



Clustering dan Visualisasi Data ASN dalam Penunjang Analisis Kecukupan Data di Perangkat Daerah Pemerintah Provinsi Jawa Timur

Zhulfi Bajra Wikjatmiko^{1,2}, Endang Setyati^{2,*}, Yosi Kristian^{2,3}

¹UPT Pusat Penilaian Pegawai - Badan Kepegawaian Daerah Provinsi Jawa Timur,
Jalan Jemur Andayani No 1, Wonocolo, Surabaya, Indonesia
wikjatmiko@gmail.com

²Teknologi Informasi, Institut Sains dan Teknologi Terpadu Surabaya (ISTTS) Surabaya,
Jalan Ngagel Jaya Tengah 73-77, Surabaya, Indonesia
endang@stts.edu

³Informatika, Institut Sains dan Teknologi Terpadu Surabaya (ISTTS) Surabaya,
Jalan Ngagel Jaya Tengah 73-77, Surabaya, Indonesia
yosi@stts.edu

*Corresponding Author

Abstrak

Penelitian ini mengimplementasikan K-Means, Gaussian Mixture Model (GMM), dan Hierarchical Clustering untuk menganalisis kecukupan data 18.962 ASN Pemerintah Provinsi Jawa Timur dari 232 unit kerja di 38 wilayah administratif. Dataset terdiri dari 8 jenis perangkat daerah dengan Satpol PP (7.231 ASN) dan UPT (6.961 ASN) sebagai kontributor terbesar. Variabel clustering mencakup 12 atribut kelengkapan dokumen kepegawaian dalam format biner: foto ½ badan, foto full body, akta lahir, KTP, NPWP, sumpah jabatan PNS, nota BKN, SPMT, kartu ASN virtual, nomor NPWP, nomor BPJS, dan nomor KK. Metodologi penelitian meliputi preprocessing data dengan normalisasi Min-Max, penghapusan 287 duplikat, dan encoding biner. Metode Elbow menghasilkan cluster optimal k=7 untuk K-Means (distortion score 119.496), k=4 untuk GMM (BIC 119.839), dan k=3 untuk Hierarchical Clustering. Evaluasi menggunakan Silhouette Score, Calinski-Harabasz Index, dan Davies-Bouldin Index menunjukkan K-Means memiliki performa terbaik (0.332, 3412.783, 1.224). K-Means mengidentifikasi 35% ASN kategori High (>80%), 45% Medium (70-79%), dan 20% Low (<70%). GMM menghasilkan distribusi 40% High, 55% Medium, 5% Low plus 14 outlier. Hierarchical Clustering menghasilkan 52% High, 47% Medium, 1% Low. Temuan menunjukkan unit kerja Surabaya memiliki kelengkapan tertinggi (54.27%) dibanding kabupaten lain (<5%). PNS memiliki kelengkapan 90% lebih baik dari PPPK/CPNS. Kartu ASN Virtual dan Nomor KK merupakan dokumen dengan kelengkapan terendah (<40%). Visualisasi melalui dashboard interaktif, heatmap, scatter plot PCA, dan dendrogram memfasilitasi identifikasi prioritas pembenahan data. Model ini dapat diadaptasi untuk mendukung transformasi digital birokrasi di instansi pemerintah lainnya.

Kata Kunci: Clustering dan Visualisasi Data, ASN Pemerintah Provinsi Jawa Timur, Pemutakhiran data, Sistem informasi kepegawaian, Metode Elbow

Abstract

This research implements K-Means, Gaussian Mixture Model (GMM), and Hierarchical Clustering to analyze data completeness of 18,962 Civil Service Employees (ASN) from East Java Provincial Government across 232 work units in 38 administrative regions. The dataset comprises 8 types of regional agencies with Civil Service Police Unit (7,231 ASN) and Technical Implementation Units (6,961 ASN) as the largest contributors. Clustering variables include 12 personnel document completeness attributes in binary format: half-body photo, full-body photo, birth certificate, ID card, tax ID number, civil servant oath, civil service agency note, duty statement letter, virtual ASN card, tax ID registration number, social insurance number, and family card number. The research methodology includes data preprocessing with Min-Max normalization, removal of 287 duplicates, and binary encoding. The Elbow method yields optimal clusters of $k=7$ for K-Means (distortion score 119,496), $k=4$ for GMM (BIC 119,839), and $k=3$ for Hierarchical Clustering. Evaluation using Silhouette Score, Calinski-Harabasz Index, and Davies-Bouldin Index shows K-Means performs best (0.332, 3412.783, 1.224). K-Means identifies 35% ASN in High category ($>80\%$), 45% Medium (70-79%), and 20% Low ($<70\%$). GMM produces distribution of 40% High, 55% Medium, 5% Low plus 14 outliers. Hierarchical Clustering yields 52% High, 47% Medium, 1% Low. Findings reveal Surabaya work units have the highest completeness (54.27%) compared to other districts ($<5\%$). Permanent civil servants show 90% better completeness than government contract employees. Virtual ASN Card and Family Card Number are documents with lowest completeness ($<40\%$). Visualization through interactive dashboard, heatmap, PCA scatter plot, and dendrogram facilitates identification of data improvement priorities. This model can be adapted to support digital transformation of bureaucracy in other government institutions.

Keywords: Clustering and Data Visualization, East Java Provincial Government ASN, Data updating, Personnel information system, Elbow Method

1. Pendahuluan

Transformasi digital dalam tata kelola pemerintahan telah membawa perubahan fundamental dalam cara organisasi sektor publik mengelola sumber daya dan memberikan pelayanan kepada masyarakat. Pengelolaan data Aparatur Sipil Negara (ASN) menjadi salah satu aspek krusial yang menentukan keberhasilan reformasi birokrasi, mengingat perannya sebagai tulang punggung dalam perencanaan formasi pegawai, evaluasi kinerja, dan pengembangan karier [1]. Namun demikian, Pemerintah Provinsi Jawa Timur yang mengelola 232 unit kerja dengan ribuan pegawai masih menghadapi kendala serius dalam memastikan kelengkapan dan kemutakhiran data kepegawaian. Persoalan ini tercermin dari minimnya kelengkapan dokumen penting seperti Surat Pernyataan Melaksanakan Tugas (SPMT), Sumpah Jabatan PNS, Kartu ASN Virtual, dan data kepesertaan BPJS [2].

Ketidaklengkapan data kepegawaian membawa konsekuensi yang lebih luas dari sekadar permasalahan administratif. Kondisi ini berpotensi mengganggu proses pengambilan keputusan strategis, mengakibatkan distribusi pegawai yang tidak merata, serta menghambat upaya peningkatan transparansi dan akuntabilitas institusi pemerintah [3]. Sistem pengelolaan data konvensional yang masih mengandalkan pencatatan manual dan tersebar di berbagai unit kerja terbukti tidak lagi memadai untuk mengakomodasi kebutuhan analisis data yang cepat, akurat, dan komprehensif di era digital saat ini [4].

Perkembangan teknologi pembelajaran mesin membuka peluang baru dalam mengatasi kompleksitas pengelolaan data kepegawaian. Teknik unsupervised learning, khususnya metode clustering, menawarkan kemampuan untuk mengidentifikasi pola dan struktur tersembunyi dalam dataset besar tanpa memerlukan pelabelan data sebelumnya [5], [6]. Algoritma K-Means telah terbukti efektif dalam melakukan segmentasi data berdasarkan karakteristik yang mirip, termasuk dalam konteks data kepegawaian [7], [8]. Sementara itu, Gaussian Mixture Model (GMM) dan Hierarchical Clustering memberikan fleksibilitas lebih dalam menangani data dengan distribusi yang kompleks dan tidak seragam [9], [10].

Penelitian ini mengusulkan pengembangan model clustering terintegrasi dengan sistem visualisasi interaktif untuk data ASN. Model ini diharapkan mampu mengidentifikasi unit kerja atau kelompok ASN yang memerlukan prioritas pembenahan data, sekaligus menyediakan sistem monitoring untuk memantau progres perbaikan secara berkelanjutan [11],[12],[13],[14],[15],[16].

2. Metode

2.1. Dataset dan Variabel Penelitian

Penelitian ini menggunakan dataset Aparatur Sipil Negara (ASN) Pemerintah Provinsi Jawa Timur yang terdiri dari 18.962 entri pegawai aktif per Februari 2025. Dataset dikumpulkan dari 232 unit kerja yang tersebar di 38 wilayah administratif, mencakup 29 kabupaten dan 9 kota. Data diperoleh dari sistem informasi kepegawaian Badan Kepegawaian Daerah dalam format Excel/CSV.

Tabel 1: Distribusi Dataset ASN Berdasarkan Perangkat Daerah

No	Nama Perangkat Daerah	Tingkat	Jumlah Unit	Data ASN	Persentase
1	Sekretariat Daerah	Induk	9	1.032	5,44%
2	Sekretariat DPRD	Induk	1	228	1,20%
3	Badan	Induk	14	262	1,38%
4	Dinas	Induk	24	246	1,30%
5	UPT	UPT	168	6.961	36,71%
6	Cabang Dinas	Cabang	14	904	4,77%
7	Inspektorat Provinsi	Induk	1	2.098	11,07%
8	Satuan Polisi Pamong Praja	Induk	1	7.231	38,14%
Total			232	18.962	100%

Sumber: Sistem Informasi Manajemen ASN terintegrasi BKD Provinsi Jawa Timur (2025)

Pada Tabel 1, distribusi dataset menunjukkan bahwa dari 232 unit kerja yang tersebar di 38 wilayah administratif, Satuan Polisi Pamong Praja (Satpol PP) memiliki jumlah ASN terbesar dengan 7.231 pegawai (38,14%), diikuti oleh Unit Pelaksana Teknis (UPT) dengan 6.961 pegawai (36,71%).

Variabel yang digunakan untuk clustering terdiri dari 12 atribut kelengkapan dokumen kepegawaian dalam format biner (1=ada, 0=tidak ada), yaitu: foto ½ badan, foto full body, akta lahir, KTP, NPWP, sumpah jabatan PNS, nota BKN, SPMT, kartu ASN virtual, nomor NPWP, nomor BPJS, dan nomor KK.

2.2 Tahapan Preprocessing Data

Preprocessing data dilakukan melalui beberapa tahap sistematis. Pertama, pembersihan data dengan mengidentifikasi dan menghapus 287 data duplikat berdasarkan NIP. Kedua, penanganan missing values dengan mengisi atribut dokumen kosong dengan nilai 0 (tidak ada), serta mengeluarkan

153 entri dengan missing values lebih dari 50% untuk menjaga kualitas clustering. Ketiga, transformasi data melalui encoding status kelengkapan dokumen ke format biner.

Normalisasi data dilakukan menggunakan metode Min-Max dengan rumus:

$$X_{\text{norm}} = (X - X_{\text{min}}) / (X_{\text{max}} - X_{\text{min}}) \quad (1)$$

Feature engineering dilakukan dengan membuat variabel agregat "Completeness_Percentage" yang menghitung persentase kelengkapan dokumen per pegawai:

$$\text{Completeness}_{\text{Percentage}} = \left(\frac{\sum \text{dokumen}_{\text{ada}}}{12} \right) \times 100\% \quad (2)$$

2.3 Implementasi Algoritma Clustering

2.3.1 K-Means Clustering

Algoritma K-Means bertujuan meminimalkan within-cluster sum of squares (WCSS) melalui fungsi objektif.

$$J = \sum_{k=1}^K \sum_{x_i \in C_k} ||x_i - \mu_k||^2 \quad (3)$$

dimana J adalah fungsi objektif, K adalah jumlah cluster, C_k adalah himpunan data dalam cluster ke-k, x_i adalah data point ke-i, dan μ_k adalah centroid cluster ke-k.

2.3.2 Gaussian Mixture Model (GMM)

GMM mengasumsikan data berasal dari campuran distribusi Gaussian dengan model probabilistik,

$$p(x) = \sum_{k=1}^K \pi_k \times N(x|\mu_k, \Sigma_k) \quad (4)$$

dimana π_k adalah mixing coefficient dengan $\sum \pi_k = 1$, dan $N(x|\mu_k, \Sigma_k)$ adalah distribusi Gaussian dengan mean μ_k dan kovarians Σ_k .

2.3.3 Hierarchical Clustering

Hierarchical Clustering menggunakan pendekatan aglomeratif dengan metode single linkage.

$$d(A, B) = \min\{\text{dist}(a, b) \mid a \in A, b \in B\} \quad (5)$$

dimana $d(A, B)$ adalah jarak antara cluster A dan B.

2.4 Metrik Evaluasi Clustering

Evaluasi menggunakan tiga metrik internal yang sesuai untuk *unsupervised learning*. Silhouette Score mengukur seberapa mirip suatu objek dengan clusternya sendiri dibandingkan dengan cluster lain, dengan nilai berkisar -1 hingga 1 dimana nilai positif menunjukkan clustering yang baik. Calinski-Harabasz Index mengukur rasio antara between-cluster dispersion dan within-cluster dispersion, dimana nilai yang lebih tinggi menunjukkan cluster yang lebih terpisah dengan baik. Davies-Bouldin Index mengukur rata-rata similaritas maksimum antar cluster, dengan nilai yang lebih rendah menunjukkan clustering yang lebih baik.

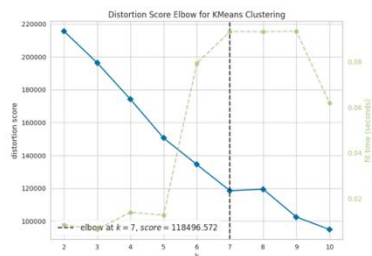
2.5 Implementasi Sistem

Sistem clustering data ASN diimplementasikan menggunakan Map Data Viewer berbasis web dengan teknologi HTML/JavaScript untuk frontend, Nodejs/Express.js untuk backend, dan Python untuk algoritma clustering. Sistem memungkinkan upload file Excel data ASN, menjalankan algoritma clustering (K-Means, GMM, Hierarchical) melalui script Python, dan menampilkan hasilnya pada peta interaktif 38 wilayah Jawa Timur. Visualisasi hasil mencakup scatter plot cluster, PCA, dan dendrogram yang dapat diakses melalui interface web. Sistem di-deploy pada VPS Ubuntu 22.04 LTS dengan data 18.962 ASN disimpan dalam format Excel.

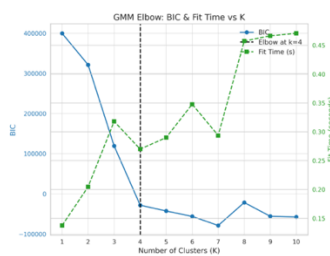
3. Hasil dan Pembahasan

3.1 Penentuan Jumlah Cluster Optimal

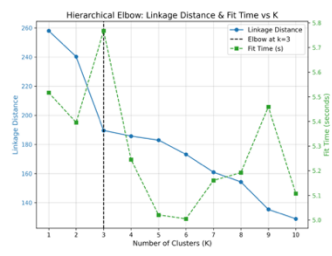
Penentuan jumlah cluster optimal dilakukan menggunakan metode elbow dengan kriteria evaluasi yang berbeda untuk setiap algoritma.



Gambar 1: Grafik Elbow Untuk K-Means



Gambar 2: Grafik Elbow Untuk Gaussian Mixture



Gambar 3: Grafik Elbow Untuk Hierarchical Clustering

Pada Gambar 1 Grafik Elbow K-Means, Gambar 1 menunjukkan penurunan distortion score terhadap jumlah cluster. Titik elbow terlihat pada k=7 dengan distortion score 119.496, dimana penurunan mulai melambat signifikan.

Pada Gambar 2 Grafik Elbow untuk Gaussian Mixtur Model, Gambar 2 Menunjukkan Bayesian Information Criterion (BIC) mencapai nilai minimum 119.839 pada k=4, menunjukkan keseimbangan optimal antara kompleksitas model dan goodness of fit.

Gambar 3 Grafik Elbow Method untuk Hierarchical Clustering ,Gambar 3 menampilkan linkage distance yang menunjukkan peningkatan signifikan setelah 3 cluster, mengindikasikan struktur alami data terbagi dalam 3 kelompok utama.

3.2. Hasil Evaluasi dan Visualisasi Clustering

Penentuan jumlah cluster optimal menggunakan metode elbow menghasilkan k=7 untuk K-Means (distortion score 119.496), k=4 untuk GMM (BIC 119.839), dan k=3 untuk Hierarchical Clustering.

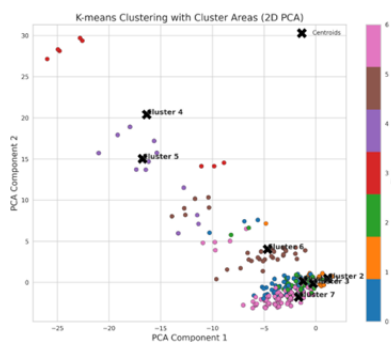
Tabel 2: Hasil Evaluasi Metrik Clustering

Algoritma	K	Silhouette Score	Calinski-Harabasz Index	Davies-Bouldin Index
K-Means	7	0.332	3412.783	1.224
GMM	4	0.263	2241.997	1.992
Hierarchical	3	0.282	3195.008	1.710

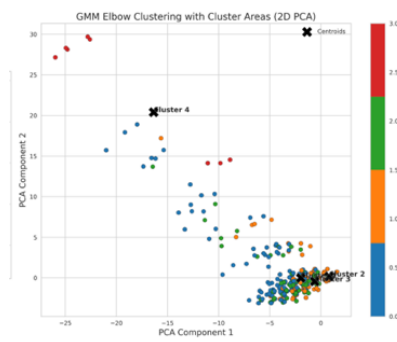
Pada Tabel 2, hasil evaluasi metrik clustering menunjukkan perbandingan nilai metrik evaluasi untuk ketiga algoritma clustering. K-Means memiliki performa terbaik berdasarkan ketiga metrik dengan Silhouette Score tertinggi (0.332), Calinski-Harabasz Index tertinggi (3412.783), dan Davies-Bouldin Index terendah (1.224).

3.3. Visualisasi PCA 2D Hasil Clustering

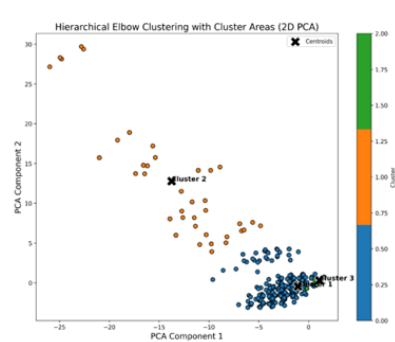
Visualisasi dilakukan menggunakan teknik Principal Component Analysis (PCA) 2D untuk memproyeksikan 12 atribut kelengkapan dokumen ASN ke dalam dua dimensi utama. Setiap algoritma clustering divisualisasikan dengan pewarnaan berbeda per cluster, serta penandaan centroid (untuk K-Means) dan area overlap (untuk GMM).



Gambar 4 : Visualisasi K-Means Clustering dalam PCA 2D



Gambar 5: Visualisasi GMM Clustering dalam PCA 2D



Gambar 6: Visualisasi Hierarchical Clustering dalam PCA 2D

Pada Gambar 4. Visualisasi PCA 2D K-Means Clustering, Gambar 4 menampilkan 7 cluster dengan separasi yang jelas. Cluster 1-2 (warna ungu-coklat) merepresentasikan kelompok High, cluster 3-5 (warna hijau-oranye) kelompok Medium, dan cluster 6-7 (warna biru) kelompok Low. Centroid ditandai dengan simbol khusus.

Pada Gambar 5, GMM menampilkan 4 cluster dengan batas probabilistik dan area overlap yang menandakan ASN dengan status ambigu. Cluster 4 dengan jumlah minimal menunjukkan outlier yang terdeteksi oleh pendekatan probabilistik GMM.

Pada Gambar 6, Hierarchical Clustering menghasilkan 3 cluster dengan separasi paling jelas dan minimal overlap, mencerminkan struktur hierarkis natural data kepegawaian

3.4. Hasil Klasifikasi K-Means Clustering

K-Means Clustering berhasil mengelompokkan 18.962 ASN ke dalam 7 cluster dengan karakteristik yang berbeda. Algoritma ini menghasilkan segmentasi yang paling detail dibandingkan metode lain, memungkinkan identifikasi kelompok ASN dengan tingkat kelengkapan dokumen yang bervariasi dari sangat tinggi hingga sangat rendah.

Tabel 3: Klasifikasi Cluster K-Means Berdasarkan Kelengkapan

Cluster	Jumlah ASN	Rata-rata Kelengkapan	Klasifikasi	Karakteristik Dominan
1	2.019	92.5%	High	PNS senior, unit kerja induk
2	4.613	88.3%	High	PNS terlantik penuh
3	4.216	76.8%	Medium	Campuran PNS/PPPK
4	3.847	74.2%	Medium	PPPK dominan
5	2.103	71.5%	Medium	CPNS masa transisi
6	1.218	45.3%	Low	PPPK baru
7	946	32.7%	Low	Data tidak lengkap

Pada Tabel 3, Tabel menunjukkan distribusi bahwa status kepegawaian (PNS vs PPPK) dan masa kerja menjadi faktor penentu utama dalam kelengkapan dokumen.

3.5 Hasil Klasifikasi GMM Clustering

Gaussian Mixture Model menghasilkan 4 cluster dengan pendekatan probabilistik yang memungkinkan identifikasi ASN dengan status ambigu. Metode ini memberikan fleksibilitas dalam menangani data yang berada di antara dua kategori dan berhasil mendeteksi outlier yang tidak teridentifikasi oleh metode lain.

Tabel 4: Klasifikasi Cluster GMM Berdasarkan Kelengkapan

Cluster	Jumlah ASN	Rata-rata Kelengkapan	Klasifikasi	Karakteristik Dominan
1	7.584	85.7%	High	PNS Senior
2	6.895	75.3%	Medium	PNS Junior
3	4.469	72.1%	Medium	PNS dan PPPK Parsial
4	14	15.2%	Low	Pegawai Khusus

Pada Tabel 4, Tabel menunjukkan temuan penting berupa identifikasi 14 ASN sebagai outlier dengan kelengkapan sangat rendah.

3.6 Hasil Klasifikasi Hierarchical Clustering

Hierarchical Clustering menghasilkan struktur 3 cluster yang mencerminkan hierarki alami organisasi kepegawaian. Metode ini memberikan gambaran makro yang sederhana namun informatif, memudahkan pemangku kepentingan dalam memahami distribusi kelengkapan data ASN secara keseluruhan.

Tabel 5: Klasifikasi Cluster Hierarchical Clustering

Cluster	Jumlah ASN	Rata-rata Kelengkapan	Klasifikasi	Karakteristik Dominan
1	8.997	82.4%	High	PNS dominan, dokumen lengkap
2	78	45.6%	Low	Kasus khusus, perlu investigasi
3	9.845	73.8%	Medium	Campuran status, kelengkapan sedang

Pada Tabel 5, Tabel menampilkan struktur 3 cluster yang memberikan gambaran makro yang lebih sederhana namun tetap informatif.

3.7 Analisis Pola Kelengkapan Dokumen

Analisis pola kelengkapan 12 dokumen kepegawaian dilakukan untuk mengidentifikasi prioritas intervensi pembaruan data.

Tabel 6: Analisis Kelengkapan per Jenis Dokumen

Dokumen	Kelengkapan Cluster High	Kelengkapan Cluster Medium	Kelengkapan Cluster Low
Foto ½ Badan	98.5%	89.3%	45.2%
Foto Full Body	97.8%	86.7%	42.1%
Akta Lahir	99.2%	91.5%	67.3%
KTP	99.7%	94.2%	71.5%
NPWP	95.3%	78.4%	31.2%
Sumpah Jabatan PNS	98.1%	82.3%	12.4%
Nota BKN	97.6%	79.8%	15.7%
SPMT	89.4%	65.2%	8.3%
Kartu ASN Virtual	78.3%	51.4%	5.1%
No NPWP	94.7%	76.8%	28.9%
No BPJS	96.2%	83.5%	52.3%
No KK	72.5%	48.3%	11.2%

Pada Tabel 6, tabel ini menunjukkan perbandingan kelengkapan 12 jenis dokumen kepegawaian antar cluster. Dokumen identitas dasar (KTP 99.7%, Akta Lahir 99.2%) dan dokumen pelantikan (Sumpah Jabatan PNS 98.1%, Nota BKN 97.6%) memiliki kelengkapan tertinggi di cluster High. Sementara Kartu ASN Virtual (5.1%) dan SPMT (8.3%) menunjukkan kelengkapan terendah di cluster Low, mengindikasikan kesulitan dalam penerapan sistem informasi kepegawaian.

3.8 Temuan Utama dan Implikasi

Analisis clustering menghasilkan empat temuan utama yang berimplikasi pada strategi pengelolaan data ASN.

Tabel 7: Ringkasan Temuan Berdasarkan Analisis Clustering

Aspek	Temuan	Implikasi Strategis
Distribusi Kelengkapan	High: 40-52%, Medium: 45-55%, Low: 1-20%	Mayoritas ASN memerlukan pembaruan dokumen parsial
Status Kepegawaian	PNS: 90% kelengkapan High, PPPK/CPNS: 70% kelengkapan Medium-Low	Prioritas program pembaruan untuk PPPK dan CPNS
Unit Kerja	Induk Surabaya: 54.27% High, Kabupaten tertentu: <5% High	Disparitas regional memerlukan intervensi targeted
Dokumen Kritis	Kartu ASN Virtual: <40% lengkap, No KK: <50% lengkap	Akselerasi digitalisasi dan integrasi sistem

Pada Tabel 7, Tabel ini merangkum disparitas antara PNS dan PPPK/CPNS serta kesenjangan regional yang menjadi fokus prioritas intervensi.

3.9 Perbandingan Kinerja Metode Clustering

Pemilihan metode optimal dilakukan melalui evaluasi terbobot lima kriteria adalah kualitas clustering, kecepatan komputasi, interpretabilitas, fleksibilitas, dan scalability.

Tabel 8: Matriks Keputusan Pemilihan Metode

Kriteria	Bobot	K-Means	GMM	Hierarchical
Kualitas Clustering (Metrik)	30%	9/10	6/10	7/10
Kecepatan Komputasi	20%	10/10	7/10	5/10
Interpretabilitas	25%	8/10	6/10	10/10
Fleksibilitas	15%	5/10	9/10	7/10
Scalability	10%	9/10	7/10	4/10
Skor Total Terbobot	100%	8.15	6.85	7.25

Pada tabel 8, tabel ini menunjukkan K-Means memperoleh skor tertinggi karena keunggulan dalam kualitas clustering dan kecepatan komputasi.

4. Kesimpulan

Penelitian berhasil mengimplementasikan tiga metode clustering untuk menganalisis kelengkapan data 18.962 ASN Pemerintah Provinsi Jawa Timur. Metode elbow menghasilkan cluster optimal k=7 untuk K-Means, k=4 untuk GMM, dan k=3 untuk Hierarchical Clustering. K-Means menunjukkan performa terbaik dengan Silhouette Score 0.332.

K-Means mengidentifikasi 35% ASN kategori High, 45% Medium, dan 20% Low. Temuan kritis menunjukkan PNS memiliki kelengkapan 90% lebih baik dari PPPK/CPNS, unit kerja Surabaya mencapai 54.27% dibanding kabupaten <5%, serta Kartu ASN Virtual dan Nomor KK memiliki kelengkapan terendah.

Tabel 9: Ringkasan Rekomendasi Strategis Berdasarkan Hasil Clustering

Kategori	Jumlah ASN	Prioritas	Strategi Intervensi	Target Peningkatan
High (>80%)	7.000-9.800	Pemeliharaan	Sistem pengingatan, berbagi praktik terbaik	Pertahankan >95%
Medium (70-79%)	8.500-10.400	kecepatan	Program pembaruan target, pendampingan	+15% per kuartal
Low (<70%)	950-3.800	Intensif	Tim khusus, integrasi sistem nasional	+20% per kuartal

Pada tabel 9, tabel ini menunjukkan implementasi sistem monitoring berbasis clustering yang memungkinkan instansi kepegawaian melakukan intervensi efektif sesuai tingkat urgensi.

Model clustering ini memungkinkan instansi kepegawaian melakukan intervensi efektif melalui dashboard dan dapat direplikasi untuk mendukung transformasi digital birokrasi nasional.

Implementasi model ini menghadapi beberapa kendala, terutama keterbatasan aplikasi yang user-friendly untuk pengguna non-teknis serta minimnya panduan operasional yang mudah dipahami bagi operator di tingkat unit kerja. Hal ini dapat menghambat adopsi sistem oleh pengguna yang tidak memiliki latar belakang teknologi informasi.

Penelitian lanjutan disarankan mengembangkan antarmuka pengguna yang lebih intuitif dan menyediakan modul pelatihan terstruktur untuk meningkatkan literasi digital pengelola data kepegawaian di berbagai tingkatan.

Pustaka

- [1] A. Ardiansyah *et al.*, "K-Means Clustering Application of Open Unemployment in 2020 Caused by COVID-19 in West Java Province," *Enthusiastic: Int. J. Appl. Stat. Data Sci.*, pp. 1–12, 2024.
- [2] M. Barata, I. S. Ayuni, A. Y. Kartini, and Z. Alawi, "Algoritma K-Means dalam Clustering Produk Skincare untuk Menentukan Strategi Pemasaran," *J. Inform. Polinema*, vol. 10, no. 3, pp. 421–428, 2024.
- [3] A. Basalamah and R. Setyadi, "Penerapan algoritma K-Means clustering pada tingkat penyelesaian pendidikan di Provinsi Indonesia," *J. Inform. Teknol. Komput. (J-ICOM)*, vol. 4, no. 2, pp. 114–121, 2023.
- [4] M. A. M. Bisri and R. Kurniawan, "Analisis Pengelompokan Data Aparatur Sipil Negara Kabupaten Cirebon Berdasarkan Perangkat Daerah dan Usia Menggunakan Metode K-Means Clustering," *JATI (J. Mahas. Tek. Inform.)*, vol. 8, no. 2, pp. 1500–1506, 2024.
- [5] R. H. Carver, *Cluster Analysis in the Public Sector: Exploratory Data Analysis (EDA), Principal Component Analysis (PCA), and Cluster Analysis*. Cary, NC: JMP Statistical Discovery LLC, 2022.
- [6] B. Dash *et al.*, "A Hybridized K-means Clustering Approach for High Dimensional Dataset," *Int. J. Eng. Sci. Technol.*, vol. 2, no. 2, pp. 59–66, 2010.
- [7] J. Dingelstad, R. T. Borst, and A. Meijer, "Hybrid Data Competencies for Municipal Civil Servants: An Empirical Analysis of The Required Competencies for Data-Driven Decision-Making," *Public Pers. Manag.*, vol. 51, no. 4, pp. 458–490, 2022.
- [8] G. D. Zion and B. K. Tripathy, "Comparative Analysis of Tools for Big Data Visualization and Challenges," in *Data Visualization: Trends and Challenges Toward Multidisciplinary Perception*, Springer, 2020, pp. 33–52.
- [9] E. Emigawaty *et al.*, "K-Means Clustering Algorithm for Partitioning the Openness Levels of Open Government Data Portals," *JOIV: Int. J. Inform. Vis.*, vol. 7, no. 3, pp. 977–983, 2023.
- [10] R. A. Faujia and M. Z. Subarkah, "Analisis Cluster K-Means dan Visualisasi Data Spasial Berdasarkan Karakteristik Persebaran Covid-19 dan Pelanggaran Protokol Kesehatan di Jawa Tengah," in *Seminar Nasional Official Statistics*, vol. 2022, no. 1, pp. 813–822, Nov. 2022.
- [11] S. A. Hermawan *et al.*, "Visualisasi data pengelompokan kelulusan mahasiswa dengan algoritma clustering k-means," *J. Spektrum*, vol. 11, no. 1, 2024.
- [12] E. K. Korir, "Comparative clustering and visualization of socioeconomic and health indicators: A case of Kenya," *Socio Econ. Plan. Sci.*, vol. 95, p. 101961, 2024.
- [13] A. D. Kuswanto, A. N. Fadhila, P. T. Setiawan, M. K. Setiawan, and D. R. Syahputra, "Penerapan K-Means Clustering Untuk Menentukan Jumlah Pengangguran Berdasarkan Umur: Studi Kasus Di Badan Statistik Provinsi DKI Jakarta 2020-2022," *Repeater: Publikasi Teknik Informatika dan Jaringan*, vol. 2, no. 3, pp. 135–146, 2024.

- [14] B. Lund and J. Ma, "A Review of Cluster Analysis Techniques and Their Uses in Library and Information Science Research: K-Means and K-Medoids Clustering," *Perform. Meas. Metrics*, vol. 22, no. 3, pp. 161–173, 2021.
- [15] E. D. Madyatmadja et al., "Data Visualization of Internet Usage in The Jabodetabek Area," *Infotech: J. Technol. Inform.*, vol. 7, no. 1, pp. 55–62, 2021.
- [16] E. D. Madyatmadja et al., "Data Visualization of Priority Region Based on Community Complaints in Government," *ICIC Express Lett. Part B Appl.*, vol. 12, pp. 957–964, 2021.