yang digunakan adalah algoritma K-Means dengan pengukuran kualitas klaster melalui Davies-Bouldin Index (DBI). Data penelitian berasal dari dataset publik *Student Performance* yang memuat informasi terkait aktivitas atau durasi belajar, tingkat kehadiran, riwayat kegagalan belajar, serta nilai siswa. Hasil klasterisasi menghasilkan empat kelompok utama, yaitu siswa dengan risiko akademik tinggi, siswa berperforma sedang, siswa berprestasi tinggi, dan kelompok dengan tingkat ketidakhadiran yang tinggi. Nilai DBI sebesar 0.0072 menunjukkan bahwa kualitas klaster yang terbentuk sangat baik. Temuan ini diharapkan dapat menjadi dasar dalam perancangan strategi pembelajaran yang lebih efektif

Kata Kunci : clustering, K-Means, DBI, learning analytics, dataset publik

ABSTRACT

Advances in information technology are driving the development of the field of Educational *Data Mining*(EDM), which focuses on the use of educational data to analyze student learning behavior. The use of public datasets *Learning Analytics*In EDM, when combined with clustering methods, it can assist in the process of grouping student learning characteristics. This study aims to uncover student learning behavior patterns by applying clustering techniques to public datasets.*learning The* approach used is the K-Means algorithm with cluster quality measured using the Davies-Bouldin Index (DBI). The research data comes from public datasets. *Student Performance*which contains information related to learning activities or duration, attendance levels, history of learning failure, and student grades. The clustering results yielded four main groups: students at high academic risk, students with moderate performance, students with high achievement, and a group with a high absenteeism rate. The DBI value of 0.0072 indicates that the quality of the formed clusters is very good. These findings are expected to form the basis for designing more effective learning strategies..

Keywords: clustering, K-Means, DBI, learning analytics, public dataset

© The Author(s). 2025 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

Perkembangan teknologi informasi telah mendorong munculnya bidang *Educational Data Mining* (EDM) yang memanfaatkan data pendidikan untuk memahami perilaku belajar siswa dan memungkinkan pengumpulan data pembelajaran dalam skala besar. Salah satu pendekatan yang sering digunakan dalam EDM adalah *clustering*, yaitu teknik pengelompokan data berdasarkan kemiripan karakteristik. Penerapan *clustering* memungkinkan proses identifikasi adanya pola tersembunyi dalam data aktivitas belajar, sebagai

contohnya adalah perbedaan perilaku antara siswa yang aktif dan pasif, atau antara siswa yang konsisten dan tidak konsisten dalam mengikuti pembelajaran. Data ini dapat dianalisis untuk memahami perilaku belajar siswa dan mengidentifikasikan pola yang dapat digunakan untuk meningkatkan hasil belajar. *Learning analytics* merupakan bidang yang berkonsentrasi pada pengukuran, pengumpulan, analisis, dan pelaporan data tentang peserta didik dan bertujuan untuk memahami serta mengoptimalkan pembelajaran dalam lingkungan tersebut berada (Seng & Chuan, 2023). Teknik clustering seperti K-Means telah banyak digunakan dalam analisis data pendidikan karena kemampuannya dalam mengelompokkan data tanpa pengawasan (unsupervised). Evaluasi cluster menggunakan Davies-Bouldin Index (DBI) membantu dalam menentukan kualitas pengelompokan yang dihasilkan. Penelitian sebelumnya oleh (Hendrastuty, 2024) dan (Bili et al., 2024) telah menunjukkan efektivitas K-Means dalam mengidentifikasi pola belajar siswa. Terkait penerapan teknik clustering, penelitian dengan topik penentuan kelompok dalam aktivitas belajar khususnya minat siswa memilih sekolah kejuruan atau SMK (Hartanti et al., 2024) dengan menerapkan metode K-Means Clustering dan Elbow, menghasilkan 3 pola cluster untuk wilayah kabupaten Magelang.

Identifikasi fitur hasil program pembelajaran yang membentuk cluster tertentu dalam bidang pendidikan, salah satu penelitian oleh (Mohd Talib et al., 2023) dengan menerapkan teknik clustering dan melalui pendekatan Support Vector Machine (SVM), dihasilkan 3 tipe pola utama yaitu kinerja rendah, rata-rata, dan tinggi. Hal ini menunjukkan kinerja K-Means clustering dalam identifikasi pola belajar siswa dengan fitur hasil program pembelajaran.

Penelitian terkait teknik clustering dengan pendekatan AHP oleh (Utami et al., 2024) melakukan clustering perilaku siswa agar dapat mengidentifikasi karakteristik kelompok siswa dan menyoroti cluster yang berisiko mengalami kegagalan pembelajaran. Clustering dengan kombinasi pendekatan Long Short-Term Memory (LSTM) oleh (Chen & Cui, 2020). Penelitian ini memanfaatkan pengolahan deret waktu frekuensi klik harian dan mampu mendeteksi lebih awal siswa yang berisiko dengan akurasi prediksi sedang.

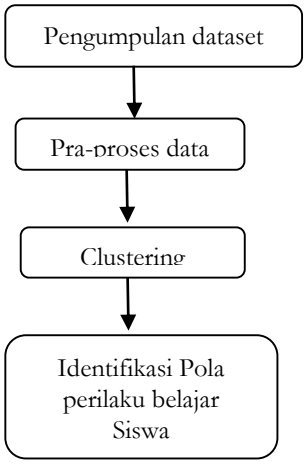
Penelitian learning analytics oleh (Franke et al., 2024) menyoroti peran learning analytics dalam meningkatkan proses kualifikasi pendidikan di institusi melalui jalur pembelajaran adaptif. Hasil penelitian berupa profil yang terbentuk akan dapat membantu untuk memahami bagaimana siswa berbeda dalam motivasi, keterlibatan, dan cara belajar mereka. Penerapan teknik clustering dengan metode evaluasi cluster, salah satunya adalah DBI (Davies-Bouldin Index) telah banyak mengalami perkembangan. Salah satunya terdapat pada penelitian oleh (Abdul Mukhsyi et al., 2025).

Adapun tujuan dari penelitian Penerapan Teknik Clustering untuk Identifikasi Pola Belajar Siswa Menggunakan Dataset Publik Learning Analytics adalah Menerapkan teknik K-Means Clustering pada dataset publik learning analytics, mengidentifikasi pola perilaku belajar siswa berdasarkan hasil klasterisasi,

dan memberikan interpretasi sederhana yang dapat digunakan untuk mendukung perbaikan proses pembelajaran.

METODE

Penelitian ini menggunakan pendekatan kuantitatif dengan metode data mining berbasis teknik *clustering*. Proses perhitungan dengan teknik *clustering* diawali dengan dataset yang diperoleh dari dataset publik *Learning Analytic*, yaitu dataset *Student Academic Performance* dari Kaggle. Adapun langkah dalam pelaksanaan penelitian ini terlampir di Gambar 1.



Gambar 1. Metodologi Penelitian

Pada tahap pra-proses data akan dilakukan cleaning data atau pembersihan data, dengan cara menormalisasikan dataset yang telah diperoleh sebelumnya, Student Academic Performance sebanyak 395 siswa. Tabel 2 merupakan hasil normalisasi datanya.

Tabel 1. Dataset *Student Academic Performance*

No	Studytime	Failures	Absences	G1	G2	G3
1	2	0	6	5	6	6
2	2	0	4	5	5	6
3	2	3	10	7	8	10
4	3	0	2	15	14	15
5	2	0	4	6	10	10
6	2	0	10	15	15	15
7	2	0	0	12	12	11
8	2	0	6	6	5	6
9	2	0	0	16	18	19
10	2	0	0	14	15	15
11	2	0	0	10	8	9
12	3	0	4	10	12	12
13	1	0	2	14	14	14
14	2	0	2	10	10	11
15	3	0	0	14	16	16
---	---	---	---	---	---	---

394	1	0	0	11	12	10
395	1	0	5	8	9	9

Dataset akan dilakukan proses normalisasi dengan menggunakan persamaan berikut,

$$Zi = \frac{Xi - \mu}{\sigma}$$ Persamaan (1)

Keterangan

Zi : nilai hasil normalisasi data ke-i

Xi : nilai asli data ke-i

μ : nilai rata-rata (mean) dari seluruh data

σ : standar deviasi dari seluruh data

Proses perhitungan selanjutnya akan menggunakan data hasil normalisasi dari dataset dengan menerapkan persamaan (1). Kemudian dilanjutkan dengan menerapkan persamaan (2) untuk menentukan nilai Elbow berdasarkan nilai SSE (Sum of Squared Errors). Oleh karena itu tahap pra-proses data akan diakhiri dengan memberikan nilai hasil perhitungan nilai Elbownya. Persamaan untuk menghitung nilai Elbow seperti berikut:

$$SSE = \sum_{k}^{k} = 1 \sum x \in S_k \ \ \|x_1 - C_k\|_2^2$$ Persamaan (2)

Keterangan

SSE : Sum of Squared Errors

xi : jarak antara objek ke-i dan C_k yang merupakan pusat klaster ke-i.

Tahap selanjutnya melibatkan proses pengelompokan data dengan memanfaatkan algoritma K-Means. Metode ini termasuk dalam kategori *unsupervised learning* yang digunakan untuk memisahkan data ke dalam beberapa klaster berdasarkan tingkat kesamaannya. Pembentukan klaster dilakukan melalui proses berulang yang terdiri atas dua langkah inti, yaitu menempatkan setiap data pada klaster terdekat dan memperbarui posisi pusat klaster (centroid) hingga mencapai kondisi stabil. Tujuan utama dari pendekatan ini adalah memastikan setiap data berada sedekat mungkin dengan centroid klaster yang mewakilinya, sehingga pola pengelompokan yang terbentuk menjadi optimal. Tahapan dalam melakukan proses clustering K-Means dilakukan melalui beberapa langkah sebagai berikut (Zubair et al., 2024)

1. Inisialisasi: Memilih K centroid awal secara acak dari dataset
2. Assignment: Setiap data dialokasikan ke klaster yang centroid-nya merupakan yang terdekat menurut perhitungan jarak Euclidean
3. Update: Menghitung ulang centroid sebagai rata-rata dari semua data point dalam cluster

4. Iterasi: Mengulangi langkah 2 dan 3 hingga konvergen (perubahan centroid < threshold) atau mencapai iterasi maksimum

Penilaian kualitas klaster dilakukan melalui perhitungan Davies-Bouldin Index, sehingga dapat menentukan nilai K optimal. Proses ini meminimalkan within-cluster sum of squares sehingga mencapai struktur cluster yang kompak dan terpisah dengan baik.

Proses diawali dengan menentukan jumlah cluster k yang akan dibentuk, kemudian melakukan inisialisasi centroid secara acak. Selanjutnya, dilakukan perhitungan jarak antara setiap data terhadap centroid menggunakan rumus *Euclidean Distance* sebagai berikut:

$$d(p,q)=\sqrt{(p1-q1)2+(p2-q2)2+\cdots+(pn-qn)^2}$$ Persamaan (3)

Keterangan

p,q : merupakan dua titik dalam ruang berdimensi n

pi : nilai ke-i dari titik p

qi : nilai ke-i dari titik q

(pi-qi)2 : selisih kuadrat dari masing-masing

Berdasarkan hasil perhitungan nilai SSE dengan metode Elbow dan hasil clustering maka langkah berikutnya adalah evaluasi cluster dengan menggunakan DBI (Davies-Bouldin Index), yang merupakan sebuah metrik evaluasi untuk mengukur kualitas hasil clustering. Metrik ini bersifat internal, artinya evaluasi dilakukan hanya berdasarkan data yang di cluster saja, tanpa menggunakan label kebenaran (ground truth) seperti pada evaluasi eksternal seperti accuracy atau F1-Score(Ramadhan et al., 2025). Persamaan untuk menentukan DBI adalah:

$$DB = \frac{1}{k}\sum_{i=1}^{k}\max_{j\neq i}\left(\frac{S_i+S_j}{d_{ij}}\right)$$ Persamaan (4)

Langkah terakhir dalam penelitian adalah pengelompokan berdasarkan hasil perhitungan dengan menerapkan persamaan (1), (2), (3) dan (4). Adapun hasilnya adalah berupa cluster atau kelompok siswa dengan beberapa kategori sesuai dengan hasil akhir dari proses clustering.

HASIL DAN PEMBAHASAN

Tahap Pra Proses Data

Merujuk Gambar yang menggambarkan uraian semua langkah dalam penelitian, berikut adalah hasil dari proses perlangkah dalam penelitian. Langkah awal dimulai dari penentuan dataset yang kemudian data tersebut ditetapkan dengan nilai numerik yaitu dengan normalisasi dengan persamaan (1). Adapun dataset yang digunakan adalah terlampir Pada dataset tersebut terdapat sebanyak 395 siswa yang belajar matematika (student-mat) yang diperoleh dari data Student Academic Performance dari Kaggle. Setelah

dilakukan proses normalisasi dengan menggunakan persamaan (1), maka dataset akan menjadi data di Tabel 2.

Tabel 2. Normalisasi Data

No	Studytime	Failures	Absences	G1	G2	G3
1	-0,045	-0,451	0,036	-1,785	-1,256	-0,966
2	-0,045	-0,451	-0,214	-1,785	-1,522	-0,966
3	-0,045	3,588	0,537	-1,181	-0,724	-0,092
4	1,148	-0,451	-0,464	1,232	0,874	1,001
5	-0,045	-0,451	-0,214	-1,483	-0,191	-0,092
6	-0,045	-0,451	0,537	1,232	1,140	1,001
7	-0,045	-0,451	-0,714	0,327	0,341	0,127
8	-0,045	-0,451	0,036	-1,483	-1,522	-0,966
9	-0,045	-0,451	-0,714	1,534	1,938	1,875
10	-0,045	-0,451	-0,714	0,930	1,140	1,001
11	-0,045	-0,451	-0,714	-0,276	-0,724	-0,310
12	1,148	-0,451	-0,214	-0,276	0,341	0,346
13	-1,238	-0,451	-0,464	0,930	0,874	0,783
14	-0,045	-0,451	-0,464	-0,276	-0,191	0,127
15	1,148	-0,451	-0,714	0,930	1,406	1,220
---	---	---	---	---	---	---
---	---	---	---	---	---	---
---	---	---	---	---	---	---
394	-1,238	-0,451	-0,714	0,025	0,341	-0,092
395	-1,238	-0,451	-0,089	-0,880	-0,457	-0,310

Proses perhitungan selanjutnya akan menggunakan data dari Tabel 2 tersebut, karena data telah dinormalisasikan. Hal ini bertujuan untuk memudahkan dalam proses perhitungan secara numerik, dikarenakan proses *clustering* akan mudah dalam perhitungan data dengan data bersifat numerik. Sebelum melakukan proses *clustering*, dari data yang telah dinormalisasi, maka proses selanjutnya adalah menghitung nilai Elbow dari data ternormalisasi tersebut. Oleh karena itu tahap pra-proses data akan diakhiri dengan hasil perhitungan nilai Elbownya. Persamaan untuk menghitung nilai Elbow seperti pada persamaan (2). Hasil perhitungan dengan menerapkan metode Elbow terlampir pada Tabel 3.

Tabel 3. Nilai SSE untuk 3 Cluster 4 Cluster dan 5 Cluster

Uji Coba ke-	Jumlah Data	Cluster		
		3	4	5
1	150	1,384	1,174	0,447
2	100	42,652	1,079	0,901
3	120	1,137	0,501	1,353

Berdasarkan Tabel 3 terlihat bahwa terdapat 3 kali uji coba untuk perhitungannya. Uji coba ke-1 menggunakan 150 data, uji coba ke-2 menggunakan 100 data, dan di uji coba terakhir menggunakan 120 data. Dalam setiap uji coba tersebut, akan diproses berdasarkan jumlah cluster nya. Dari uji coba ke-1 dengan menggunakan 150 data untuk 3 cluster diperoleh nilai SSE = 1,384. Uji coba ke-2 dengan menggunakan 100 data untuk 3 cluster diperoleh nilai SSE = 42,652 dan uji coba ke-3 dengan

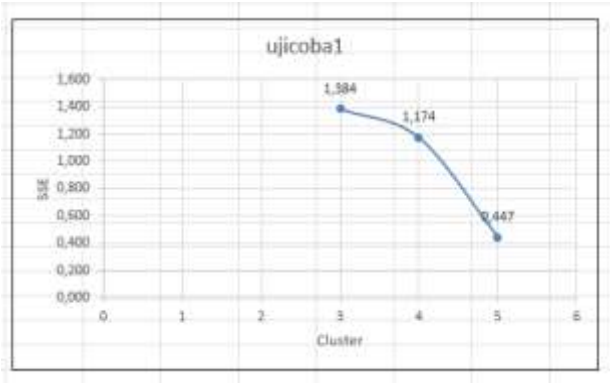
menggunakan 120 data untuk 3 cluster diperoleh nilai SSE = 1,137. Begitu juga dalam setiap uji coba dengan menggunakan sejumlah data tertentu untuk 4 cluster dan 5 cluster akan dihasilkan nilai SSE seperti pada Tabel 3.

Tabel 4. Nilai Selisih SSE untuk Setiap Cluster di Uji coba ke1

Jml Cluster	Nilai SSE	Selisih SSE
3	1,384	1,384
4	1,174	0,210
5	0,447	0,728

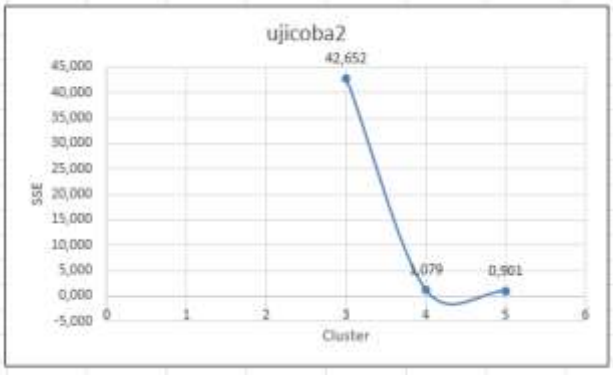
Merujuk Tabel 4 terdapat nilai selisih SSE untuk uji coba ke-1 dari jumlah cluster = 3, 4 dan 5. Diketahui nilai SSE di uji coba ke-1 untuk 3 cluster = 1,384 dan untuk 4 cluster = 1,174 maka selisihnya adalah 1,174 – 1,384 = 0,210. Sedangkan untuk selisih 3 cluster dengan sebelumnya adalah berupa nilai SSE itu sendiri. Seperti pada Tabel 4, selisih SSE untuk jumlah cluster =3 adalah 1,385.

Hasil nilai selisih SSE tersebut harus diproses sampai dengan uji coba ke-3, sehingga dari hasil selisih nilai SSE mulai dari uji coba ke-1 sampai dengan uji coba ke-3 akan menghasilkan grafik seperti Gambar 2, Gambar 3 dan Gambar 4.

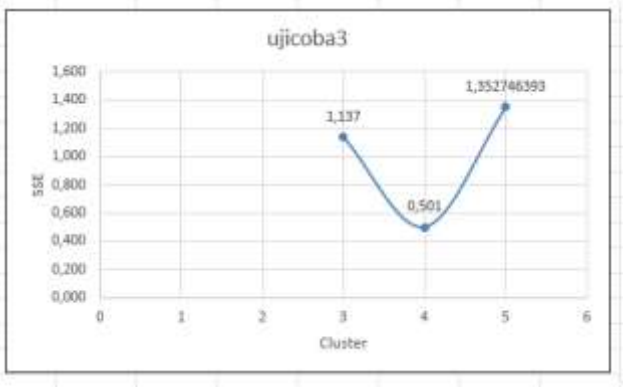


Gambar 2. Nilai SSE 1

Gambar 2 di atas merupakan grafik dari hasil nilai SSE untuk uji coba ke-1 yang berhubungan dengan hasil nilai pada Tabel 4. Sedangkan Gambar 3 dan Gambar 4 merupakan grafik dari hasil nilai SSE untuk uji coba ke-2 dan uji coba ke-3.



Gambar 3. Nilai SSE 2



Gambar 4. Nilai SSE 3

Menurut hasil nilai SSE yang terdapat pada Gambar 2, 3, dan 4 terlihat bahwa terdapat garis melengkung di posisi angka 4, sehingga nilai k untuk jumlah cluster optimal sesuai nilai SSE dengan menggunakan metode Elbow adalah 4. Oleh karena itu proses perhitungan selanjutnya akan menggunakan data cluster (k=4).

Tahap Clustering

Proses perhitungan selanjutnya adalah clustering data. Pada tahap ini menggunakan pendekatan algoritma K-Means untuk perhitungannya. Langkah ini membutuhkan data cluster yang sebelumnya telah ditentukan terlebih dahulu dengan metode Elbow di persamaan (2). Diketahui sebelumnya bahwa dengan perhitungan metode Elbow diperoleh data jumlah cluster optimal adalah 4 (k=4). Oleh karena itu, pada tahap clustering ini akan dilakukan proses cluster untuk k=4.

Data yang akan digunakan untuk proses clustering adalah data yang terlampir di Tabel 1. Selanjutnya dari data tersebut akan diproses perhitungan clusternya dengan menggunakan pendekatan algoritma K-Means, dengan nilai k=4.

Tabel 5. Nilai Centroid Awal

Cluster	Studytime	Failures	Absences	G1	G2	G3	Studytime
1	-0,05	-0,45	-0,21	-1,48	-0,19	-0,09	-0,05
2	-0,05	-0,45	-0,71	-0,58	-0,46	-0,09	-0,05
3	-0,05	-0,45	-0,21	-0,58	-0,19	-0,09	-0,05
4	-1,24	-0,45	-0,09	-0,88	-0,46	-0,31	-1,24

Melalui dataset normalisasi penentuan nilai centroid awal diambil dari nilai data ke-5, ke-200, ke-340, dan data ke-395. Setelah ditentukan nilai centroid awal, langkah berikutnya adalah menghitung jarak semua centroid dengan menerapkan persamaan (3). Hasil perhitungannya terlampir pada Tabel 6.

Tabel 6. Jarak Antar Centroid dan Penentuan Cluster Iterasi 1

No	Jarak 1	Jarak2	Jarak3	Jarak4	Cluster
1	1,43	1,85	1,85	1,82	1
2	1,62	1,90	2,00	1,95	1
3	4,15	4,28	4,19	4,28	1
4	3,34	2,78	2,66	3,71	3

5	0,00	1,07	0,90	1,39	1
6	3,30	2,93	2,61	3,25	3
7	1,96	1,23	1,18	2,02	3
8	1,61	1,81	1,85	1,83	1
9	4,21	3,75	3,62	4,26	3
10	3,01	2,45	2,34	3,06	3
11	1,43	0,46	0,82	1,50	2
12	1,83	1,61	1,41	2,67	3
13	3,03	2,51	2,38	2,53	3
14	1,25	0,52	0,45	1,48	3
---	---	---	---	---	---
---	---	---	---	---	---
---	---	---	---	---	---
394	2,06	1,56	1,52	1,38	4
395	1,39	1,40	1,28	0,00	4

Jarak antar centroid di atas akan digunakan sebagai penentuan centroid baru setelah sebelumnya sudah ditentukan centroid awal. Dengan menentukan nilai rata-rata dari setiap jarak di atas, maka dapat diperoleh data centroid baru sebagai berikut, terlampir pada Tabel 7.

Tabel 7. Centroid Baru

Centroid	Study time	Failures	Absences	G1	G2	G3
1	-0,129	1,709	0,214	-1,26	-1,064	-0,981
2	0,429	-0,313	-0,432	-0,75	-0,853	-0,966
3	0,149	-0,332	-0,062	0,716	0,688	0,666
4	-1,036	0,370	0,638	-0,70	-0,611	-0,440

Sesuai dengan perhitungan penentuan nilai centroid baru tersebut, maka dapat diperoleh jumlah atau banyaknya data dalam setiap cluster yang ditampilkan pada Tabel 8.

Tabel 8. Jumlah Data Tiap Cluster Iterasi1

Cluster 1	Cluster 2	Cluster 3	Cluster 4
43	78	215	59

Centroid baru telah diperoleh dan jumlah data untuk setiap cluster sudah dihasilkan, maka langkah selanjutnya adalah perhitungan K-Means iterasi2 yaitu dengan cara kembali menentukan jarak terdekat setiap centroid, kemudian dari hasil jarak terdekat tersebut akan diperoleh cluster dan dari cluster akan dihasilkan jumlah data di setiap clusternya. Langkah ini akan terus dilakukan berulang kali sampai jumlah data dalam setiap cluster tidak ada perubahan atau tetap. Adapun hasil dari perhitungan K-Means iterasi 2 ditampilkan pada Tabel 9 dan 10.

Tabel 9. Jarak Antar Centroid dan Penentuan Cluster Iterasi 2

No	Jarak 1	Jarak2	Jarak3	Jarak4	Cluster
1	2,24	1,30	3,57	1,97	2
2	2,31	1,34	3,72	2,16	2
3	2,13	4,17	4,68	3,42	1
4	4,55	3,37	1,26	3,83	3

5	2,54	1,42	2,50	1,81	2
6	4,45	3,60	1,00	3,25	3
7	3,36	2,03	1,02	2,40	3
8	2,23	1,20	3,53	1,93	2
9	5,53	4,63	2,04	4,51	3
10	4,38	3,32	0,92	3,36	3
11	2,66	1,00	2,10	1,92	2
12	3,35	1,99	1,50	2,81	3
13	4,25	3,42	1,48	2,88	3
14	2,85	1,45	1,50	1,89	2
---	---	---	---	---	---
---	---	---	---	---	---
---	---	---	---	---	---
394	3,35	2,38	1,88	2,03	3
395	2,64	1,88	2,60	1,15	4

Tabel 10. Jumlah Data Tiap Cluster Iterasi 2

Cluster 1	Cluster 2	Cluster 3	Cluster 4
40	99	194	62

Langkah di atas akan terus dilakukan sampai jumlah data di setiap cluster tidak mengalami perubahan. Jumlah data untuk setiap cluster berhasil tidak mengalami perubahan setelah dilakukan proses perhitungan K-Means sebanyak 14 iterasi dari 395 data, 4 cluster. Adapun hasil perhitungan untuk setiap iterasi ditampilkan pada Tabel 11.

Tabel 11. Hasil Clustering Tiap Iterasi

Iterasi	Cluster 1	Cluster 2	Cluster 3	Cluster 4
1	43	78	215	59
2	40	99	194	62
3	38	112	180	65
4	40	122	173	60
5	43	132	165	55
6	43	137	161	54
7	45	139	159	52
8	47	140	157	51
9	48	141	155	51
10	49	143	153	50
11	50	143	153	49
12	50	144	153	48
13	**50**	**145**	**153**	**47**
14	**50**	**145**	**153**	**47**

Pada Tabel 11 terlihat bahwa jumlah data di setiap cluster tidak mengalami perubahan terletak di Iterasi 13 dan 14, sehingga proses perhitungan clustering dengan K-Means akan berhenti di Iterasi 14. Pada iterasi terakhir yaitu iterasi 14 dihasilkan data cluster seperti terlampir pada Tabel 12.

Tabel 12. Centroid Iterasi 14

Centroid	Study time	Failures	Absences	G1	G2	G3
1	-0,52	2,16	-0,17	-1,16	-1,12	-1,25
2	0,14	-0,40	-0,29	-0,61	-0,54	-0,47
3	0,13	-0,34	-0,24	0,99	0,94	0,88
4	-0,32	0,04	1,86	-0,13	-0,19	-0,09

Tahap Evaluasi Cluster

Pada hasil K-Means di iterasi terakhir (iterasi ke 14) pada Tabel 12 selanjutnya akan dievaluasi dengan menggunakan DBI sesuai persamaan (4) untuk mengukur seberapa baik pemisahan (separation) antar cluster dan seberapa padat (cohesion) setiap cluster tersebut. Hasil dari proses DBI = 0,007283. Dari hasil tersebut, bisa dikatakan bahwa nilai DBI < 1 yang artinya cluster cukup baik, pemisahan antar cluster jelas. Jika nilai DBI mendekati nol maka cluster tersebut sangat baik, namun apabila nilai DBI > 1 maka cluster tersebut saling tumpang tindih atau pemisahan antar cluster tidak baik.

KESIMPULAN

Pada penelitian dengan judul Penerapan Teknik Clustering untuk Identifikasi Pola Perilaku Belajar Siswa Menggunakan Dataset Publik Learning Analytics ini bertujuan untuk mengidentifikasi pola perilaku belajar siswa menggunakan teknik clustering berdasarkan dataset publik learning analytics. Teknik clustering yang digunakan adalah dengan pendekatan algoritma K-Means. Dengan melalui beberapa langkah penelitian yang tercantum dalam metodolgi penelitian, maka dapat disimpulkan beberapa hal berikut ini.

1. Data yang digunakan diambil dari dataset publik learning analytics, dengan 395 data, 4 cluster dan dibutuhkan sebanyak 14 iterasi dalam proses clustering.
2. Cluster 1 adalah kelompok dengan risiko akademik tinggi. Siswa dalam cluster1 menunjukkan mempunyai performa akademik yang buruk, dengan ditandai rendahnya waktu belajar (studytime) serta tingginya jumlah kegagalan akademik (failures). Walaupun absensi (absences) tidak bermasalah, rendahnya nilai ujian (G1,G2, dan G3) menunjukkan bahwa mereka berada dalam kategori risiko akademik tinggi.
3. Cluster 2 merupakan kelompok performa menengah. Siswa pada cluster ini memiliki disiplin kehadiran baik dan jarang mengalami kegagalan dalam mata pelajaran. Namun, meskipun waktu belajar cukup baik, nilai yang dihasilkan belum optimal dan masih berada sedikit di bawah rata-rata. Mereka termasuk kelompok dengan performa akademik menengah yang cenderung stabil namun belum maksimal.
4. Cluster 3 merupakan kelompok siswa berprestasi tinggi. Cluster ini merupakan kelompok siswa yang memiliki performa terbaik. Disiplin kehadiran tinggi, jarang mengalami kegagalan dalam mata pelajaran, dan memiliki nilai paling tinggi pada seluruh tahapan penilaian (G1–G3). Mereka termasuk kategori siswa berprestasi tinggi.

5. Cluster 4 adalah absensi tinggi. Siswa dalam cluster 4 ini menunjukkan tingkat absensi yang sangat tinggi, namun performa akademik tidak serendah cluster 1. Nilai mereka mendekati rata-rata meskipun waktu belajar rendah. Kelompok ini termasuk kategori siswa dengan performa campuran, yaitu kehadiran buruk tetapi nilai masih bertahan.

REFERENCES

Abdul Mukhsyi, S., Irma Purnamaari, A., & Bahtiar, A. (2025). *Improving Student Achievement Clustering Model Using K-Means Algorithm in Pasundan Majalaya Vocational School* (Vol. 4, Issue 2). https://ioinformatic.org/

Bili, N., Thimotius Abineno, R., Arini Aha Pekuwali, dan, Studi Teknik Informatika, P., Sains Dan Teknologi, F., Kristen Wira Wacana Sumba JlRSuprato No, U., Waingapu, K., Sumba Timur, K., & Tenggara Timur, N. (2024). *Penerapan Algoritma K-MEANS Clustering untuk Pengelompokan Performa Siswa Pada Pembelajaran Bahasa Indonesia (Studi Kasus: SD INPRES Waingapu 3)*.

Chen, F., & Cui, Y. (2020). Utilizing student time series behaviour in learning management systems for early prediction of course performance. *Journal of Learning Analytics*, *7*(2), 1–17. https://doi.org/10.18608/JLA.2020.72.1

Franke, S., Trommer, M., Hellbach, S., & Teich, T. (2024). Clustering of Learners' Data - A Case Study on Learning Analytics. *IFAC-PapersOnLine*, *58*(26), 118–123. https://doi.org/10.1016/j.ifacol.2024.10.281

Hartanti, N. T., Seniwati, E., Pramitasari, R., Wulandari, I. R., & Yogyakarta, U. A. (2024). Jumlah Cluster Optimal dalam Pengelompokan Siswa SMK dengan Metode Elbow K-Means Clustering. In *Journal of Accounting Information System* (Vol. 4, Issue 2). http://jurnal.bsi.ac.id/index.php/jais

Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa. *Jurnal Ilmiah Informatika Dan Ilmu Komputer (JIMA-ILKOM)*, *3*(1), 46–56. https://doi.org/10.58602/jima-ilkom.v3i1.26

Mohd Talib, N. I., Abd Majid, N. A., & Sahran, S. (2023). Identification of Student Behavioral Patterns in Higher Education Using K-Means Clustering and Support Vector Machine. *Applied Sciences (Switzerland)*, *13*(5). https://doi.org/10.3390/app13053267

Ramadhan, B. F., Abdillah, B. M., Hidayatullah, M., Faizal, M. N., Winata, R. A., & Alfian, Z. (2025). *Analisis Performa Clustering: K-Means dan Similarity Matrix dalam Evaluasi Silhouette, DBI, CHI, dan Dunn Index*. https://jptam.org/index.php/jptam/article/view/30265/19983

Seng, E., & Chuan, N. C. (2023). Integrating Learning Analytics into the Flipped Classroom: A Study on Data-Enabled Flipped Learning. *OALib*, *10*(09), 1–18. https://doi.org/10.4236/oalib.1110692

Utami, I. Q., Hwang, W.-Y., & Ningrun, R. A. (2024). Student's Behavior Clustering based on Ubiquitous Learning Log Data using Unsupervised Machine Learning. *Journal of Advanced Technology and*

Multidiscipline (JATM), *3*(1), 13–20. https://e-journal.unair.ac.id/JATM/article/view/55572/29485

Zubair, M., Iqbal, M. A., Shil, A., Chowdhury, M. J. M., Moni, M. A., & Sarker, I. H. (2024). An Improved K-means Clustering Algorithm Towards an Efficient Data-Driven Modeling. *Annals of Data Science*, *11*(5), 1525–1544. https://doi.org/10.1007/s40745-022-00428-2