



ANALISIS CLUSTERING KRIMINALITAS BERDASARKAN TINGKAT PENDIDIKAN WARGA BINAAN PEMASYARAKATAN MENGUNAKAN METODE AGGLOMERATIVE HIERARCHICAL CLUSTERING

Marhama M. Baburahman Saifuddin¹, Junaidi Noh², Sigit So ijoyo³

¹maynakahsaifuddin@gmail.com, ²junski576@gmail.com, ³Sigitsoijoyo@ummu.ac.id

^{1,2,3}Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Maluku Utara

Abstrak

Kriminalitas merupakan permasalahan sosial yang memengaruhi keamanan dan ketertiban masyarakat, dan tingkat pendidikan diduga berkaitan dengan pola kriminalitas pada Warga Binaan Pemasyarakatan (WBP). Penelitian ini menganalisis pengelompokan kriminalitas berdasarkan tingkat pendidikan WBP di Rutan Kelas II Ternate menggunakan metode *Agglomerative Hierarchical Clustering* (AHC) dan mengevaluasi kualitas cluster menggunakan *Silhouette Coefficient*. Data yang digunakan adalah data sekunder WBP tahun 2018-2023 sebanyak 1.490 data yang mencakup tingkat pendidikan, jenis tindak pidana, usia, dan tahun perkara, yang diproses melalui *preprocessing*, transformasi data, pembentukan cluster dengan AHC, visualisasi *dendrogram*, serta evaluasi menggunakan *Silhouette Coefficient*. Hasil penelitian menunjukkan AHC membentuk tiga cluster utama dengan nilai *Silhouette Coefficient* sebesar 0,3294, termasuk kategori *weak structure* yang masih dapat diterima dalam penelitian eksploratif dan menunjukkan pola pengelompokan awal pada data. Penelitian ini menghasilkan visualisasi tren cluster tahun 2018-2023 yang menggambarkan dinamika distribusi cluster dari waktu ke waktu, sehingga dapat menjadi dasar penyusunan strategi pembinaan yang lebih tepat sasaran.

Kata Kunci: *Agglomerative Hierarchical Clustering*, Kriminalitas, *Silhouette Coefficient*, Tingkat Pendidikan, Warga Binaan Pemasyarakatan

Abstract

Criminality is a social problem that affects public safety and order. Education level is considered related to criminal patterns among correctional inmates (WBP). This study analyzes criminality clustering based on WBP education level at Rutan Kelas II Ternate using Agglomerative Hierarchical Clustering (AHC), while evaluating cluster quality using the Silhouette Coefficient. The data used is secondary WBP data from 2018-2023 consisting of 1,490 records covering education level, type of criminal offense, age, and case year. The research stages include data preprocessing, transformation, cluster formation using AHC, dendrogram visualization, and evaluation using the Silhouette Coefficient. Results show that AHC formed three main clusters, with a Silhouette Coefficient value of 0.3294, classified as weak structure. This value remains acceptable in exploratory research, as it indicates an initial clustering pattern in the data. The study also produced a cluster trend visualization for 2018-2023 showing the dynamics of cluster distribution over time.

Keywords: *Agglomerative Hierarchical Clustering*, Criminality, *Silhouette Coefficient*, Education Level, Correctional Inmates

1. Pendahuluan

Tindak kriminalitas berpengaruh terhadap keamanan masyarakat serta mengancam ketenteraman lahir dan batin. Kriminalitas merupakan segala bentuk tindakan yang merugikan secara ekonomis maupun psikologis serta melanggar hukum yang berlaku di Indonesia dan norma-norma sosial serta agama [1]. Tingkat kriminalitas yang tinggi dapat mengancam keamanan dan ketertiban masyarakat [2], sehingga penting untuk melakukan analisis dan pengelompokan (*clustering*) terhadap data kriminalitas guna memahami pola dan karakteristik kejahatan yang terjadi.

Salah satu faktor yang dapat memengaruhi tingkat kriminalitas adalah tingkat pendidikan. Tingkat pendidikan berkaitan dengan pembentukan perilaku dan kesadaran masyarakat terhadap hukum dan aturan

[3]. Melalui penelitian ini diharapkan muncul wawasan yang lebih mendalam tentang pola kriminalitas di lingkungan lembaga pemasyarakatan, dengan fokus pada tingkat pendidikan para narapidana.

Analisis pengelompokan dengan *Agglomerative Hierarchical Clustering* (AHC) digunakan sebagai alat utama untuk menelaah apakah ada pola perilaku tertentu yang dapat diidentifikasi berdasarkan tingkat pendidikan WBP. AHC merupakan metode pengelompokan hierarkis *bottom-up*: setiap observasi mulanya dianggap sebagai satu *cluster*, lalu pasangan *cluster* yang paling mirip digabungkan secara bertahap hingga terbentuk struktur hierarki yang dapat divisualisasikan melalui *dendrogram* [4].

Sejumlah penelitian terdahulu menganalisis kriminalitas dengan pendekatan K-Means [5], [6], [7]. Namun, kajian yang secara khusus menganalisis pengelompokan kriminalitas berdasarkan tingkat pendidikan WBP di Rutan Kelas II Ternate dengan pendekatan AHC masih sangat terbatas. Perbedaan mendasar penelitian ini dengan penelitian-penelitian terdahulu terletak pada pendekatan yang digunakan: penelitian sebelumnya umumnya menerapkan metode K-Means yang mengelompokkan data secara partisi tanpa memperlihatkan struktur hierarki kemiripan antar data, serta belum secara khusus mengaitkan hasil pengelompokan dengan tingkat pendidikan WBP. Penelitian ini mengisi celah tersebut dengan menerapkan metode AHC yang menghasilkan struktur hierarkis berupa *dendrogram*, sehingga pola kemiripan antar kelompok data dapat diamati secara lebih rinci sekaligus dikaitkan secara langsung dengan tingkat pendidikan WBP di Rutan Kelas II Ternate. Penelitian ini bertujuan untuk: (1) menganalisis hasil cluster pengelompokan tingkat pendidikan WBP menggunakan metode AHC, dan (2) mengevaluasi kualitas *cluster* menggunakan *Silhouette Coefficient*.

2. Metode Penelitian

2.1 Data Mining

Data Mining merupakan serangkaian proses untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual. *Data Mining* adalah proses penggalian informasi dari *dataset* yang besar melalui penggunaan algoritma dan teknik yang diambil dari bidang statistik, *machine learning*, dan *database* sistem manajemen [8], [9].

2.2 Analisis Clustering

Clustering atau klasterisasi adalah metode pengelompokan data yang merupakan proses partisi satu set objek data ke dalam himpunan bagian yang disebut dengan *cluster*. Objek yang berada di dalam *cluster* memiliki kemiripan karakteristik antar satu sama lain dan berbeda dengan *cluster* yang lain [4], [10].

2.3 Agglomerative Hierarchical Clustering (AHC)

AHC merupakan algoritma yang bertujuan untuk mengelompokkan data ke dalam *cluster* hierarkis. Pendekatan ini dimulai dengan memperlakukan setiap data sebagai sebuah *cluster* tunggal, kemudian secara bertahap menggabungkan dua *cluster* yang memiliki kesamaan tertinggi berdasarkan suatu metrik jarak hingga seluruh data tergabung menjadi satu *cluster* besar. Proses penggabungan tersebut bergantung pada metode *linkage* yang digunakan, di antaranya *single linkage* (jarak minimum antar anggota *cluster*), *complete linkage* (jarak maksimum antar anggota *cluster*), *average linkage* (rata-rata jarak antar anggota), serta *Ward linkage* yang meminimalkan peningkatan variansi di dalam *cluster* setelah penggabungan. Jarak antar data dapat dihitung menggunakan *Euclidean Distance* maupun *Manhattan Distance* [10]. Hasil proses AHC divisualisasikan dalam bentuk *dendrogram*, yaitu diagram pohon yang menggambarkan urutan dan tingkat kemiripan penggabungan antar *cluster*, sehingga jumlah *cluster* optimal dapat ditentukan dengan memotong *dendrogram* pada tingkat jarak tertentu.

Secara algoritmik, proses AHC dapat diringkas ke dalam beberapa tahapan utama, yaitu: (1) menghitung matriks jarak antar seluruh pasangan data menggunakan *Euclidean Distance*; (2) menggabungkan dua *cluster* dengan jarak terdekat berdasarkan kriteria *linkage* yang dipilih; (3) memperbarui matriks jarak setelah penggabungan dilakukan; dan (4) mengulangi proses penggabungan tersebut hingga seluruh data tergabung ke dalam satu *cluster* tunggal. Pada penelitian ini digunakan *Ward Linkage*, yaitu kriteria penggabungan yang meminimalkan peningkatan jumlah kuadrat galat (*sum of squared error*) di dalam *cluster* pada setiap tahap penggabungan, sehingga cenderung menghasilkan *cluster* yang berukuran relatif seimbang dan kompak. Dibandingkan dengan metode partisi seperti K-Means, AHC memiliki keunggulan karena tidak memerlukan penentuan jumlah *cluster* di awal proses serta mampu memperlihatkan hubungan hierarkis antar data melalui *dendrogram*, meskipun kompleksitas komputasinya lebih tinggi sehingga kurang efisien untuk diterapkan pada data berukuran sangat besar.

2.4 Silhouette Coefficient

Silhouette Coefficient (SC) merupakan indikator evaluasi *cluster* yang mengevaluasi *cluster* berdasarkan ukuran rata-rata jarak antara satu titik data dengan titik data lainnya dalam *cluster* yang sama (kohesi) dan jarak rata-rata antar *cluster* yang berbeda (pemisahan). Nilai SC berkisar antara -1 hingga 1; semakin mendekati 1, semakin baik kualitas *cluster* yang terbentuk [11]. Interpretasi rentang nilai *Silhouette Coefficient* dapat dilihat pada Tabel 1.

Tabel 1. Interpretasi Rentang Nilai *Silhouette Coefficient*

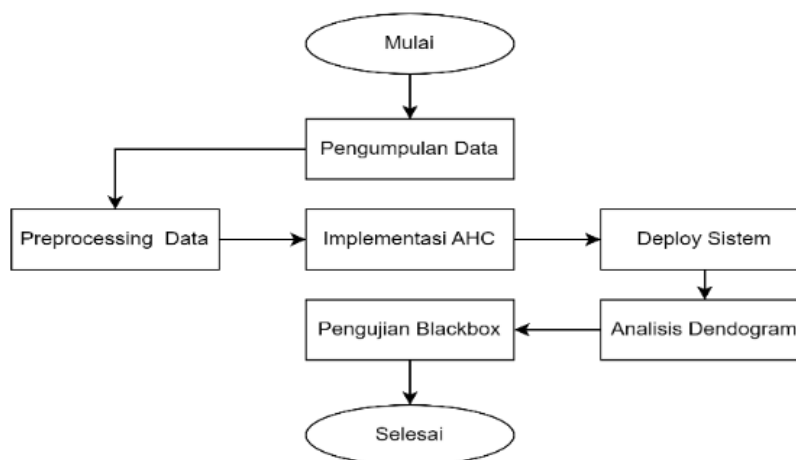
Rentang Nilai	Interpretasi
0,71 - 1,00	<i>Strong Structure</i>
0,51 - 0,70	<i>Reasonable Structure</i>
0,26 - 0,50	<i>Weak Structure</i>
< 0,25	<i>No Substantial Structure</i>

2.5 Objek dan Data Penelitian

Objek penelitian adalah Warga Binaan Pemasyarakatan (WBP) di Rutan Kelas II Ternate. Data yang digunakan merupakan data sekunder WBP periode tahun 2018-2023, sebanyak 1.490 data yang mencakup atribut tingkat pendidikan, jenis tindak pidana, usia, dan tahun perkara. Batasan penelitian mencakup tingkat pendidikan dari SD hingga S3 serta kategori tidak sekolah, dengan jumlah *cluster* yang dievaluasi sebanyak tiga *cluster*.

2.6 Alur Penelitian

Tahapan penelitian meliputi:



Gambar 1. Alur Penelitian

(1) Pengumpulan data dari Rutan Kelas II Ternate, (2) *preprocessing* data yang mencakup pembersihan, normalisasi, dan *encoding* variabel kategorikal, (3) implementasi metode AHC dengan *Euclidean distance* dan *Ward Linkage*, (4) visualisasi *dendrogram*, serta (5) evaluasi kualitas *cluster* menggunakan *Silhouette Coefficient*.

2.7 Preprocessing Data

Tahap *preprocessing* mencakup: normalisasi variasi penulisan tingkat pendidikan (SMA, SMK, STM menjadi SMA/SMK; D1, D2, D3 menjadi Diploma; D4 dan S1 menjadi D4/S1), pengelompokan jenis kejahatan ke dalam tujuh kategori utama (Ekonomi, Kekerasan, Seksual & Anak, Properti, Narkotika, Sosial, dan Lainnya), anonimisasi nama WBP menjadi inisial, serta transformasi data menggunakan *One-Hot Encoding* dan *StandardScaler*.

3. Hasil dan Pembahasan

Bagian ini menyajikan hasil penelitian yang disusun mengikuti tahapan pada alur penelitian (Gambar 1), yaitu analisis data awal, analisis hubungan tingkat pendidikan dan jenis kejahatan, implementasi metode AHC, analisis *dendrogram*, evaluasi *Silhouette Coefficient*, *deploy* sistem beserta pengujian *Black Box*, hingga analisis *trend cluster* tahun 2018-2023.

3.1 Analisis Data

Dataset yang digunakan terdiri atas 1.490 data WBP selama periode 2018-2023. Distribusi kategori kejahatan menunjukkan bahwa sebagian besar WBP terlibat dalam kategori narkoba (604 data), diikuti ekonomi (187 data), lainnya (175 data), properti (169 data), kekerasan (167 data), seksual & anak (137 data), dan sosial (51 data). Distribusi ini mengindikasikan variasi karakteristik tindak kriminal yang memengaruhi proses pembentukan *cluster*.

3.2 Hubungan Tingkat Pendidikan dan Jenis Kejahatan

Berdasarkan Tabel 2, kelompok pendidikan SMA/SMK mendominasi hampir seluruh jenis kejahatan, khususnya pada kejahatan narkoba dengan 452 kasus (49,08%). Kelompok pendidikan Diploma dan D4/S1 cenderung pada kejahatan ekonomi, sedangkan kelompok pendidikan rendah seperti Tidak Sekolah dan SD memiliki distribusi yang lebih merata.

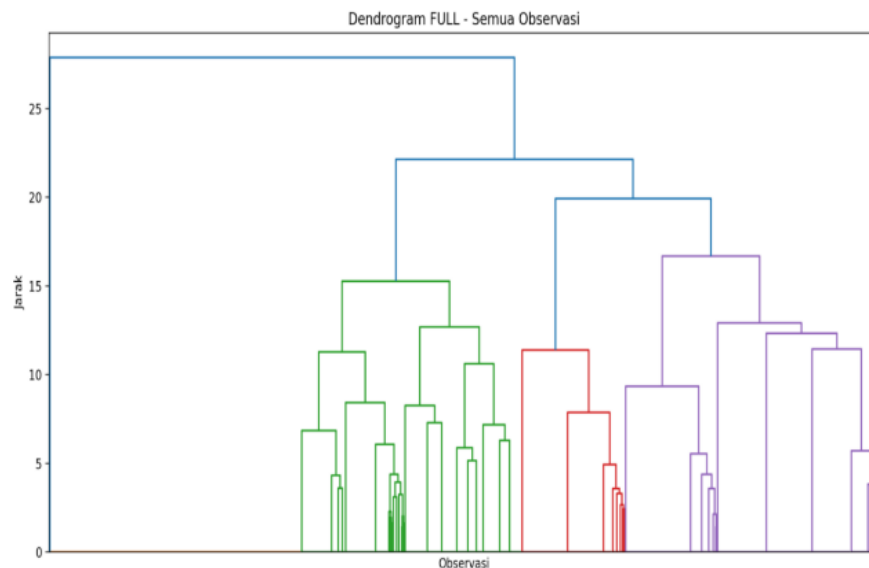
Tabel 2. Hubungan Tingkat Pendidikan dan Jenis Kejahatan (Data Dominan)

Pendidikan	Narkoba	Ekonomi	Properti	Kekerasan	Seksual
SMA/SMK	452	82	88	117	70
D4/S1	54	64	6	6	7
SMP	54	17	40	20	26
SD	24	3	22	12	18
Diploma	11	12	1	0	0

3.3 Implementasi AHC

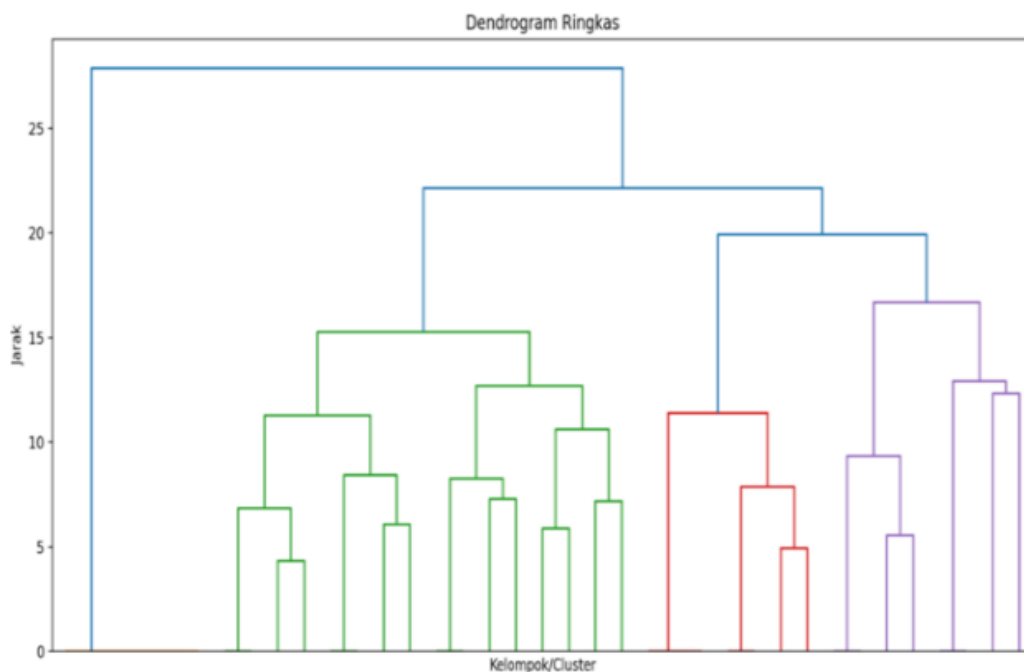
Proses *clustering* dilakukan menggunakan pendekatan *bottom-up* dengan pengukuran jarak *Euclidean Distance* dan metode penggabungan *Ward Linkage*. Pemilihan *Ward Linkage* didasarkan pada kemampuannya meminimalkan variansi di dalam *cluster* sehingga menghasilkan kelompok yang relatif homogen. Jumlah *cluster* ditetapkan sebanyak tiga *cluster* berdasarkan hasil analisis *dendrogram*. Implementasi dilakukan menggunakan *library scikit-learn* pada *Python*.

3.4 Analisis Dendrogram



Gambar 2. Dendrogram

Analisis *dendrogram* penuh menunjukkan bahwa proses penggabungan data terjadi secara bertahap dari bawah ke atas. Pada bagian bawah *dendrogram*, banyak observasi bergabung pada jarak yang rendah, yang menunjukkan kemiripan yang tinggi antar data. Pada bagian atas terlihat adanya lonjakan jarak penggabungan yang signifikan, menandakan perbedaan yang cukup besar antar kelompok data.



Gambar 3. Dendrogram Peringkasan

Dendrogram terpotong (*truncated*) menampilkan pemisahan data menjadi tiga kelompok utama yang ditunjukkan oleh tiga cabang besar. Perbedaan jarak antar cabang utama menunjukkan bahwa masing-masing *cluster* memiliki karakteristik yang berbeda satu sama lain. Hal ini mengkonfirmasi bahwa jumlah *cluster* optimal adalah $k = 3$.

3.5 Evaluasi *Silhouette Coefficient*

Hasil pengujian menunjukkan nilai *Silhouette Coefficient* sebesar 0,3294. Berdasarkan tabel interpretasi, nilai tersebut termasuk dalam kategori *weak structure* (0,26-0,50). Hal ini menunjukkan bahwa pemisahan antar *cluster* belum sepenuhnya optimal dan masih terdapat tumpang tindih antar *cluster*. Kondisi ini dipengaruhi oleh karakteristik data sosial yang heterogen, di mana tingkat pendidikan tidak sepenuhnya menjadi faktor pembeda utama dalam pola kriminalitas.

Meskipun demikian, nilai tersebut masih dapat diterima dalam penelitian eksploratif. Beberapa penelitian menyatakan bahwa pada data sosial nyata (*real-world data*), nilai *Silhouette* di atas 0,3 sudah dapat menunjukkan adanya pola pengelompokan awal (*initial clustering pattern*).

3.6 Deploy Sistem dan Pengujian *Black Box*

Sistem analisis *clustering* dikembangkan berbasis web menggunakan *framework Streamlit* dengan bahasa pemrograman *Python*. Sistem mampu melakukan *preprocessing* data, proses *clustering* AHC, menampilkan visualisasi *dendrogram* (penuh dan terpotong), tabel hasil *clustering* per WBP, serta mengunduh hasil dalam format CSV. Pengujian *Black Box* terhadap lima skenario (*upload dataset*, proses *clustering*, visualisasi *dendrogram*, tampilan tabel hasil, unduh CSV) seluruhnya menunjukkan hasil berhasil.



Gambar 4. Tampilan Unggah *Dataset* dan Proses *Clustering* AHC

Gambar 4 menampilkan antarmuka utama sistem, tempat pengguna mengunggah dataset melalui panel unggah pada sisi kiri dan mengatur parameter jumlah *cluster* (*k*) pada bagian Parameter AHC. Setelah proses *clustering* dijalankan, sistem menampilkan ringkasan hasil berupa jumlah data yang dianalisis, jumlah *cluster* yang terbentuk, metode *linkage* dan metrik jarak yang digunakan, serta nilai *Silhouette Coefficient*.



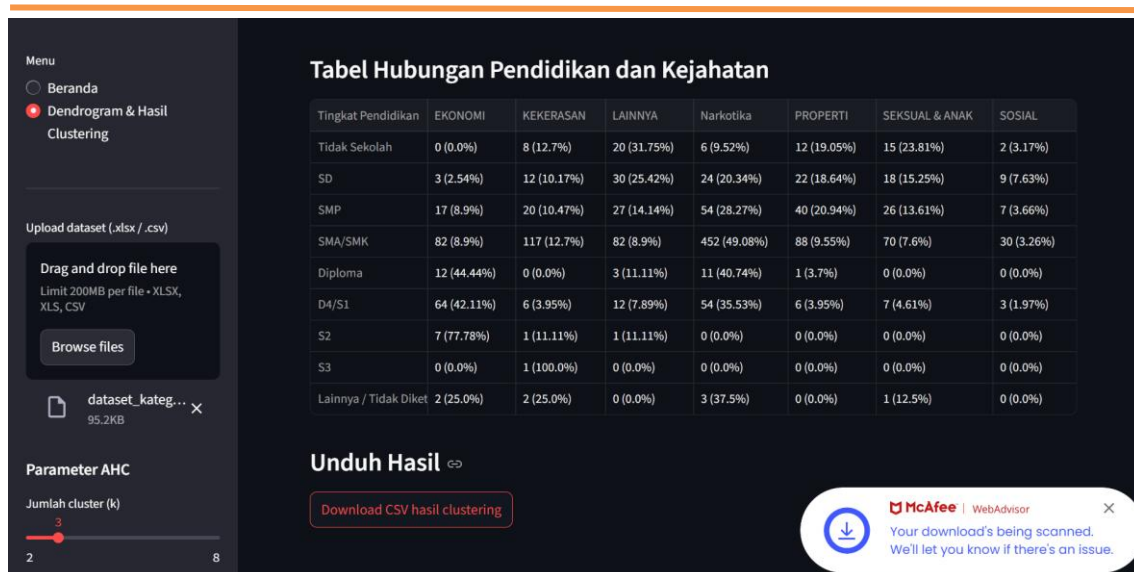
Gambar 5. Visualisasi *Dendrogram*

Gambar 5 menampilkan visualisasi *dendrogram* hasil proses *clustering* AHC, terdiri atas *dendrogram* penuh dan *dendrogram* ringkas (terpotong). Melalui tampilan ini, pengguna dapat mengamati struktur hierarki penggabungan data secara langsung serta melihat ringkasan *cluster* beserta nilai *Silhouette Coefficient* yang dihasilkan.



Gambar 6. Tampilan Tabel Hasil *Clustering* per WBP

Gambar 6 menampilkan tabel hasil *clustering* yang merangkum hubungan antara tingkat pendidikan dan jenis kejahatan pada setiap WBP, lengkap dengan persentase distribusinya pada masing-masing kategori. Pada bagian bawah tabel tersedia tombol unduh yang memungkinkan pengguna menyimpan hasil *clustering* dalam format CSV.



Gambar 7. Tampilan Unduh Hasil

Gambar 7 menunjukkan proses pengunduhan hasil *clustering* setelah tombol Download CSV hasil *clustering* ditekan, di mana sistem menghasilkan berkas CSV berisi data hasil *clustering* yang siap disimpan dan digunakan lebih lanjut oleh pengguna.

Sistem dijalankan pada lingkungan lokal (*localhost*) menggunakan *Streamlit* sebagai antarmuka pengguna, sehingga pengguna dapat mengunggah *dataset*, menjalankan proses *clustering*, melihat visualisasi *dendrogram* serta tren *cluster*, dan mengunduh hasil *clustering* secara langsung melalui *browser*. Hasil pengujian *Black Box* terhadap lima skenario utama pada sistem disajikan pada Tabel 3.

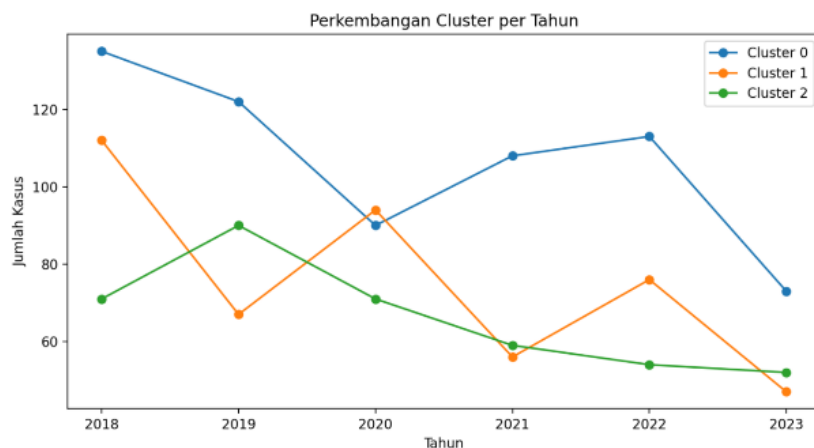
Tabel 3. Hasil Pengujian *Black Box* Sistem

Skenario Pengujian	Hasil yang Diharapkan	Status
Unggah dataset (upload CSV)	Sistem membaca dan menampilkan data WBP sesuai format	Berhasil
Proses <i>clustering</i> AHC	Sistem membentuk tiga <i>cluster</i> sesuai parameter yang ditentukan	Berhasil
Visualisasi <i>dendrogram</i>	Dendrogram penuh dan terpotong tampil sesuai hasil <i>clustering</i>	Berhasil
Tampilan tabel hasil	Tabel hasil <i>clustering</i> per WBP tampil lengkap dan sesuai	Berhasil
Unduh hasil (download CSV)	Berkas CSV hasil <i>clustering</i> berhasil diunduh	Berhasil

Berdasarkan Tabel 3, seluruh skenario pengujian menunjukkan hasil yang sesuai dengan yang diharapkan, sehingga sistem dinyatakan telah berfungsi dengan baik dan siap digunakan untuk mendukung proses analisis *clustering* kriminalitas berdasarkan tingkat pendidikan WBP.

3.7 Tren Cluster Tahun 2018-2023

Grafik Cluster per Tahun



Gambar 8. Tren Cluster

Visualisasi *trend cluster* berdasarkan tahun perkara menunjukkan bahwa setiap *cluster* memiliki pola distribusi yang berbeda pada tiap tahun. Beberapa *cluster* mengalami peningkatan pada periode tertentu, sementara *cluster* lainnya cenderung stabil atau menurun. Dinamika ini mengindikasikan adanya perubahan karakteristik kriminalitas WBP dari waktu ke waktu, sehingga hasil *clustering* tidak hanya menggambarkan kemiripan karakteristik data, tetapi juga perubahan pola kriminalitas secara temporal.

4. Kesimpulan dan Saran

4.1 Kesimpulan

Berdasarkan hasil implementasi dan pembahasan, dapat disimpulkan bahwa: (1) metode *Agglomerative Hierarchical Clustering* (AHC) berhasil diterapkan untuk menganalisis pola kriminalitas berdasarkan tingkat pendidikan WBP di Rutan Kelas II Ternate, menghasilkan tiga cluster utama ($k = 3$) yang ditentukan berdasarkan analisis *dendrogram*; (2) evaluasi kualitas *clustering* menggunakan *Silhouette Coefficient* menghasilkan nilai 0,3294 yang termasuk kategori *weak structure*, menunjukkan bahwa cluster telah terbentuk namun masih terdapat tumpang tindih antar *cluster* akibat heterogenitas data sosial; (3) kelompok pendidikan SMA/SMK mendominasi hampir seluruh jenis kejahatan, terutama narkoba, sedangkan kelompok berpendidikan tinggi (Diploma dan D4/S1) cenderung terlibat dalam kejahatan ekonomi; dan (4) sistem analisis berbasis web menggunakan *Streamlit* berhasil dikembangkan dan seluruh fungsionalitas sistem lulus pengujian *Black Box*. Secara keseluruhan, penelitian ini menunjukkan bahwa metode AHC mampu digunakan sebagai pendekatan eksploratif dalam mengidentifikasi pola kriminalitas berdasarkan tingkat pendidikan WBP sehingga dapat menjadi referensi dalam penyusunan program pembinaan di lembaga pemasyarakatan.

4.2 Saran

Penelitian selanjutnya disarankan menambahkan variabel lain di luar tingkat pendidikan, seperti kondisi ekonomi dan latar belakang sosial, serta membandingkan metode AHC dengan algoritma *clustering* lain untuk memperoleh nilai *Silhouette Coefficient* yang lebih tinggi.

Daftar Pustaka

- [1] L. Suriani, "Pengelompokan Data Kriminal Pada Poldasu Menentukan Pola Daerah Rawan Tindak Kriminal Menggunakan Data Mining Algoritma K-Means Clustering," *Jurnal Sistem Komputer dan Informatika (JSON)*, vol. 1, no. 2, Jan. 2020.

-
- [2] Badan Pusat Statistik, "Statistik Kriminal 2023," Katalog: 4401002, Dec. 2023. [Online]. Available: <https://www.bps.go.id/id/publication/2023/12/12/5edba2b0fe5429a0f232c736/statistik-kriminal-2023.html>.
 - [3] A. D. Putra, G. S. Martha, M. Fikram, and R. J. Yuhan, "Faktor-Faktor yang Memengaruhi Tingkat Kriminalitas di Indonesia Tahun 2018," *Indonesian Journal of Applied Statistics*, vol. 3, no. 2, Nov. 2021.
 - [4] Y. Mayona, R. Buaton, and M. Simanjutak, "Data Mining Clustering Tingkat Kejahatan Dengan Metode Algoritma K-Means (Studi Kasus: Kejaksaan Negeri Binjai)," *Jurnal Informatika Kaputama (JIK)*, vol. 6, no. 3, Aug. 2022.
 - [5] I. Z. Muzakkiy *et al.*, "Penggunaan Algoritma K-Means Pada Metode Clustering Untuk Menganalisa Tindak Kriminal," *Journal of Informatics and Advanced Computing (JIAC)*, vol. 4, no. 1, May 2023.
 - [6] N. Septiani and K. Erwansyah, "Analisis Data Mining Pengelompokkan Kasus Tindak Kejahatan Yang Terjadi Di Kecamatan Medan Polonia Dengan Menggunakan Metode K-Means Clustering," *Jurnal Cyber Tech*, vol. 3, no. 2, Feb. 2020.
 - [7] U. T. Suryadi and Y. Supriatna, "Sistem Clustering Tindak Kejahatan Pencurian Di Wilayah Jawa Barat Menggunakan Algoritma K-Means," *Jurnal Teknologi Dan Komunikasi STMIK Subang*, vol. 12, no. 1, Apr. 2019.
 - [8] Kusrini and E. T. Luthfi, *Algoritma Data Mining*. Yogyakarta: Andi Offset, 2009.
 - [9] J. Han, M. Kamber, and J. Pei, *Data Mining: Concepts and Techniques*, 3rd ed. Waltham, MA: Morgan Kaufmann, 2012.
 - [10] B. S. Everitt, S. Landau, M. Leese, and D. Stahl, *Cluster Analysis*, 5th ed. Chichester: John Wiley & Sons, 2011.
 - [11] H. B. Tambunan, D. H. Barus, J. Hartono, A. S. Alam, D. A. Nugraha, and H. H. H. Usman, "Electrical Peak Load Clustering Analysis Using K-Means Algorithm and Silhouette Coefficient," in *Proc. 2nd Int. Conf. on Technology and Policy in Electric Power and Energy*, Oct. 2020.