

PERBANDINGAN METODE K-MEANS DAN HIERARCHICAL CLUSTERING DALAM MENGELOMPOKKAN KOMPETENSI PRAKTIK SISWA TSM SEBAGAI DASAR EVALUASI PEMBELAJARAN

Mochamad Ardy Burhannudin^{1*}, Nuramaliyah²

^{1,2}Program Studi Statistika, Universitas Terbuka, Jakarta, Indonesia

*Penulis korespondensi: ardymoch@gmail.com

ABSTRAK

Evaluasi kompetensi praktik siswa SMK secara konvensional masih bersifat subjektif sehingga program intervensi pembelajaran tidak selalu tepat sasaran. Penelitian ini membandingkan metode K-Means Clustering dan Hierarchical Clustering dalam mengelompokkan kompetensi praktik siswa Teknik Sepeda Motor (TSM) SMK Baitul Izza sebagai dasar evaluasi pembelajaran. Data sekunder berupa enam variabel kompetensi praktik diperoleh dari 120 siswa kelas XI dan XII TSM. Jumlah cluster optimal ditentukan dengan Elbow Method dan kinerja kedua metode dibandingkan menggunakan Silhouette Index. Hasil analisis menggunakan software R menunjukkan bahwa K=3 merupakan jumlah cluster optimal. K-Means menghasilkan tiga cluster dengan 45 siswa berkompetensi tinggi, 32 siswa berkompetensi sedang, dan 43 siswa berkompetensi rendah. Hierarchical Clustering menghasilkan 45 siswa berkompetensi tinggi, 30 siswa berkompetensi sedang, dan 45 siswa berkompetensi rendah. K-Means menunjukkan kinerja sedikit lebih baik dengan Silhouette Index sebesar 0,5502 dibandingkan Hierarchical Clustering sebesar 0,5438. Meskipun selisihnya kecil, hasil ini menunjukkan bahwa K-Means lebih sesuai untuk data kompetensi praktik siswa pada penelitian ini.

Kata kunci: k-means clustering, hierarchical clustering, kompetensi praktik, evaluasi pembelajaran, silhouette index

1 PENDAHULUAN

Sekolah Menengah Kejuruan (SMK) berperan strategis dalam menyiapkan tenaga kerja terampil, khususnya pada jurusan Teknik Sepeda Motor (TSM) yang permintaannya terus meningkat seiring pertumbuhan kendaraan bermotor di Indonesia. Keberhasilan pembelajaran TSM sangat ditentukan oleh kompetensi praktik siswa yang mencakup servis berkala, perbaikan mesin, sistem kelistrikan, dan keselamatan kerja. Namun, evaluasi kompetensi praktik yang dilakukan secara konvensional masih bersifat subjektif dan tidak terstruktur, sehingga pengelompokan kemampuan siswa tidak didasarkan pada analisis data yang objektif dan menyeluruh (Sugiyono, 2019).

Perkembangan teknik data mining membuka peluang pengelompokan data akademik secara lebih objektif. Clustering merupakan teknik data mining yang mengelompokkan data berdasarkan kemiripan karakteristik tanpa memerlukan label kelas sebelumnya (Han, Kamber, & Pei, 2012). K-Means Clustering merupakan metode yang paling banyak digunakan karena sederhana dan efisien (Hastie, Tibshirani, & Friedman, 2009), sedangkan Hierarchical Clustering menghasilkan visualisasi hierarki cluster yang informatif melalui dendrogram (Kaufman & Rousseeuw, 2005). Penerapan clustering pada data akademik siswa dikenal sebagai Educational Data Mining yang terbukti efektif mendukung pengambilan keputusan berbasis data (Tan, Steinbach, & Kumar, 2019).

Beberapa penelitian terdahulu membuktikan efektivitas metode clustering pada data siswa. Hutagalung (2022) berhasil memetakan siswa kelas unggulan menggunakan K-Means dengan hasil yang representatif. Kurniawan dkk. (2025) dan Fauzie dkk. (2023) menerapkan K-Means untuk mengelompokkan nilai siswa SMK menjadi tiga cluster kompetensi sebagai

dasar rekomendasi pembelajaran. Dewi, Defit, dan Yuhandri (2021) membuktikan bahwa K-Means efektif memetakan kelompok belajar siswa menuju prestasi dengan tingkat akurasi yang baik. Hendrastuty (2024) serta Ishak dan Amiruddin (2024) turut mengonfirmasi bahwa K-Means dapat diandalkan untuk mengevaluasi hasil pembelajaran dan mengelompokkan prestasi akademik siswa secara objektif. Srirahmawati dkk. (2025) juga menunjukkan bahwa pengelompokan prestasi akademik menggunakan K-Means menghasilkan cluster yang bermakna sebagai panduan pembelajaran. Ripai, Baskoro, dan Imelda (2025) membandingkan K-Means dan Hierarchical Clustering pada data nilai siswa SMK dan menemukan K-Means sebagai metode terbaik berdasarkan Silhouette Index. Novaldi dan Wijayanto (2023) serta Abdulpatah dkk. (2025) menegaskan bahwa perbandingan dua metode clustering menghasilkan analisis yang lebih komprehensif. Putra dan Fadhillah (2025) juga membuktikan bahwa kinerja K-Means dan Hierarchical Clustering dapat bervariasi tergantung karakteristik data, sehingga perbandingan keduanya diperlukan untuk menentukan metode yang paling sesuai.

Penelitian ini mengangkat studi kasus di SMK Baitul Izza dengan menggunakan enam variabel kompetensi praktik dari 120 siswa kelas XI dan XII TSM. Penelitian ini bertujuan untuk mengelompokkan kompetensi praktik siswa Teknik Sepeda Motor SMK Baitul Izza menggunakan metode K-Means Clustering dan Hierarchical Clustering, membandingkan kinerja kedua metode berdasarkan Silhouette Index, serta merumuskan rekomendasi evaluasi pembelajaran berbasis data.

2 METODE

2.1 Cakupan Penelitian

Penelitian ini merupakan penelitian kuantitatif deskriptif komparatif dengan pendekatan data mining. Penelitian dilaksanakan di SMK Baitul Izza pada semester ganjil tahun pelajaran 2025/2026. Populasi adalah seluruh siswa TSM kelas XI dan XII SMK Baitul Izza berjumlah 120 siswa (4 rombongan belajar). Teknik sampling yang digunakan adalah total sampling, sehingga seluruh populasi dijadikan sampel.

Tabel 1. Variabel penelitian

Kode	Variabel	Skala	Sumber
X1	Nilai Praktik Harian	Rasio	Rekap nilai
X2	Kehadiran Bengkel (%)	Rasio	Absensi
X3	Nilai Jobsheet / Laporan Praktik	Rasio	Arsip guru
X4	Nilai Ujian Praktik Akhir	Rasio	Dok. ujian
X5	Nilai Teori / Pengetahuan TSM	Rasio	Rekap nilai
X6	Nilai Sikap / Kedisiplinan	Rasio	Wali kelas

Sumber: (SMK Baitul Izza, 2025)

2.2 Alat Analisis

Analisis data dilakukan menggunakan software R versi 4.5.3 dengan antarmuka RStudio Desktop. Package khusus yang digunakan meliputi cluster (v2.1.6) untuk analisis clustering dan Silhouette Index, factoextra (v1.0.7) untuk visualisasi Elbow Method dan hasil clustering, serta ggplot2 (v3.5.1) dan dplyr (v1.1.4) untuk visualisasi dan manipulasi data.

2.3 Tahapan Analisis Data

Analisis dilakukan melalui lima tahapan berurutan sebagai berikut.

2.3.1. Preprocessing Data

Dilakukan pengecekan missing value pada seluruh variabel dilanjutkan normalisasi Min-Max untuk menyeragamkan skala data ke rentang [0,1] Normalisasi diperlukan agar variabel dengan satuan berbeda tidak mendominasi perhitungan jarak antar data (Han, Kamber, & Pei, 2012). Rumus normalisasi Min-Max yang digunakan adalah:

$$X' = \frac{X - X_{min}}{X_{max} - X_{min}}$$

Keterangan:

X' = nilai normalisasi

X = nilai asli

X_{min} = nilai minimum

X_{max} = nilai maksimum.

2.3.2. Elbow Method

Penentuan jumlah cluster optimal K dengan menghitung nilai Within-Cluster Sum of Squares (WCSS) untuk K=1 hingga K=10. Nilai K pada titik siku grafik WCSS ditetapkan sebagai K optimal karena penambahan cluster di atas nilai tersebut tidak memberikan penurunan WCSS yang signifikan (Hastie, Tibshirani, & Friedman, 2009; Fauji & Farokhah, 2025).

2.3.3. K-Means Clustering

Analisis pengelompokan menggunakan fungsi `kmeans()` dengan parameter `nstart=25` dan `iter.max=100`. Jarak antar data dihitung menggunakan Euclidean Distance (Tan, Steinbach, & Kumar, 2019):

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

2.3.4. Hierarchical Clustering

Analisis menggunakan metode Agglomerative dengan Ward's D2 `hclust()`, divisualisasikan dalam bentuk dendrogram, dan dipotong pada K sesuai hasil Elbow Method menggunakan `cutree()`. Metode Ward's D2 dipilih karena menghasilkan cluster yang lebih kompak dibandingkan metode linkage lainnya (Kaufman & Rousseeuw, 2005).

2.3.5. Evaluasi Silhouette Index

Kinerja kedua metode dibandingkan menggunakan Silhouette Index (SI). Kinerja kedua metode dibandingkan menggunakan Silhouette Index (SI). Nilai SI untuk setiap data ke- i dihitung menggunakan rumus berikut (Kaufman & Rousseeuw, 2005):

$$s = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}$$

Keterangan:

$a(i)$ = rata-rata jarak ke sesama anggota cluster

$b(i)$ = rata-rata jarak ke anggota cluster terdekat lainnya

$s(i)$ = nilai silhouette untuk data ke- i

Nilai SI berkisar -1 hingga 1 ; nilai $> 0,50$ mengindikasikan struktur cluster yang baik. Metode dengan SI tertinggi dinyatakan lebih optimal pada tingkat kepercayaan 95% Nilai SI berkisar -1 hingga 1 . Metode dengan SI tertinggi dinyatakan lebih optimal pada tingkat kepercayaan 95%. Interpretasi nilai SI mengacu pada kriteria berikut (Kaufman & Rousseeuw, 2005):

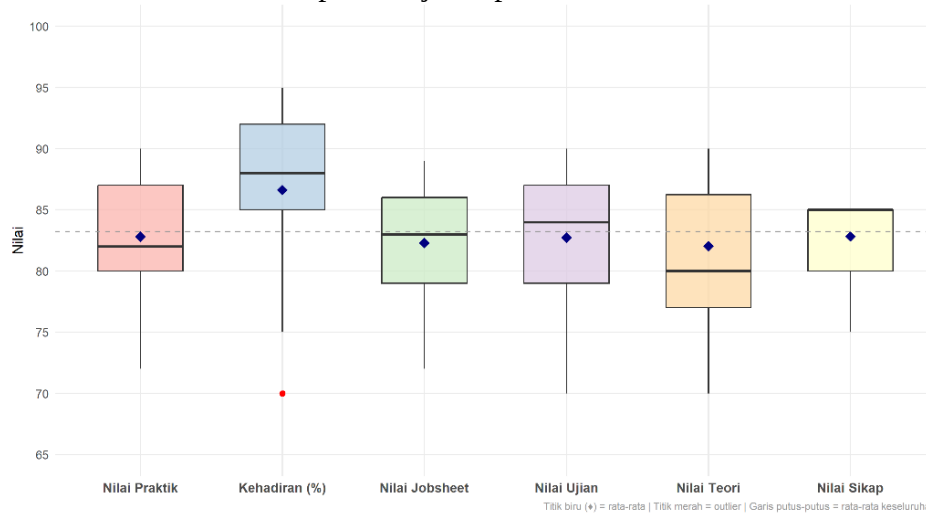
Tabel 2. Kriteria Interpretasi Silhouette Index

Rentang Nilai	Interpretasi	Kualitas
0,71 – 1,00	Struktur kuat	Sangat baik
0,51 – 0,70	Struktur cukup	Baik
0,26 – 0,50	Struktur lemah	Cukup
< 0,25	Tidak ada struktur	Kurang

3 HASIL DAN PEMBAHASAN

3.1 Statistik Deskriptif

Penelitian menggunakan data kompetensi praktik dari 120 siswa TSM SMK Baitul Izza. Hasil pengecekan menunjukkan tidak terdapat missing value pada seluruh variabel sehingga data siap dianalisis. Statistik deskriptif disajikan pada Gambar 1.

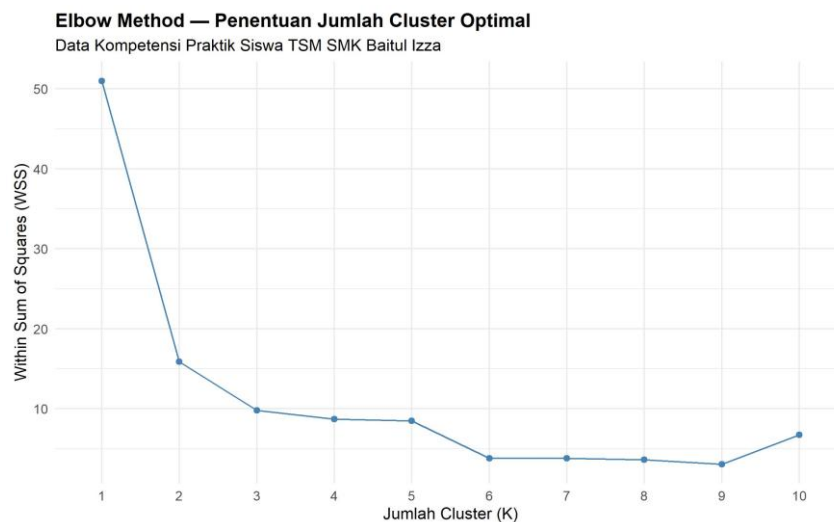


Gambar 1. Boxplot Distribusi Kompetensi Praktik Siswa TSM SMK Baitul Izza (n=120)

Berdasarkan Gambar 1, variabel kehadiran bengkel memiliki sebaran nilai paling lebar (IQR terbesar) dibandingkan variabel lainnya, mengindikasikan variasi kehadiran yang paling tinggi antar siswa. Sebaliknya, variabel nilai sikap memiliki kotak IQR paling sempit menunjukkan nilai sikap yang paling seragam. Tidak terdapat outlier yang ekstrem pada seluruh variabel.

3.2 Penentuan Jumlah Cluster Optimal

Penentuan jumlah cluster optimal dilakukan menggunakan Elbow Method dengan menghitung nilai Within Sum of Squares (WSS) untuk K=1 hingga K=10. Hasilnya disajikan pada Gambar 2.

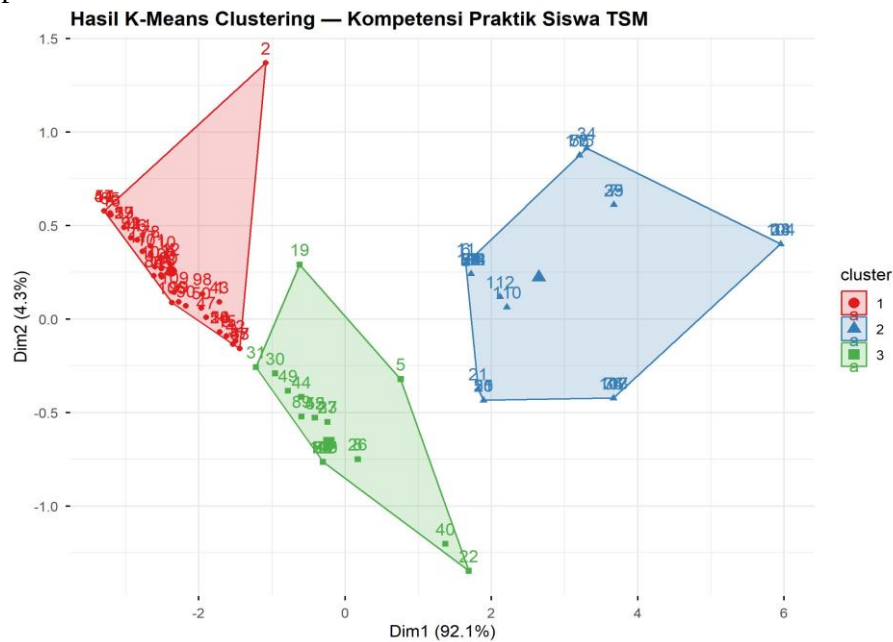


Gambar 2. Elbow Method — Penentuan Jumlah Cluster Optimal

Berdasarkan Gambar 2, penurunan nilai WSS yang signifikan terjadi dari K=1 (WSS≈51) ke K=2 (WSS≈15), kemudian berlanjut ke K=3 (WSS≈10). Setelah K=3, penurunan WSS mulai melandai membentuk sudut siku. Dengan demikian, K=3 ditetapkan sebagai jumlah cluster optimal yang digunakan pada kedua metode analisis.

3.3 Hasil K-Means Clustering

Analisis K-Means Clustering dengan K=3 menghasilkan tiga cluster kompetensi praktik siswa TSM SMK Baitul Izza. Visualisasi hasil clustering disajikan pada Gambar 3 dan profil tiap cluster pada Tabel 3.



Gambar 3. Scatter Plot Hasil K-Means Clustering

Tabel 3. Profil Cluster K-Means Clustering

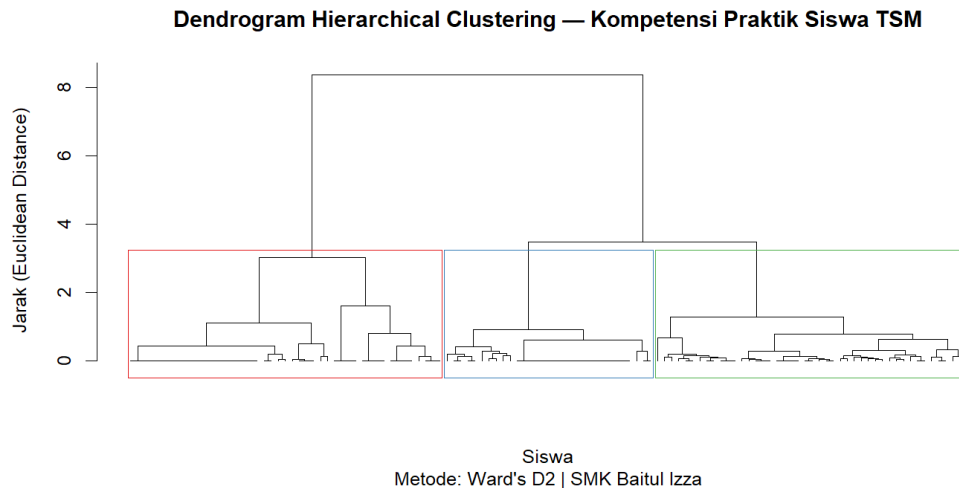
Cluster	n	Praktik	Hadir	Jobsheet	Ujian	Teori	Sikap	Kategori
1	45	87.62	93.13	86.82	87.49	87.96	84.89	Kompetensi Tinggi
2	43	78.07	79.98	77.37	77.42	76.35	79.16	Kompetensi Rendah
3	32	82.41	86.38	82.56	83.22	81.41	84.88	Kompetensi Sedang

Sumber: Output Software RStudio

Berdasarkan Tabel 3, Cluster 1 terdiri dari 45 siswa (37,5%) dengan rata-rata nilai tertinggi pada semua variabel (Nilai Praktik=87,62; Kehadiran=93,13%) dikategorikan sebagai kelompok kompetensi tinggi. Cluster 2 terdiri dari 43 siswa (35,8%) dengan rata-rata nilai terendah (Nilai Praktik=78,07; Kehadiran=79,98%) dikategorikan sebagai kompetensi rendah. Cluster 3 terdiri dari 32 siswa (26,7%) dengan nilai menengah (Nilai Praktik=82,41; Kehadiran=86,38%) dikategorikan sebagai kompetensi sedang. Visualisasi pada Gambar 3 menunjukkan ketiga cluster terpisah dengan baik, terutama pada dimensi pertama (Dim1) yang menjelaskan 92,1% variansi data.

3.4 Hasil Hierarchical Clustering

Analisis Hierarchical Clustering menggunakan metode Agglomerative dengan Ward's D2 menghasilkan dendrogram yang disajikan pada Gambar 4 dan profil cluster pada Tabel 4.



Gambar 4. Dendrogram Hierarchical Clustering (Metode Ward's D2)

Tabel 4. Profil Cluster Hierarchical Clustering

Cluster	n	Praktik	Hadir	Jobsheet	Ujian	Teori	Sikap	Kategori
1	45	87.62	93.13	86.82	87.49	87.96	84.89	Kompetensi Tinggi
2	30	82.63	86.8	82.87	83.57	81.57	84.87	Kompetensi Sedang
3	45	78.11	79.98	77.4	77.44	76.47	79.42	Kompetensi Rendah

Sumber: Output Software RStudio

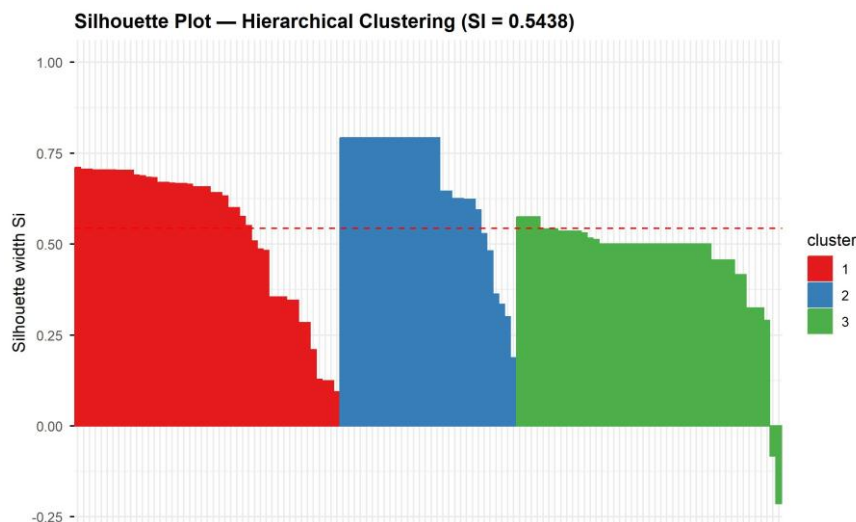
Hierarchical Clustering menghasilkan komposisi yang sedikit berbeda. Cluster 1 terdiri dari 45 siswa (37,5%) dengan profil kompetensi tinggi yang identik dengan K-Means. Cluster 2 terdiri dari 30 siswa (25,0%) dengan kompetensi sedang (Nilai Praktik=82,63; Kehadiran=86,80%). Cluster 3 terdiri dari 45 siswa (37,5%) dengan kompetensi rendah (Nilai Praktik=78,11; Kehadiran=79,98%). Dendrogram pada Gambar 4 menunjukkan pemisahan yang jelas antara kelompok kompetensi tinggi (kanan) dengan kelompok kompetensi sedang dan rendah pada jarak Euclidean sekitar 8.

3.5 Perbandingan Kinerja Metode

Perbandingan kinerja K-Means dan Hierarchical Clustering dilakukan menggunakan Silhouette Index. Visualisasi dan hasil perbandingan disajikan pada Gambar 5 Gambar 6, dan Tabel 5.



Gambar 5. Silhouette Plot K-Means Clustering (SI = 0,5502)



Gambar 6. Silhouette Plot Hierarchical Clustering (SI = 0,5438)

Tabel 5. Perbandingan Silhouette Index

Metode	Silhouette Index	Keterangan
K-Means Clustering	0,5502	Metode Terbaik
Hierarchical Clustering	0,5438	

Sumber: Output Software Rstudio

Berdasarkan Tabel 5, kedua metode menghasilkan Silhouette Index di atas 0,50 yang mengindikasikan struktur cluster yang baik. K-Means Clustering memiliki nilai Silhouette Index sedikit lebih tinggi dibandingkan Hierarchical Clustering, sehingga pada data ini K-Means dapat dipilih sebagai metode yang lebih sesuai. Namun, karena selisih nilai relatif kecil, kedua metode pada dasarnya menghasilkan kualitas cluster yang hampir sebanding. Pada Gambar 5, cluster K-Means menunjukkan sebagian besar anggota memiliki nilai silhouette positif, terutama Cluster 1 dan Cluster 3. Sedangkan pada Gambar 6, Hierarchical Clustering memiliki beberapa anggota Cluster 3 dengan nilai silhouette negatif yang mengindikasikan sedikit ketidaktepatan pengelompokan. Dengan demikian, K-Means Clustering dinyatakan

sebagai metode yang lebih optimal dalam mengelompokkan kompetensi praktik siswa TSM SMK Baitul Izza.

Terdapat perbedaan keanggotaan cluster pada sejumlah siswa antara kedua metode. Perbedaan ini terjadi pada siswa yang berada di zona perbatasan antara kompetensi sedang dan rendah, yaitu siswa dengan nilai yang tidak jauh berbeda dari rata-rata kedua cluster tersebut. K-Means mengelompokkan siswa berdasarkan jarak Euclidean ke centroid yang terus diperbarui secara iteratif, sehingga bersifat fleksibel. Sementara Hierarchical Clustering menggunakan pendekatan Ward's D2 yang bersifat hierarkis dan tidak dapat merevisi keputusan penggabungan yang sudah dibuat. Perbedaan ini merupakan hal yang wajar dan justru menjadi alasan pentingnya membandingkan dua metode sekaligus, sebagaimana dinyatakan oleh Novaldi dan Wijayanto (2023) bahwa perbandingan dua metode clustering menghasilkan perspektif yang lebih komprehensif.

3.6 Rekomendasi Evaluasi Pembelajaran

Berdasarkan hasil pengelompokan K-Means sebagai metode terbaik, rekomendasi program evaluasi pembelajaran untuk masing-masing cluster disajikan pada Tabel 6.

Tabel 6. Rekomendasi Evaluasi Pembelajaran Berdasarkan Hasil Clustering

Cluster	n	Profil Siswa	Rekomendasi Program
Cluster 1 (Tinggi)	45	Nilai & kehadiran sangat baik	Program pengayaan, kompetisi keterampilan otomotif, asisten instruktur junior
Cluster 3 (Sedang)	32	Nilai & kehadiran cukup baik	Pembinaan reguler, pendampingan intensif, mentoring teman sebaya
Cluster 2 (Rendah)	43	Nilai & kehadiran perlu peningkatan	Program remedial intensif, konseling akademik, peningkatan motivasi belajar

Sumber: Output Software RStudio

4 KESIMPULAN

Berdasarkan hasil analisis, kompetensi praktik siswa Teknik Sepeda Motor SMK Baitul Izza dapat dikelompokkan menjadi tiga cluster, yaitu kompetensi tinggi, sedang, dan rendah. Jumlah cluster optimal ditentukan menggunakan Elbow Method dengan nilai $K = 3$. Hasil K-Means Clustering membentuk kelompok kompetensi tinggi sebanyak 45 siswa, kompetensi sedang sebanyak 32 siswa, dan kompetensi rendah sebanyak 43 siswa, sedangkan Hierarchical Clustering menghasilkan kelompok kompetensi tinggi sebanyak 45 siswa, kompetensi sedang sebanyak 30 siswa, dan kompetensi rendah sebanyak 45 siswa.

Perbandingan metode menunjukkan bahwa K-Means Clustering memiliki nilai Silhouette Index sebesar 0,5502, sedikit lebih tinggi dibandingkan Hierarchical Clustering sebesar 0,5438. Dengan demikian, K-Means dapat dipilih sebagai metode yang lebih sesuai untuk data penelitian ini, meskipun kedua metode sama-sama menghasilkan struktur cluster yang tergolong baik.

Hasil pengelompokan ini dapat digunakan sebagai dasar evaluasi pembelajaran berbasis data. Kelompok kompetensi tinggi dapat diarahkan pada program pengayaan dan kompetisi keterampilan, kelompok sedang dapat memperoleh pembinaan reguler dan mentoring, sedangkan kelompok rendah memerlukan remedial intensif, penguatan motivasi belajar, serta pemantauan kehadiran dan penyelesaian jobsheet.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Program Studi Statistika Universitas Terbuka, dosen pembimbing, serta SMK Baitul Izza Tulungagung yang telah memberikan dukungan dan menyediakan data dalam penyusunan artikel ilmiah ini.

DAFTAR PUSTAKA

- Abdulpatah, T., Sari, B. N., & Susilawati. (2025). Perbandingan Algoritma K-Means dan Agglomerative Hierarchical Clustering untuk Pengelompokan Daerah Penghasil Padi di Indonesia. *JITET: Jurnal Informatika dan Teknik Elektro Terapan*, 13(3).
- Dewi, S., Defit, S., & Yuhandri, Y. (2021). Akurasi Pemetaan Kelompok Belajar Siswa Menuju Prestasi Menggunakan Metode K-Means. *Jurnal Sistim Informasi dan Teknologi*, 3(1), 28–33.
- Fauji, F., & Farokhah, L. (2025). Penerapan Data Mining Menggunakan K-Means Clustering dalam Mengelompokkan Tingkat Kesulitan Mata Pelajaran. *JOSH: Journal of Information System Research*, 6(3), 1705–1714.
- Fauzie, M. A. A., Yuliadi, & Putra, J. A. (2023). Clustering Data Menggunakan Metode K-Means untuk Rekomendasikan Pembelajaran Akademik bagi Siswa Aktif dalam Ekstrakurikuler. *KLIK: Kajian Ilmiah Informatika dan Komputer*, 4(1), 642–648.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning* (2nd ed.). Springer.
- Hendrastuty, N. (2024). Penerapan Data Mining Menggunakan Algoritma K-Means Clustering dalam Evaluasi Hasil Pembelajaran Siswa. *JIMA-ILKOM*, 3(1), 46–56.
- Hutagalung, J. (2022). Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering. *JATISI: Jurnal Teknik Informatika dan Sistem Informasi*, 9(1), 606–620.
- Ishak, R., & Amiruddin. (2024). Clustering Prestasi Akademik Lulusan Menggunakan Metode K-Means. *Jambura Journal of Electrical and Electronics Engineering*, 6(1).
- Kaufman, L., & Rousseeuw, P. J. (2005). *Finding Groups in Data: An Introduction to Cluster Analysis*. Wiley.
- Kurniawan, S., Purba, R., Manullang, S. O., & Sitorus, H. (2025). Implementasi Clustering Data Nilai Siswa Menggunakan Algoritma K-Means: Studi Kasus di SMK Nasional Namoterasi. *Global Research and Innovation Journal*, 1(3), 788–798.
- Novaldi, J., & Wijayanto, A. W. (2023). Analisis Cluster Kualitas Pemuda di Indonesia dengan Agglomerative Hierarchical dan K-Means. *Komputika: Jurnal Sistem Komputer*, 12(2).
- Putra, I. P., & Fadhillah, A. (2025). Perbandingan Metode K-Means dan Hierarchical Clustering dalam Pengelompokan Data Penduduk Miskin di Kabupaten Cianjur. *LANCAH: Jurnal Inovasi dan Tren*, 3(1), 227–234.
- Ripai, A., Baskoro, Y., & Imelda. (2025). Perbandingan Algoritma K-Means, Fuzzy C-Means, dan Hierarchical Clustering pada Klasterisasi Tingkat Pemahaman Siswa dalam Pembelajaran Berdiferensiasi. *Jurnal Algoritma*, 22(2), 692–703.
- Srirahmawati, E. O., Huda, M., & Pratama, D. (2025). Pengelompokan Prestasi Akademik Siswa SD Menggunakan Algoritma K-Means. *JIRE: Jurnal Informatika dan Rekayasa Elektronik*, 8(1), 81–86.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.
- Tan, P. N., Steinbach, M., & Kumar, V. (2019). *Introduction to Data Mining* (2nd ed.). Pearson.