

PENERAPAN METODE K-MEANS CLUSTERING UNTUK PENGELOMPOKAN SISWA BERDASARKAN PRESTASI DI SD N 03 SANGGANG SUKOHARJO

Dika Restyanto¹⁾; Kumaratih Sandradewi²⁾; Retno Tri Vulandari^{3)*}

¹⁾ Program Studi Informatika Universitas Tiga Serangkai Surakarta, Indonesia

²⁾ Program Studi Manajemen Universitas Tiga Serangkai Surakarta, Indonesia

³⁾ Program Studi Teknologi Informasi Universitas Tiga Serangkai Surakarta, Indonesia

¹⁾18500087.dika@sinus.ac.id ; ²⁾kumaratih@tsu.ac.id ; ^{3*)}retnotv@tsu.ac.id

ABSTRACT

The rapid growth of data science has made efficient data processing essential in education. In schools, student achievement is assessed through both theoretical and practical subjects. However, manual grouping of achievement levels (very good, good, sufficient) is often inefficient, making it difficult for homeroom teachers to identify student performance early. This study applies the K-Means Clustering method to classify students' achievement levels at SD N Sanggang 03. The goal is to provide a system that helps teachers categorize students into performance groups more effectively, thereby supporting data-driven decision-making in academic evaluation. The system was developed using PHP and MySQL, and its clustering quality was evaluated with the Modified Partition Coefficient.

Keywords: Data Mining, K-Means Clustering, Student Achievement Grouping, Educational Data Analysis, Modified Partition Coefficient

I. PENDAHULUAN

Perkembangan pesat dalam bidang data science menjadikan pengolahan data sebagai kebutuhan utama di berbagai sektor, termasuk pendidikan [1]. Dalam dunia pendidikan, pemanfaatan sumber daya dan sistem informasi berperan penting untuk meningkatkan efisiensi serta mendukung pengambilan keputusan strategis [2]. Di SD N Sanggang 03, proses penentuan kelompok prestasi siswa (sangat baik, baik, cukup) masih dilakukan secara manual. Penilaian manual yang tidak efisien ini menghambat deteksi dini terhadap prestasi siswa dan memperlambat intervensi pembelajaran yang seharusnya dapat diberikan tepat waktu [3] [4]. Akibatnya, potensi siswa yang membutuhkan perhatian khusus seringkali terlambat teridentifikasi.

Sejumlah penelitian sebelumnya telah menerapkan metode K-Means Clustering untuk mengelompokkan data prestasi siswa, namun sebagian besar fokus pada tingkat pendidikan menengah [5] [6]. Padahal, konteks Sekolah Dasar memiliki karakteristik yang berbeda: data prestasi siswa lebih bervariasi karena meliputi aspek dasar seperti membaca, menulis, berhitung, serta keterampilan praktis; selain itu, kebutuhan intervensi di SD bersifat lebih mendesak karena berkaitan langsung dengan pembentukan fondasi belajar siswa. Dengan demikian, kurangnya penelitian

serupa di tingkat SD, khususnya di SD N Sanggang 03, menunjukkan adanya gap yang signifikan dan perlu segera diisi.

Oleh karena itu, penelitian ini bertujuan mengembangkan sistem pengelompokan prestasi siswa di SD N Sanggang 03 dengan memanfaatkan metode K-Means Clustering. Metode ini dipilih karena efisien dalam mengolah data numerik berskala besar serta memiliki tingkat presisi yang lebih tinggi dibandingkan metode alternatif seperti Hierarchical Clustering atau K-Nearest Neighbor [5].

II. TINJAUAN PUSTAKA

2.1 Data Mining

Data mining merupakan serangkaian proses untuk mengekstrak informasi berharga dari kumpulan data yang besar, yang sebelumnya tidak dapat diidentifikasi secara manual [7]. Proses ini menggabungkan metode statistik, *machine learning*, dan *database systems* untuk mendukung pengambilan keputusan berbasis data [8]. Dalam konteks pendidikan, data mining banyak digunakan untuk menganalisis perilaku belajar, mengevaluasi prestasi akademik, dan mengelompokkan siswa berdasarkan kriteria tertentu [9].

2.2 K-Means Clustering

K-Means Clustering merupakan metode analisis klaster non-hirarki yang membagi data

DOI : <https://doi.org/10.30646/tikomsin.v13i2.1007>

ISSN Online : 2620-7532

ke dalam sejumlah kelompok (*cluster*) berdasarkan kedekatan jarak antar data terhadap pusat klaster (*centroid*) [10]. Algoritma K-Means bekerja secara iteratif dengan langkah-langkah sebagai berikut [11]:

1. Menentukan jumlah cluster (K) yang diinginkan.
2. Memilih pusat cluster secara acak dari data yang ada.
3. Menghitung jarak setiap data terhadap pusat cluster menggunakan rumus Euclidean Distance.

$$D(x, y) = \sqrt{\sum (x_i - y_i)^2}$$

4. Menentukan keanggotaan cluster berdasarkan jarak terdekat.
5. Memperbarui pusat cluster berdasarkan rata-rata posisi data dalam cluster.

$$C_j = \frac{1}{n_j} \sum x_i$$

6. Mengulangi proses hingga tidak ada lagi perubahan signifikan pada pusat cluster [12].

Metode ini dipilih karena efisien dalam mengolah data numerik berskala besar serta memiliki tingkat akurasi yang baik dalam pengelompokan data [13].

2.3 MySQL

MySQL adalah sistem manajemen basis data (*Database Management System / DBMS*) yang menggunakan bahasa SQL (*Structured Query Language*) sebagai dasar pengelolaan datanya [14]. MySQL bersifat open-source dan mendukung pengolahan data secara cepat, aman, dan dapat diintegrasikan dengan berbagai bahasa pemrograman, seperti PHP, untuk membangun aplikasi berbasis web [15].

2.4 Modified Partition Coefficient (MPC)

Modified Partition Coefficient (MPC) merupakan salah satu metode untuk mengukur validitas hasil klasterisasi. MPC digunakan untuk menilai kualitas pembagian data ke dalam cluster yang telah dibentuk. Nilai MPC berada pada rentang 0 hingga 1, di mana nilai mendekati 1 menunjukkan kualitas klasterisasi yang lebih baik [16].

$$MPC = \frac{1}{n} \sum \sum u_{ij}^2$$

dengan

n : jumlah data

c : Jumlah cluster

u_{ij} : Derajat keanggotaan data ke- i pada cluster ke- j

III. METODE PENELITIAN

3.1 Analisis Sistem

Analisis sistem dilakukan untuk memahami kebutuhan, permasalahan, dan rancangan solusi dalam membangun aplikasi pengelompokan siswa berprestasi di SD Negeri Sanggang 03. Tahapan analisis sistem ini mencakup beberapa aspek penting sebagai berikut:

3.1.1 Identifikasi Masalah

Proses penentuan siswa berprestasi sebelumnya dilakukan secara manual oleh wali kelas dengan mempertimbangkan nilai akademik, partisipasi ekstrakurikuler, dan kehadiran siswa. Proses manual ini memiliki beberapa kelemahan, antara lain:

- Rentan terhadap kesalahan perhitungan karena dilakukan secara subjektif.
- Memerlukan waktu lama untuk mengolah data, terutama jika jumlah siswa semakin meningkat.
- Tidak ada sistem yang dapat menyimpan dan mengelompokkan data secara terstruktur.

3.1.2 Analisis Kebutuhan Sistem

Berdasarkan identifikasi masalah, sistem yang dikembangkan harus memenuhi kebutuhan berikut:

- Input Data Siswa: Meliputi NIS, Nama, Nilai Tugas, Nilai UTS, Nilai UAS, Keikutsertaan Ekstrakurikuler, dan Jumlah Ketidakhadiran.
- Proses Pengelompokan: Sistem dapat memproses data numerik dan mengelompokkan siswa dalam kategori prestasi (sangat baik, baik, cukup).
- Laporan Hasil Pengelompokan: Menampilkan hasil dalam bentuk tabel atau grafik yang dapat dicetak.
- Manajemen Data: Menyediakan fitur tambah, ubah, dan hapus data siswa.

3.1.3 Analisis Proses Bisnis

Proses bisnis sistem dirancang agar sesuai dengan alur kerja wali kelas dan pihak sekolah:

1. Admin (wali kelas) melakukan login ke sistem.
2. Data siswa diinput atau diperbarui pada basis data.
3. Sistem memproses data sesuai parameter prestasi yang telah ditentukan.
4. Hasil pengelompokan ditampilkan dalam dashboard aplikasi.
5. Laporan dapat dicetak sebagai bahan evaluasi.

3.1.4 Analisis Kelayakan

DOI : <https://doi.org/10.30646/tikomsin.v13i2.1007>

ISSN Online : 2620-7532

Analisis kelayakan dilakukan dalam tiga aspek utama:

- Kelayakan Teknis: Sistem berbasis web menggunakan PHP dan MySQL dapat diimplementasikan pada server sekolah dengan spesifikasi minimum (RAM 4 GB, prosesor Dual-Core).
- Kelayakan Ekonomis: Biaya pengembangan relatif rendah karena menggunakan perangkat lunak open-source.
- Kelayakan Operasional: Wali kelas dapat dengan mudah mengoperasikan sistem tanpa pelatihan khusus karena antarmuka dirancang sederhana.

3.2 Pengumpulan Data

Pengumpulan data dilakukan melalui beberapa metode berikut:

- Observasi – Peninjauan langsung ke lokasi penelitian untuk memperoleh data faktual mengenai prestasi siswa serta sistem evaluasi yang digunakan.
- Wawancara – Dilakukan dengan pihak sekolah, termasuk wali kelas dan kepala sekolah, untuk mendapatkan informasi mendalam mengenai kriteria penilaian prestasi.
- Studi Literatur – Mengkaji sumber-sumber akademis, seperti buku, artikel ilmiah, dan jurnal terkait pengelompokan data dan sistem informasi pendidikan.

3.3 Metode Pengembangan Sistem

Pengembangan sistem dilakukan menggunakan model Software Development Life Cycle (SDLC) dengan pendekatan Waterfall karena sesuai untuk proyek dengan spesifikasi kebutuhan yang jelas. Tahapan Waterfall meliputi:

1. Analisis Kebutuhan
Mengidentifikasi kebutuhan sistem, termasuk kebutuhan fungsional seperti pengelompokan data siswa, penyimpanan basis data, dan penyajian informasi berbasis web.
2. Perancangan Sistem (System Design)
Menentukan rancangan basis data, arsitektur sistem, antarmuka pengguna, serta pemodelan proses menggunakan UML agar lebih mudah dipahami oleh pengembang.
3. Implementasi (Coding)
Proses pengkodean sistem dilakukan menggunakan bahasa pemrograman berbasis web (misalnya PHP) dan sistem manajemen basis data MySQL untuk

menyimpan data siswa.

4. Pengujian Sistem (System Testing)

Pengujian sistem dilakukan menggunakan metode black box testing untuk memastikan fungsi sistem berjalan sesuai kebutuhan pengguna tanpa memperhatikan kode program secara internal. Fokus pengujian adalah pada kesesuaian input dan output sistem dengan spesifikasi yang telah ditentukan. Beberapa kategori kasus uji yang digunakan antara lain:

1. Data valid (normal case): Memasukkan data prestasi siswa lengkap dan sesuai format untuk memastikan sistem dapat melakukan proses clustering dengan benar.
2. Data kosong (empty case): Menguji respons sistem ketika input tidak diisi, untuk melihat apakah sistem memberikan peringatan atau pesan error yang tepat.
3. Data tidak valid (invalid case): Menguji input dengan format salah (misalnya nilai bukan angka) untuk memastikan sistem menolak data dan menampilkan pesan kesalahan.
4. Batasan input (boundary case): Menguji input pada kondisi minimum dan maksimum (misalnya nilai 0 atau 100) untuk melihat apakah sistem tetap dapat memproses dengan benar.

Skala pengujian dilakukan dalam dua tahap. Pertama, pengujian awal oleh pengembang untuk memastikan setiap fungsi berjalan sesuai rancangan. Kedua, pengujian langsung melibatkan pengguna (wali kelas) melalui User Acceptance Testing (UAT). Tahap ini bertujuan menilai apakah sistem benar-benar mudah digunakan, sesuai kebutuhan, dan membantu dalam pengelompokan prestasi siswa.

5. Pemeliharaan Sistem (Maintenance)

Tahap ini mencakup perbaikan bug, pembaruan fitur, dan penyesuaian sistem sesuai kebutuhan sekolah di masa mendatang.

3.4 Desain Sistem

Desain sistem mencakup pemodelan struktur aplikasi, alur proses, serta mekanisme pengelompokan siswa menggunakan Unified Modeling Language (UML) dan metode K-Means. Beberapa elemen yang digunakan antara lain:

1. Use Case Diagram
Menggambarkan interaksi antara pengguna (wali kelas/admin) dengan sistem. Use case mencakup: login, input data siswa, proses pengelompokan, melihat hasil, dan mencetak laporan.
2. Activity Diagram
Menjelaskan alur aktivitas mulai dari input data, proses analisis, hingga penyajian output berupa pengelompokan siswa.
3. Class Diagram
Menggambarkan struktur data pada sistem, termasuk atribut siswa seperti Nama, NIS, Nilai Tugas, Nilai UTS, Nilai UAS, Ekstrakurikuler, dan Jumlah Ketidakhadiran.
4. Arsitektur Sistem
Sistem dirancang berbasis web dengan arsitektur client-server. Browser bertindak sebagai klien, sedangkan server menjalankan aplikasi dan basis data.
5. Desain Basis Data
Basis data disusun dalam bentuk tabel relasional untuk menyimpan data siswa, hasil pengelompokan, dan riwayat laporan. Normalisasi dilakukan untuk menghindari redundansi data.
6. Penentuan Jumlah Kelompok (K)
Sistem mengelompokkan siswa ke dalam tiga kategori prestasi: *sangat baik*, *baik*, dan *cukup*. Penetapan jumlah cluster ($K=3$) tidak dilakukan secara arbitrer, tetapi divalidasi menggunakan metode evaluasi seperti Elbow Method atau Silhouette Analysis. Dengan pendekatan ini, nilai K yang dipilih terbukti optimal secara matematis sekaligus relevan secara akademis.
7. Interpretasi Hasil Clustering
Output K-Means berupa pembagian siswa ke dalam tiga cluster. Untuk memberi label pada cluster, sistem menghitung nilai rata-rata dari setiap cluster pada atribut akademik (nilai tugas, UTS, UAS) dan non-akademik (ekstrakurikuler, kehadiran). Cluster dengan nilai rata-rata tertinggi ditetapkan sebagai kelompok *sangat baik*, cluster menengah sebagai *baik*, dan cluster dengan rata-rata terendah sebagai *cukup*. Dengan mekanisme ini, hasil pengelompokan tidak hanya berbasis algoritme, tetapi juga memiliki interpretasi yang jelas dan logis bagi guru.

IV. HASIL DAN PEMBAHASAN

Penelitian ini menggunakan data dari 44 siswa SD Negeri Sanggang 03. Sebelum proses iterasi dalam algoritma K-Means dilakukan, terlebih dahulu dilakukan tahap inisialisasi untuk memberikan nilai awal (default) yang diperlukan dalam perhitungan. Inisialisasi ditentukan berdasarkan nilai atau jumlah dari masing-masing variabel penelitian yang mencakup aspek akademik maupun non-akademik.

4.1 Algoritma K-Means

Variabel yang digunakan dalam sistem pengelompokan siswa berprestasi meliputi: Nama, Nilai Tugas, Nilai Ujian Tengah Semester (UTS), Nilai Ujian Akhir Semester (UAS), Keikutsertaan Ekstrakurikuler, Jumlah Ketidakhadiran. Setelah inisialisasi dilakukan, pusat cluster (centroid) awal ditentukan secara acak berdasarkan data hasil inisialisasi dari setiap cluster. Data centroid awal penelitian disajikan pada Tabel 1.

Tabel 1 Centroid Awal

Cluster	Tugas	UTS	UAS	Ekstra	Absensi
C1	1	1	1	2	2
C2	2	3	2	2	4
C3	4	4	4	1	4

Data pada baris ke-1 dengan nama Aprilian Rizki Mutha dihitung dengan titik centroid ke-1 sebagai C1, centroid ke-2 sebagai C2 dan centroid ke-3 sebagai C3,

$$C_1 = \sqrt{(3-1)^2 + (4-1)^2 + (4-1)^2 + (1-2)^2 + (4-2)^2}$$

$$C_1 = 5,1962$$

$$C_2 = \sqrt{(3-2)^2 + (4-3)^2 + (4-2)^2 + (1-2)^2 + (4-4)^2}$$

$$C_2 = 2,6458$$

$$C_3 = \sqrt{(3-4)^2 + (4-4)^2 + (4-4)^2 + (1-1)^2 + (4-4)^2}$$

$$C_3 = 1$$

Kemudian diperoleh hasil perhitungan jarak ecludian yang bisa pada table 2.

Tabel 2 Hasil iterasi 1

No	Nama	Jarak	Cluster
1	Siswa 01	1.0000	C3
2	Siswa 02	1.0000	C3
3	Siswa 03	0.0000	C3
4	Siswa 04	2.8284	C1
5	Siswa 05	2.4495	C2
6	Siswa 06	2.0000	C3
7	Siswa 07	1.4142	C3
8	Siswa 08	1.4142	C3
9	Siswa 09	2.2361	C1
10	Siswa 10	1.7321	C2
11	Siswa 11	0.0000	C3
12	Siswa 12	1.4142	C3
13	Siswa 13	2.2361	C3
14	Siswa 14	1.4142	C2
15	Siswa 15	1.0000	C2
16	Siswa 16	2.0000	C2
17	Siswa 17	1.0000	C1

No	Nama	Jarak	Cluster
18	Siswa 18	1.7321	C3
19	Siswa 19	1.0000	C3
....			
44	Siswa 44	2.0000	C1

Setelah diketahui anggota setiap *cluster* kemudian pusat *cluster* baru dihitung berdasarkan data anggota tiap-tiap *cluster* sesuai dengan rumus pusat setiap *cluster*. Sehingga akan didapatkan perhitungan sebagai berikut:

- Cluster 1 berisi 15 siswa (data ke-4, 9, 17, 25, 27, 30, 31, 33, 34, 35, 36, 38, 41, 42, dan 44). Rata-rata nilai tugas, UTS, dan UAS pada klaster ini lebih tinggi dibandingkan dua klaster lainnya, dengan tingkat kehadiran yang baik serta keterlibatan aktif dalam kegiatan ekstrakurikuler. Oleh karena itu, klaster ini diinterpretasikan sebagai kelompok “sangat baik”.
- Cluster 2 berisi 17 siswa (data ke-5, 10, 14, 15, 16, 20, 21, 23, 24, 26, 28, 29, 32, 37, 39, 40, dan 43). Nilai rata-rata akademik siswa dalam klaster ini berada pada tingkat menengah. Kehadiran relatif baik, meskipun partisipasi dalam ekstrakurikuler tidak setinggi klaster pertama. Dengan demikian, klaster ini dikategorikan sebagai kelompok “baik”.
- Cluster 3 berisi 12 siswa (data ke-1, 2, 3, 6, 7, 8, 11, 12, 13, 18, 19, dan 22). Karakteristik siswa dalam klaster ini menunjukkan nilai rata-rata tugas, UTS, dan UAS yang lebih rendah dibandingkan dua klaster lainnya. Tingkat kehadiran juga cenderung lebih rendah. Oleh karena itu, klaster ini ditetapkan sebagai kelompok “cukup”.

Dengan demikian akan didapatkan pusat cluster baru yang dapat dilihat pada table 3. Tabel 3 Data Pusat Cluster Baru dari iterasi ke-2

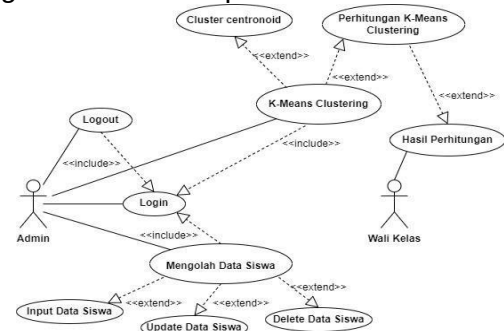
Pusat Cluster Baru					
C1	1.200	1.133	1.133	1.933	2.933
C2	2.647	2.176	2.471	2.000	3.882
C3	3.583	3.667	3.667	1.333	3.500

Selanjutnya proses iterasi berhenti pada iterasi ke-3, dikarenakan sudah tidak ada data yang berpindah atau berubah (konvergen). Cluster pertama terdapat 16 data yaitu data ke-4, 9, 16, 17, 25, 27, 30, 31, 33, 34, 35, 36, 38, 41, 42, 44. Cluster kedua terdapat 16 data yaitu data ke-5, 10, 14, 15, 20, 21, 23, 24, 26, 28, 29, 32, 37, 39, 40, 43. Cluster ketiga terdapat 12 data yaitu data ke-1, 2, 3, 6, 7, 8, 11, 12, 13,

18, 19, dan 22.

4.2 Use Case Diagram

Use case diagram pada sistem ini menggambarkan kegiatan-kegiatan pada sistem yang akan berjalan nantinya. *Use Case Diagram* bisa dilihat pada Gambar 1.

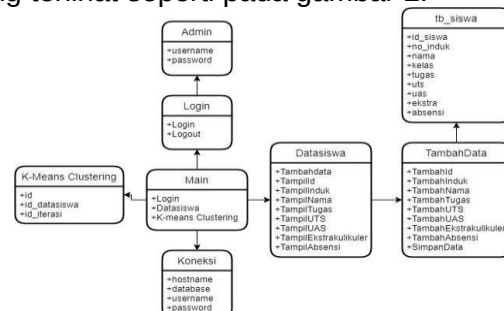


Gambar 1 Use Case Diagram

Berdasarkan Gambar 1, admin dapat masuk ke halaman proses perhitungan *k-means clustering*, setelah masuk ke proses *k-means clustering*, admin dapat melakukan input centroid awal dan melakukan proses perhitungan. Kemudian admin masuk ke halaman hasil dan dapat melanjutkan untuk melihat hasil dari perhitungan algoritma *k-means clustering*. Dan setelah semua proses selesai, admin dapat melakukan *logout* dari sistem. Sedangkan wali kelas hanya dapat melihat hasil data yang sudah dikelompokkan.

4.3 Class Diagram

Class Diagram menggambarkan struktur dan deskripsi *class* hubungan antar *class* atau data yang ada pada sistem. Sistem pengelompokan siswa berprestasi ini menggunakan 3 tabel yang saling terelasi. Berikut ini class diagram sistem aplikasi berbasis web pengelompokan siswa berprestasi di SD N Sanggang 03 yang menggunakan metode *K-Means Clustering* yang terlihat seperti pada gambar 2.



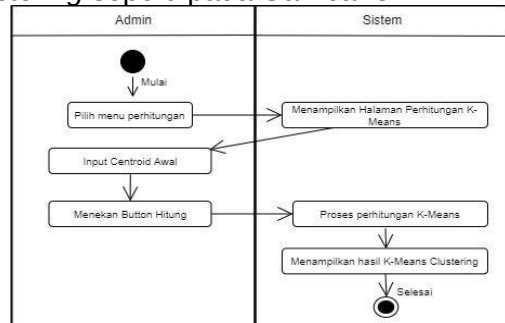
Gambar 2 Class Diagram

4.4 Activity Diagram

Admin memulai aktivitas dengan memilih menu perhitungan kemudian sistem akan menampilkan halaman *k-means clustering* dan admin memasukkan centroid awal

DOI : <https://doi.org/10.30646/tikomsin.v13i2.1007>

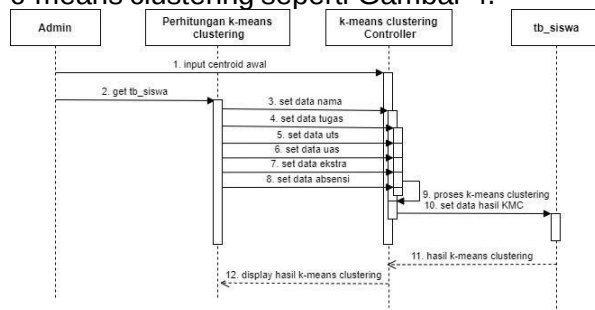
kemudian menekan button hitung, maka secara otomatis sistem akan melakukan proses *k-means* clustering dan admin akan memperoleh hasil clustering. Berikut gambar activity diagram perhitungan *K-Means clustering* seperti pada Gambar 3.



Gambar 3 Activity Diagram

4.5 Sequence Diagram

Menjelaskan alur proses kegiatan pada sistem sesuai urutan waktu dan menjelaskan penyampaian pesan dari setiap proses. Berikut sequence diagram perhitungan fuzzy c-means clustering seperti Gambar 4.



Gambar 4 Sequence Diagram

4.6 Pengujian Sistem

a. Pengujian Fungsionalitas

Pengujian fungsionalitas menggunakan metode Black Box. Dalam penelitian ini dilakukan uji fungsionalitas mulai dari proses melakukan login, menambah data, mengedit data, menghapus data. Berikut ini uji fungsionalitas pada menu pengelompokkan untuk melakukan perhitungan data dengan algoritma *K-Means Clustering* seperti yang terlihat pada Tabel 4.

Tabel 4 Uji Fungsionalitas Menu Perhitungan

Kasus dan Hasil Uji Coba			
Test Case	Yang Diharapkan	Pengamatan	Kesimpulan
Menginputkan data centroid awal	Data akan diproses dengan menggunakan algoritma <i>k-means clustering</i>	Menampilkan hasil perhitungan	Sesuai

b. Pengujian Validitas Cluster

Untuk menguji validitas dari metode *K-Means Clustering* yang diterapkan pada

ISSN Online : 2620-7532

program ini, penelitian ini menggunakan metode *Modified Partition Coefficient* (MPC). MPC merupakan metode yang digunakan untuk menguji validitas jumlah cluster. Pada umumnya jumlah cluster yang paling optimal ditentukan dari nilai MPC yang paling besar. Berikut hasil pengujian *Modified Partition Coefficient* yang ditunjukkan pada tabel 5.

Tabel 5. Hasil MPC

Jumlah Cluster	Hasil MPC
3	0,859655847
5	0,666294941

Dari tabel 5 dapat dilihat bahwa setelah dilakukan perbandingan, bahwa nilai tertinggi MPC pada saat menggunakan jumlah Cluster 3. Dari nilai hasil yang didapat bahwa jumlah cluster yang paling optimal ditentukan berdasarkan dari nilai interval $0 \leq \text{MPC} \leq 1$, sehingga dapat disimpulkan bahwa nilai MPC jika mendekati nilai 1, maka hasil cluster dinyatakan optimal atau kualitas cluster dinyatakan berada di cluster yang tepat.

V. PENUTUP

Penelitian ini berhasil merancang aplikasi berbasis web untuk mengelompokkan prestasi siswa di SD Negeri Sanggang 03 menggunakan metode *K-Means Clustering*. Sistem membagi data menjadi tiga kluster yang diinterpretasikan sebagai kategori sangat baik (15 siswa), baik (17 siswa), dan cukup (12 siswa). Variabel yang digunakan meliputi Nama, Nilai Tugas, Nilai UTS, Nilai UAS, Keikutsertaan Ekstrakurikuler, dan Jumlah Ketidakhadiran.

Hasil validasi menggunakan *Modified Partition Coefficient* (MPC) menunjukkan nilai tertinggi sebesar 0,8596 pada tiga kluster, menandakan kualitas pengelompokan yang optimal. Analisis karakteristik setiap kluster memperlihatkan bahwa kluster sangat baik memiliki rata-rata nilai akademik tertinggi dan tingkat kehadiran yang stabil; kluster baik berada pada tingkat menengah; sedangkan kluster cukup menunjukkan nilai akademik serta kehadiran yang lebih rendah dibandingkan kluster lainnya.

Sistem ini secara efektif mendukung wali kelas dalam melakukan evaluasi prestasi siswa berdasarkan data akademik dan non-akademik. Selain itu, penelitian ini membuka peluang pengembangan lebih lanjut, misalnya penambahan fitur pencetakan laporan, visualisasi grafik interaktif, serta dukungan akses multiuser dan perangkat mobile.

DOI : <https://doi.org/10.30646/tikomsin.v13i2.1007>

Penelitian selanjutnya juga dapat mengeksplorasi metode clustering lain sebagai pembanding untuk meningkatkan kualitas hasil.

DAFTAR PUSTAKA

- [1] T. S. Sujana, R. Astuti, W. Prihartono, and R. Hamonangan, "Implementation of Data Mining Using the K-Means Algorithm to Group Students Based on Academic Performance," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 2, pp. 1527–1531, 2025.
- [2] S. Abdul Mukhsyi, A. I. Purnamaari, and A. Bahtiar, "Improving Student Achievement Clustering Model Using K-Means Algorithm in Pasundan Majalaya Vocational School," *Journal of Artificial Intelligence and Engineering Applications (JAIEA)*, vol. 4, no. 2, 2025.
- [3] A. P. Riani, A. Voutama, and T. Ridwan, "Penerapan K-Means Clustering Dalam Pengelompokan Hasil Belajar Peserta Didik Dengan Metode Elbow," *Journal of Information System Research (JOSH)*, vol. 6, no. 3, pp. 1705–1714, 2025.
- [4] Dewi A., N. Wulandari, Riski Annisa, Lestari Yusuf, and dan Titin Prihatin, "Educational Data Mining for Student Academic Prediction Using K-Means Clustering and Naïve Bayes Classifier," *Jurnal Pilar Nusa Mandiri*, vol. 16, no. 2, pp. 155–160, 2020.
- [5] K. Setiawan and Y. Y. A. Saputry, "Clustering Data Calon Siswa Baru Menggunakan Metode K-Means di Pusat Pengembangan Anak Fajar Baru Cengkareng," *Jurnal JTIK*, vol. 8, no. 1, pp. 75–83, 2024.
- [6] Mochamad H. and T. Utomo, "Clustering Data Siswa Putus Sekolah dengan Algoritma K-Means dan DBSCAN," *Jurnal Komputer dan Informatika*, vol. 18, no. 2, 2023.
- [7] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA, USA: Morgan Kaufmann, 2017.
- [8] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Mag*, vol. 17, no. 3, pp. 37–54, 2016.
- [9] R. Agrawal and R. Srikant, "Educational Data Mining: Techniques for Student Performance Evaluation," *International Journal of Data Science*, vol. 5, no. 2, pp. 101–110, 2020.
- [10] A. K. Jain, "Data Clustering: 50 Years Beyond K-Means," *Pattern Recognit Lett*, vol. 31, no. 8, pp. 651–666, 2019.
- [11] R. T. Vulandari, *Data Mining: Teori dan Aplikasi Rapidminer*. Yogyakarta: CV. Gava Media, 2017.
- [12] Y. Zhao and G. Karypis, "Criterion Functions for Document Clustering: Experiments and Analysis," *Machine Learning Journal*, vol. 6, no. 4, pp. 311–331, 2021.
- [13] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*, 2nd ed. New York, NY, USA: Springer, 2021.
- [14] P. DuBois, *MySQL*, 5th ed. Indianapolis, IN, USA: Addison-Wesley, 2018.
- [15] A. F. Abdullah and N. A. Rahman, "Implementation of Web-Based Student Data Management System Using PHP and MySQL," *Journal of Information Technology and Computer Science*, vol. 14, no. 1, pp. 35–42, 2022.
- [16] S. Modha and W. S. Spangler, "Clustering Validity Indices for High-Dimensional Data: An Empirical Study," *IEEE Trans Pattern Anal Mach Intell*, vol. 27, no. 3, pp. 473–482, 2020.