

Pemodelan Kepribadian Mahasiswa Santri dengan *Multi-View Clustering* untuk Pembinaan Karakter

Eko Prasetyo Widhi^{*1}, Aziz Musthafa², Triana Harmini³, Miftakhurrokhmat⁴

^{1,2,3} Program Studi Teknik Informatika, Universitas Darussalam Gontor, Ponorogo

⁴ Sekolah Tinggi Manajemen Informatika dan Komputer AMIKOM Surakarta

e-mail: ^{*1}ekoprasetiowidhi@unida.gontor.ac.id, ²aziz@unida.gontor.ac.id,

³triana@unida.gontor.ac.id, ⁴miftah@dosen.amikomsolo.ac.id

Abstrak

Pesantren merupakan lembaga pendidikan karakter tertua di Indonesia, namun proses pembinaan kepribadian mahasiswa santri masih didominasi oleh penilaian subjektif dan belum didukung pendekatan analitik berbasis data. Penelitian ini bertujuan mengembangkan model pemetaan kepribadian mahasiswa santri menggunakan pendekatan *Multi-View Clustering* untuk mendukung pembinaan karakter yang lebih objektif dan personal. Penelitian menerapkan metodologi CRISP-DM pada dataset yang terdiri atas 558 responden dengan 21 atribut, meliputi 6 atribut demografis dan 15 dimensi kepribadian. Untuk memperoleh representasi data yang lebih komprehensif, dilakukan *Exploratory Factor Analysis (EFA)* sehingga diperoleh empat faktor laten yang diintegrasikan dengan atribut asli dalam proses *Multi-View Clustering*. Hasil penelitian menunjukkan bahwa *Gaussian Mixture Model Multi-View (GMM-MV)* dengan dua kluster memberikan performa terbaik berdasarkan *Silhouette Score*, *Calinski–Harabasz Index*, dan *Davies–Bouldin Index*, serta menghasilkan dua profil kepribadian utama, yaitu *Dominant–Analytic* dan *Empathic–Adaptive*, sebagai dasar rekomendasi pembinaan karakter. Evaluasi lebih lanjut menunjukkan bahwa model memenuhi aspek *fairness* dengan tidak terpengaruh oleh lokasi kampus dan memiliki kemampuan generalisasi yang memadai berdasarkan pengujian *Leave-One-Campus-Out (LOCO)*. Model yang dihasilkan selanjutnya diintegrasikan ke dalam *Decision Support System (DSS)* untuk menghasilkan rekomendasi pembinaan karakter yang lebih objektif, adaptif, dan berbasis data.

Kata kunci— Santri; Kepribadian; *Multi-View Clustering*; Data Science; Pembinaan Karakter.

1. PENDAHULUAN

Pesantren berperan penting dalam membentuk karakter dan kepribadian santri melalui integrasi pendidikan agama, disiplin, dan kehidupan asrama [1]. Namun, perkembangan teknologi digital telah mengubah pola perilaku dan interaksi santri, sehingga pendekatan pembinaan yang masih berbasis observasi subjektif menjadi kurang adaptif dan berpotensi menimbulkan bias [2]. Selain itu, keterbatasan instrumen kuantitatif serta pembinaan yang masih bersifat seragam menyebabkan pemetaan kepribadian santri belum optimal [3]. Dalam konteks data science, teknik clustering telah digunakan untuk segmentasi kepribadian [4], umumnya melalui pendekatan *single-view clustering* seperti K-Means [5]. Namun, pendekatan ini terbatas karena hanya memanfaatkan satu representasi data sehingga cenderung mengabaikan kompleksitas hubungan antar dimensi kepribadian. Untuk mengatasi hal tersebut, *multi-view clustering* (MVC) dikembangkan dengan mengintegrasikan berbagai sumber data sehingga mampu menghasilkan kluster yang lebih representatif dan robust [6]. Berbagai penelitian terdahulu menunjukkan perkembangan metode MVC, seperti *low-rank embedded multi-view clustering* [6], *one-step multi-view clustering with diverse representation* [7], pendekatan berbasis graf (*graph fusion*) [8], *fuzzy self-weighted multi-view clustering* [9], serta pendekatan *late integration* seperti CPS-merge [10], yang terbukti meningkatkan kualitas klusterisasi pada data kompleks. Secara komparatif,

single-view clustering cenderung menyederhanakan data, sedangkan MVC mampu mengintegrasikan informasi komplementer sehingga menghasilkan kluster yang lebih akurat dan stabil. Meskipun demikian, penelitian yang ada masih terbatas pada domain umum dan belum secara spesifik diterapkan pada mahasiswa santri maupun diintegrasikan ke dalam pembinaan karakter [11].

Berdasarkan hal tersebut, terdapat tiga celah penelitian utama, yaitu: (1) belum adanya model pemetaan kepribadian santri berbasis MVC yang mengintegrasikan data eksplisit dan laten; (2) belum adanya evaluasi *fairness* untuk memastikan model bebas bias; dan (3) belum adanya pengujian generalisasi pada konteks lintas pesantren. Oleh karena itu, penelitian ini mengembangkan model pemetaan kepribadian berbasis *multi-view clustering* yang diintegrasikan ke dalam sistem pendukung keputusan guna mendukung pembinaan karakter yang lebih objektif dan berbasis data.

2. METODE PENELITIAN

Penelitian ini menggunakan metodologi CRISP-DM (Cross Industry Standard Process for Data Mining) yang terdiri atas enam tahapan, yaitu *Business Understanding*, *Data Understanding*, *Data Preparation*, *Modeling*, *Evaluation*, dan *Deployment* [12]. Metodologi ini dipilih karena menyediakan kerangka kerja yang sistematis dan iteratif dalam pengembangan model analitik berbasis data.



Gambar 1. Metodologi CRISP-DM

Pada tahap *Business Understanding*, penelitian mengidentifikasi permasalahan berupa subjektivitas dalam penilaian kepribadian santri sehingga diperlukan pendekatan berbasis data untuk mendukung pengambilan keputusan yang lebih objektif [13].

Tahap *Data Understanding* menggunakan dataset sebanyak 558 responden dengan 21 atribut, yang terdiri atas 6 variabel demografis dan 15 dimensi kepribadian berbasis *self-assessment*. Struktur dataset disajikan pada Tabel 1, sedangkan definisi setiap dimensi kepribadian dirangkum pada Tabel 2.

Tabel 1. Struktur dan Contoh Dataset

ID	Jenis Kelamin	Kampus	Achievement	Order	Autonomy	Affiliation	Endurance
1	L	A	4.2	3.8	4.5	3.9	4.1
2	P	B	3.7	4.1	3.6	4.3	3.8
3	L	A	4.5	4.0	4.2	3.7	4.4

Tabel 1 menunjukkan struktur dataset yang menjadi masukan utama pada proses prapemrosesan, analisis faktor, dan pemodelan clustering. Tabel 2 menjelaskan 15 dimensi kepribadian yang digunakan sebagai variabel penelitian dan menjadi dasar dalam analisis faktor serta pembentukan kluster kepribadian.

Tabel 2. Dimensi Kepribadian dalam dataset

No	Dimensi Kepribadian	Deskripsi Singkat	No	Dimensi Kepribadian	Deskripsi Singkat
1	Achievement	Orientasi terhadap prestasi dan pencapaian	9	Dominance	Kecenderungan memimpin dan mengendalikan
2	Deference	Kecenderungan mengikuti otoritas dan aturan	10	Abasement	Kerendahan hati dan penerimaan diri
3	Order	Kedisiplinan dan kerapian dalam bekerja	11	Nurturance	Kepedulian dan keinginan membantu sesama
4	Exhibition	Keinginan menampilkan diri dan berbagi ide	12	Change	Fleksibilitas terhadap perubahan
5	Autonomy	Kemandirian dalam berpikir dan bertindak	13	Endurance	Ketekunan dan daya tahan menghadapi tantangan
6	Affiliation	Kebutuhan untuk berinteraksi dan diterima	14	Aggression	Kecenderungan menghadapi konflik secara langsung
7	Intrapersonal	Kemampuan memahami diri sendiri	15	Heterosexual	Penyesuaian sosial dengan lawan jenis
8	Succorance	Ketergantungan terhadap dukungan orang lain			

Tahap *Data Preparation* meliputi imputasi median untuk menangani nilai hilang, normalisasi menggunakan *StandardScaler* [14], serta reduksi dimensi menggunakan Exploratory Factor Analysis (EFA). Kelayakan data diuji dengan KMO dan Bartlett's Test [15]. Pada tahap *Modeling*, digunakan algoritma clustering seperti K-Means, Gaussian Mixture Model (GMM, dan Hierarchical Clustering. Pendekatan Multi-View Clustering diterapkan dengan mengintegrasikan fitur asli dan skor faktor hasil EFA untuk menangkap informasi eksplisit dan laten secara simultan [16]. Tahap *Evaluation* menggunakan metrik *Silhouette Score* [17], Calinski–Harabasz Index [18], dan Davies–Bouldin Index untuk mengukur kualitas kluster [19]. Selain itu, dilakukan evaluasi *fairness* untuk memastikan model bebas bias [20], serta uji generalisasi menggunakan metode *Leave-One-Campus-Out* (LOCO) [21]. Tahap *Deployment* dilakukan dengan mengintegrasikan model ke dalam Decision Support System (DSS) guna menghasilkan rekomendasi pembinaan karakter berbasis data.

3. HASIL DAN PEMBAHASAN

Hasil penelitian menunjukkan bahwa proses pembinaan karakter mahasiswa santri masih menghadapi beberapa kendala, terutama subjektivitas penilaian, belum tersedianya klasifikasi kepribadian yang terstandar, serta belum adanya sistem pendukung keputusan berbasis data. Ringkasan permasalahan dan kebutuhan sistem disajikan pada Tabel 3.

Table 3. Ringkasan Permasalahan dan Kebutuhan Sistem Pembinaan Karakter

Aspek	Permasalahan	Kebutuhan
Penilaian karakter	Subjektif, tidak terstandar	Sistem pemetaan kepribadian berbasis data
Analisis santri	Tidak ada klasifikasi kepribadian	Model kluster santri objektif dan stabil
Pembinaan	Tidak personal	Rekomendasi pembinaan per tipe
Pengambilan keputusan	Tidak terdokumentasi	DSS pembinaan karakter berbasis bukti

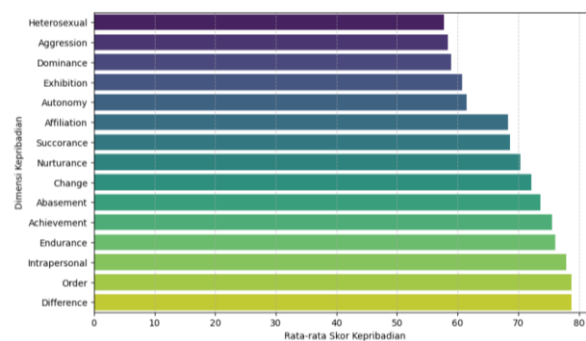
Tabel 3 menunjukkan bahwa setiap tahapan pembinaan karakter masih dilakukan secara konvensional sehingga diperlukan sistem berbasis data untuk menghasilkan pemetaan kepribadian dan rekomendasi pembinaan yang lebih objektif. Dataset penelitian terdiri atas 558 responden dengan 21 atribut, yang meliputi 6 atribut demografis dan 15 dimensi kepribadian, sebagaimana ditunjukkan pada Tabel 4.

Tabel 4 menunjukkan bahwa atribut kepribadian menjadi variabel utama dalam proses clustering, sedangkan atribut demografis dimanfaatkan sebagai informasi pendukung dan evaluasi fairness.

Table 4. Struktur Atribut Dataset Penelitian

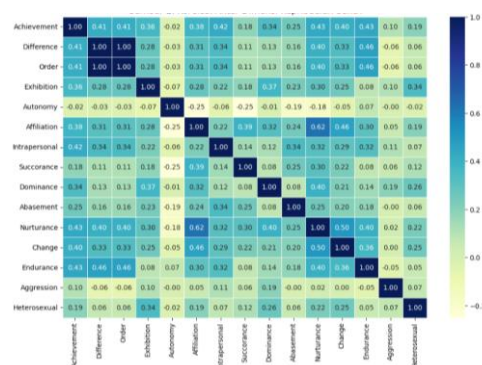
Kategori	Atribut	Deskripsi
Demografis (6)	Nama, Usia, Jenis Kelamin, Program Studi, Lokasi Kampus, Semester	Informasi profil responden
Kepribadian (15)	Achievement, Difference, Order, Exhibition, Autonomy, Affiliation, Intrapersonal, Succorance, Dominance, Abasement, Nurturance, Change, Endurance, Aggression, Heterosexual	Variabel utama analisis

Analisis deskriptif menunjukkan bahwa sebagian besar dimensi memiliki rata-rata di atas 65 dengan penyebaran data yang relatif rendah, sehingga karakteristik responden dapat dianggap cukup homogen. Distribusi nilai rata-rata setiap dimensi ditampilkan pada Gambar 2, sedangkan hubungan antar dimensi divisualisasikan pada Gambar 3.



Gambar 2. Distribusi Nilai Rata-rata Dimensi Kepribadian Santri

Gambar 2 menunjukkan bahwa skor rata-rata dimensi kepribadian berada pada kisaran 57–78. Dimensi Difference, Order, dan Intrapersonal memiliki nilai rata-rata tertinggi, sedangkan Heterosexual, Aggression, dan Dominance memiliki nilai terendah. Temuan ini menunjukkan bahwa sebagian besar mahasiswa santri memiliki karakter yang teratur, adaptif, dan berorientasi pada pengembangan diri, sehingga layak digunakan sebagai dasar pemetaan kepribadian.



Gambar 3. Korelasi Antar Dimensi Kepribadian Santri

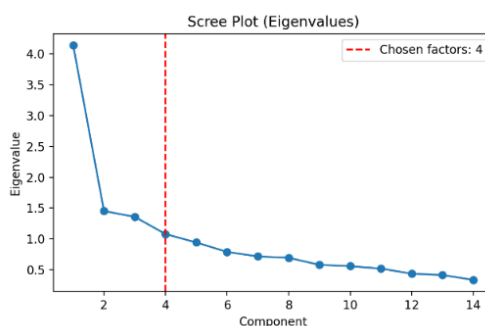
Gambar 3 memperlihatkan bahwa sebagian besar dimensi memiliki korelasi positif dengan kekuatan rendah hingga sedang. Korelasi tertinggi terjadi antara Nurturance dan Affiliation ($r = 0,62$), sedangkan korelasi negatif terendah ditemukan antara Aggression dan Endurance ($r = -0,05$). Hasil ini menunjukkan bahwa setiap dimensi saling berkaitan, namun tetap mampu merepresentasikan konstruk psikologis yang berbeda sehingga memenuhi asumsi awal untuk dilakukan analisis faktor. Untuk memverifikasi kelayakan analisis faktor, dilakukan

pengujian Kaiser-Meyer-Olkin (KMO) dan Bartlett's Test of Sphericity, sebagaimana ditunjukkan pada Tabel 5.

Table 5. Hasil Uji Kelayakan EFA

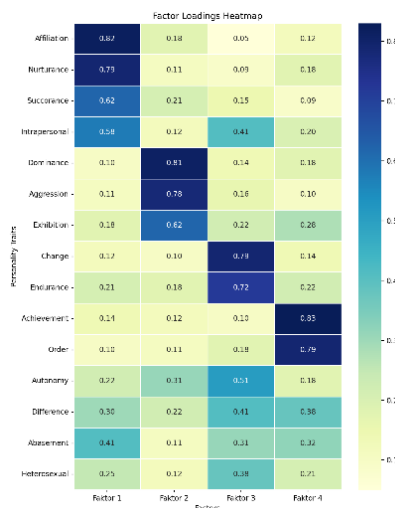
Uji	Hasil	Interpretasi
KMO	0.799	Layak (merit)
Bartlett's Test	$p < 0.001$	Variabel saling berkorelasi signifikan

Tabel 5 menunjukkan bahwa nilai KMO sebesar 0,799 dan Bartlett's Test signifikan, sehingga data memenuhi syarat untuk dilakukan Exploratory Factor Analysis (EFA). Analisis EFA berhasil mengidentifikasi empat faktor utama yang menjelaskan 57,2% total variansi data. Struktur faktor divisualisasikan pada Gambar 4 dan Gambar 5, sedangkan hasil interpretasi faktor dirangkum pada Tabel 6.



Gambar 4. Scree Plot – titik infleksi pada empat faktor utama

Gambar 4 menunjukkan titik infleksi (*elbow*) pada faktor keempat. Hal ini mengindikasikan bahwa empat faktor merupakan jumlah faktor yang paling optimal untuk dipertahankan karena faktor berikutnya hanya memberikan tambahan variansi yang relatif kecil.



Gambar 5. Heatmap Factor Loadings (Varimax Rotation)

Gambar 5 memperlihatkan bahwa setiap dimensi memiliki nilai *factor loading* dominan pada salah satu faktor dengan nilai *loading* umumnya di atas 0,50, sehingga struktur faktor yang dihasilkan mudah diinterpretasikan.

Tabel 6 menunjukkan bahwa kelima belas dimensi kepribadian berhasil diringkaskan menjadi empat faktor psikologis utama.

Table 6. Faktor Psikologis Utama

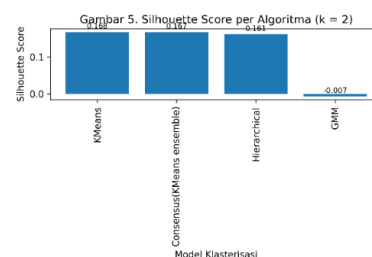
Faktor	Dimensi Dominan	Makna Psikologis
F1	Affiliation, Nurturance	Empati & sosial
F2	Dominance, Aggression	Ketegasan & kontrol diri
F3	Change, Endurance	Adaptabilitas & ketekunan
F4	Achievement, Order	Prestasi & disiplin

Reduksi dimensi ini menghasilkan representasi data yang lebih sederhana tanpa menghilangkan informasi penting, sehingga digunakan sebagai salah satu *view* pada proses clustering. Selanjutnya dilakukan pemodelan menggunakan pendekatan Multi-View Clustering dengan mengintegrasikan atribut asli dan skor faktor hasil EFA. Perbandingan performa algoritma ditampilkan pada Tabel 7.

Table 7. Evaluasi Kinerja Model Clustering

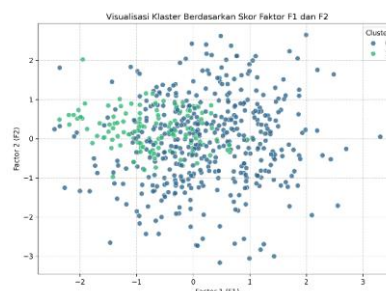
Model	k	Silhouette	CH Index	DB Index
K-Means	2	0.141	9.21	0.82
Hierarchical	2	0.134	7.89	0.91
Consensus	2	0.088	6.42	1.04
GMM-MV	2	0.192	13.60	0.74

Tabel 7 menunjukkan bahwa GMM-MV memberikan performa terbaik berdasarkan Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index. Hasil ini menunjukkan bahwa integrasi fitur asli dan faktor laten mampu menghasilkan pemisahan kluster yang lebih baik dibandingkan pendekatan clustering konvensional. Perbandingan performa ditunjukkan pada Gambar 6, sedangkan visualisasi kluster dan profil dimensi ditunjukkan pada Gambar 7 dan Gambar 8. Gambar 6 memperlihatkan bahwa algoritma GMM-MV memperoleh nilai *Silhouette Score* tertinggi sehingga menghasilkan kualitas kluster yang paling baik.



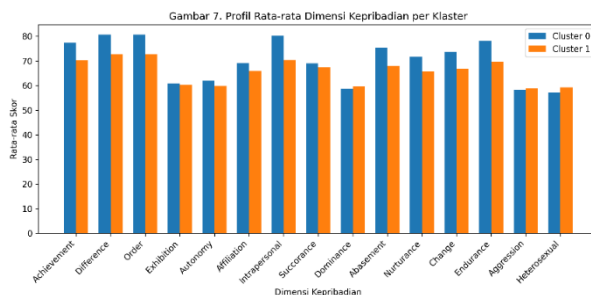
Gambar 6. Silhouette Score per Algoritma (Bar Chart)

Gambar 7 menunjukkan bahwa dua kluster utama dapat dipisahkan secara relatif jelas pada ruang faktor psikologis, meskipun masih terdapat beberapa titik yang berada pada area transisi.



Gambar 7. Scatter Plot Kluster (Factor Score F1–F2)

Gambar 8 menunjukkan bahwa kedua kluster memiliki karakteristik kepribadian yang berbeda sehingga dapat digunakan sebagai dasar penyusunan strategi pembinaan yang lebih personal. Untuk memastikan bahwa model tidak dipengaruhi oleh lokasi kampus, dilakukan evaluasi fairness sebagaimana ditunjukkan pada Tabel 8.



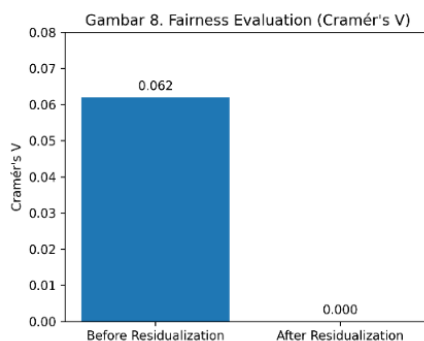
Gambar 8. Profil rata-rata dimensi kepribadian pada masing-masing kluster

Tabel 8 menunjukkan bahwa nilai *Cramér's V* menurun menjadi 0,000 setelah proses residualisasi, sehingga model dapat dianggap bebas dari bias berdasarkan lokasi kampus.

Table 8. Evaluasi Fairness

Tahap	Cramér's V	Interpretasi
Sebelum	0.062	Terdapat pengaruh kecil dari lokasi kampus
Sesudah	0.000	Model sepenuhnya bebas dari bias kampus

Gambar 9 mengkonfirmasi bahwa distribusi kluster menjadi lebih seimbang setelah proses residualisasi sehingga pengaruh asal kampus terhadap hasil clustering berhasil diminimalkan. Kemampuan generalisasi model kemudian diuji menggunakan metode *Leave-One-Campus-Out* (LOCO) sebagaimana ditunjukkan pada Tabel 9.

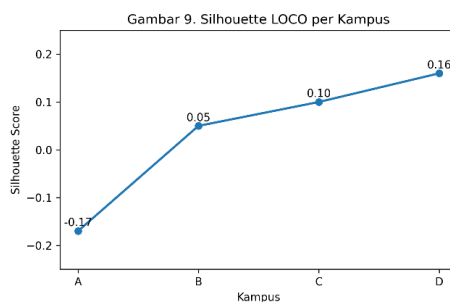


Gambar 9. Visualisasi Fairness

Tabel 9 menunjukkan bahwa model memiliki kemampuan generalisasi yang cukup baik meskipun terdapat variasi performa antar kampus. Hal ini menunjukkan bahwa karakteristik lokal masih memengaruhi hasil clustering.

Table 9. Hasil Leave-One-Campus-Out

Kampus	Silhouette	Interpretasi
A	-0.17	Rendah
B	0.05	Moderat
C	0.10	Moderat-baik
D	0.16	Stabil



Gambar 10. Generalisasi Model

Gambar 10 memperlihatkan konsistensi performa model pada setiap skenario LOCO, yang menunjukkan bahwa model tetap mampu mempertahankan kualitas clustering pada lingkungan yang berbeda. Tahap akhir penelitian mengimplementasikan model ke dalam *Decision Support System* (DSS) untuk menghasilkan rekomendasi pembinaan karakter berdasarkan hasil clustering.

Table 3 Profil Tipe Kepribadian dan Rekomendasi Pembinaan

Tipe Kepribadian	Karakteristik Utama	Rekomendasi Pembinaan
Dominant–Analytic	Achievement, Order, Dominance (tegas, terstruktur, berorientasi hasil)	Pelatihan kepemimpinan, tugas struktural, manajemen organisasi
Empathic–Adaptive	Affiliation, Nurturance, Intrapersonal (empatik, adaptif, komunikatif)	Kegiatan sosial, mentoring, program kolaboratif

Tabel 10 menunjukkan bahwa hasil clustering menghasilkan dua tipe kepribadian utama dengan karakteristik dan rekomendasi pembinaan yang berbeda. Implementasi ke dalam DSS memungkinkan proses pembinaan karakter dilakukan secara lebih personal, objektif, dan berbasis data. Secara keseluruhan, pendekatan *Multi-View Clustering* berhasil meningkatkan kualitas pemetaan kepribadian melalui integrasi fitur observasional dan faktor laten hasil EFA. Model yang dihasilkan memiliki kualitas klaster yang lebih baik, bebas dari bias lokasi kampus, serta memiliki kemampuan generalisasi yang memadai sehingga berpotensi menjadi dasar pengembangan sistem pendukung keputusan dalam pembinaan karakter mahasiswa santri.

4. KESIMPULAN

Penelitian ini berhasil mengembangkan model pemetaan kepribadian mahasiswa santri berbasis *Multi-View Clustering* yang mengintegrasikan atribut kepribadian dan faktor laten hasil *Exploratory Factor Analysis* (EFA). Pendekatan yang diusulkan mampu meningkatkan kualitas clustering dibandingkan metode konvensional, menghasilkan dua profil kepribadian utama sebagai dasar rekomendasi pembinaan, serta memenuhi aspek fairness dan generalisasi melalui pengujian residualisasi dan *Leave-One-Campus-Out* (LOCO). Hasil penelitian ini mampu menyediakan pendekatan analitik berbasis data untuk pembinaan karakter mahasiswa santri secara lebih objektif, adaptif, dan terukur.

5. SARAN

Penelitian selanjutnya diharapkan memperluas cakupan dataset pada berbagai institusi pendidikan berbasis pesantren. Pengembangan metode juga dapat dilakukan melalui pemanfaatan teknik *deep learning*, *graph neural network*, atau *advanced multi-view clustering* untuk meningkatkan kualitas representasi data dan akurasi clustering. Selain itu, implementasi model ke dalam *Decision Support System* (DSS) perlu divalidasi melalui studi lapangan guna mengevaluasi dampaknya terhadap efektivitas pembinaan karakter dan pengambilan keputusan di lingkungan pesantren.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Fakultas Sains dan Teknologi Universitas Darussalam Gontor atas dukungan hibah penelitian. Program Studi Teknik Informatika UNIDA Gontor atas fasilitasi, serta STMIK AMIKOM Surakarta atas kolaborasi dalam penelitian ini.

DAFTAR PUSTAKA

- [1] H. Fahmy Zarkasyi, F. Mas'ud, R. Agung Hidayatullah, and U. Khakim, "Strategy of Indonesian Pesantren University in Achieving Competence of Student: A Grounded Research at UNIDA Gontor," *KnE Soc. Sci.*, vol. 2024, pp. 119–138, 2024, doi: 10.18502/kss.v9i6.15260.
- [2] S. Yusuf, M. Marhumah, A. Muslim, and N. Zainuddin, "Character Education for Contemporary Urban Muslim Students: a Comparative Study At Uii and Umy Student Islamic Boarding Schools," *Akad. J. Pemikir. Islam*, vol. 30, no. 1, pp. 135–146, Jun. 2025, doi: 10.32332/akademika.v30i1.9596.
- [3] C. Gemilang, A. Siahaan, and F. Rohman, "Boarding School Management Strategies in Developing Student Character at State Islamic Senior High School (MAN) 2 Tanjung Pura," *J. Gen. Educ. Humanit.*, vol. 4, no. 3, pp. 1205–1216, Aug. 2025, doi: 10.58421/gehu.v4i3.654.
- [4] A. Kerber, M. Roth, and P. Y. Herzberg, "Personality types revisited—a literature-informed and data-driven approach to an integration of prototypical and dimensional constructs of personality description," *PLoS One*, vol. 16, no. 1 January, p. e0244849, Jan. 2021, doi: 10.1371/journal.pone.0244849.
- [5] Patrick Lay and Ary Budi Warsito, "Penerapan Algoritma K-Means Untuk Clustering Big Five Personality," *J. Tek. Mesin, Ind. Elektro dan Inform.*, vol. 3, no. 1, pp. 240–246, 2023, doi: 10.55606/jtmei.v3i1.3288.
- [6] S. Wang, L. Chen, Z. Liang, and Q. Liu, "A Novel Low-Rank Embedded Latent Multi-View Subspace Clustering Approach," *Sensors*, vol. 25, no. 9, p. 2778, Apr. 2025, doi: 10.3390/s25092778.
- [7] X. Wan *et al.*, "One-Step Multi-View Clustering with Diverse Representation," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 36, no. 3, pp. 5774–5786, Mar. 2025, doi: 10.1109/TNNLS.2024.3378194.
- [8] X. Li, H. Zhang, R. Wang, and F. Nie, "Multiview Clustering: A Scalable and Parameter-Free Bipartite Graph Fusion Method," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 44, no. 1, pp. 330–344, Jan. 2022, doi: 10.1109/TPAMI.2020.3011148.
- [9] X. Zhu, S. Zhang, Y. Zhu, W. Zheng, and Y. Yang, "Self-weighted Multi-view Fuzzy Clustering," *ACM Trans. Knowl. Discov. Data*, vol. 14, no. 4, pp. 1–17, Aug. 2020, doi: 10.1145/3396238.
- [10] L. Zhang, L. Lin, and J. Li, "Multi-view clustering by CPS-merge analysis with application to multimodal single-cell data," *PLOS Comput. Biol.*, vol. 19, no. 4, p. e1011044, Apr. 2023, doi: 10.1371/journal.pcbi.1011044.

- [11] Y. Zhang *et al.*, “Learning Uniform Latent Representation via Alternating Adversarial Network for Multi-View Clustering,” *IEEE Trans. Emerg. Top. Comput. Intell.*, vol. 9, no. 3, pp. 2244–2255, Jun. 2025, doi: 10.1109/TETCI.2025.3540426.
- [12] J. Brzozowska, J. Pizoń, G. Baytikenova, A. Gola, A. Zakimova, and K. Piotrowska, “Data Engineering in Crisp-Dm Process Production Data – Case Study,” *Appl. Comput. Sci.*, vol. 19, no. 3, pp. 83–95, Sep. 2023, doi: 10.35784/acs-2023-26.
- [13] S. Kaspi and S. Venkatraman, “Data-Driven Decision-Making (DDDM) for Higher Education Assessments: A Case Study,” *Systems*, vol. 11, no. 6, p. 306, Jun. 2023, doi: 10.3390/systems11060306.
- [14] K. Mahmud Sujon, R. Binti Hassan, Z. Tusnia Towshi, M. A. Othman, M. Abdus Samad, and K. Choi, “When to Use Standardization and Normalization: Empirical Evidence from Machine Learning Models and XAI,” *IEEE Access*, vol. 12, pp. 135300–135314, 2024, doi: 10.1109/ACCESS.2024.3462434.
- [15] N. Shrestha, “Factor Analysis as a Tool for Survey Analysis,” *Am. J. Appl. Math. Stat.*, vol. 9, no. 1, pp. 4–11, Jan. 2021, doi: 10.12691/ajams-9-1-2.
- [16] X. Wan *et al.*, “One-Step Multi-View Clustering With Diverse Representation,” *IEEE Trans. Neural Networks Learn. Syst.*, vol. 36, no. 3, pp. 5774–5786, Mar. 2025, doi: 10.1109/TNNLS.2024.3378194.
- [17] H. Y. Chuang and W. Chou, “SilhouetteScoreinR: Beyond traditional network layouts by leveraging local cohesion and nearest neighbor separation,” *MethodsX*, vol. 15, p. 103622, Dec. 2025, doi: 10.1016/j.mex.2025.103622.
- [18] L. W. Yerbury, R. J. G. B. Campello, G. C. Livingston, M. Goldsworthy, and L. O’Neil, “On the Use of Relative Validity Indices for Comparing Clustering Approaches,” *ACM Trans. Knowl. Discov. Data*, vol. 19, no. 8, pp. 1–53, Sep. 2025, doi: 10.1145/3748726.
- [19] D. Chicco, A. Campagner, A. Spagnolo, D. Ciucci, and G. Jurman, “The Silhouette coefficient and the Davies-Bouldin index are more informative than Dunn index, Calinski-Harabasz index, Shannon entropy, and Gap statistic for unsupervised clustering internal evaluation of two convex clusters,” *PeerJ Comput. Sci.*, vol. 11, p. e3309, Nov. 2025, doi: 10.7717/peerj-cs.3309.
- [20] J. P. Lalor, A. Abbasi, K. Oketch, Y. Yang, and N. Forsgren, “Should Fairness be a Metric or a Model? A Model-based Framework for Assessing Bias in Machine Learning Pipelines,” *ACM Trans. Inf. Syst.*, vol. 42, no. 4, pp. 1–41, Jul. 2024, doi: 10.1145/3641276.
- [21] S. C. Matz, C. S. Bukow, H. Peters, C. Deacons, and C. Stachl, “Using machine learning to predict student retention from socio-demographic characteristics and app-based engagement metrics,” *Sci. Rep.*, vol. 13, no. 1, p. 5705, Apr. 2023, doi: 10.1038/s41598-023-32484-w.