

Klasterisasi Puskesmas dengan *K-Means* Berdasarkan Data Kualitas Kesehatan Keluarga dan Gizi Masyarakat

Bakhtiyar Hadi Prakoso^{*1}, Ervina Rachmawati², Demiawan Rachmatta Putro Mudiono³,
Veronika Vestine⁴, Gandu Eko Julianto Suyoso⁵

^{1,2,3,4,5}Program Studi Manajemen Informasi Kesehatan, Jurusan Kesehatan, Politeknik Negeri Jember,
Jember 68121, Jawa Timur, Indonesia

Email: ¹bakhtiyar.hp@polije.ac.id, ²ervina_rachmawati@polije.ac.id,

³demiawanrpm@polije.ac.id, ⁴veronikavestine@polije.ac.id, ⁵gandu.eko.js@polije.ac.id

Abstract. *One of the fundamental principles followed by the Jember Health Office for decision-making is data. Data plays a crucial role in the decision-making process. Raw data is more difficult to interpret and needs to be analyzed. Clustering is one of the techniques used for analysis. This study discusses using K-Means to cluster Public Health Center data based on AKI, AKB, and stunting prevalence. The data is processed by reducing dimensions and normalizing them. The clustering process is performed using the K-Means method, where the maximum k-value is obtained by calculating WCSS. The clustering process results in three clusters of Public Health Centers in the Jember Regency. These clusters can serve as a reference for the Jember Health Office to formulate family health and community nutrition quality policies.*

Keywords: *data mining, K-Means, clustering, Maternal Mortality Rate, Infant Mortality Rate, the prevalence of stunting*

Abstrak. *Salah satu dasar pengambilan kebijakan oleh Dinas Kesehatan Jember adalah data. Data memiliki peran dalam proses pengambilan keputusan. Data mentah yang didapatkan lebih sulit untuk diinterpretasikan sehingga diperlukan analisis terhadap data tersebut. Salah satu analisis yang dapat digunakan adalah teknik klasterisasi. Pada penelitian ini akan dibahas penggunaan K-Means untuk klasterisasi data puskesmas berdasarkan AKI, AKB, dan prevalensi stunting. Data diproses dengan melakukan reduksi dimensi dan normalisasi. Proses klasterisasi dilakukan dengan metode K-Means dimana nilai k maksimal diperoleh dengan menghitung WCSS. Adapun hasil proses klasterisasi didapatkan tiga kelompok klaster puskesmas yang terdapat di Kabupaten Jember. Hasil klasterisasi dapat digunakan sebagai referensi Dinas Kesehatan Jember dalam mengambil kebijakan terkait kualitas kesehatan keluarga dan gizi masyarakat*

Kata Kunci: *data mining, K-Means, klasterisasi, Angka Kematian Ibu, Angka Kematian Bayi, prevalensi stunting*

1. Pendahuluan

Tren perkembangan internet saat ini telah merata diberbagai daerah. Ada akses internet yang semakin mudah dijangkau oleh masyarakat mempunyai efek berkembangnya data dalam bentuk digital. Perkembangan data digital yang signifikan telah memunculkan istilah *big data*. Penggunaan *big data* untuk menganalisis pola data kesehatan menjadi isu yang menarik untuk dibahas [1]. Saat ini pencarian pola lebih banyak dilakukan dengan menggunakan metode *data mining*. Dimana metode ini terdiri dari berbagai macam metode seperti *clustering*, *classification*, *prediction*, dan *association* [2]. *Data mining* merupakan sebuah konsep yang dipergunakan untuk mencari sebuah pola tertentu dari sebuah himpunan data. Luaran dari proses *data mining* ini berupa pengetahuan yang dipergunakan sebagai dasar untuk membuat sistem cerdas [3]. Beberapa inovasi telah dihasilkan dari proses pencarian pola dengan menggunakan *data mining* ini. Dalam bidang kesehatan, telah dihasilkan sistem cerdas yang dapat mendeteksi atau mengklasifikasikan berbagai penyakit seperti kanker payudara, *diabetes melitus*, *alzheimer*, serta penyakit jantung dan pembuluh darah [4][5][6][7]. Pada metode asosiasi *data mining* dapat digunakan untuk mengidentifikasi keterikatan faktor resiko dalam sebuah penyakit. Sedangkan dalam metode *clustering*, *data mining* dapat digunakan untuk mengklasterisasi pasien terkait efek samping dari sebuah obat [8].

Jember adalah salah satu Kabupaten di Propinsi Jawa Timur yang mempunyai luas 3.306,689 km² yang terbagi menjadi 31 kecamatan 22, kelurahan, 226 desa dan jumlah penduduk 2.536.729 jiwa [9]. Berdasarkan angka tersebut, sarana kesehatan yang dimiliki pun cukup banyak. Adapun sarana kesehatan yang dimiliki adalah: a) Puskesmas. b) Rumah Sakit c) Sarana Kesehatan Bersumberdaya Masyarakat dalam bentuk desa siaga, Poskesdes, Posnbindu. Berdasarkan data Dinas Kesehatan Kabupaten Jember saat ini Kabupaten Jember telah memiliki sebanyak 51 Posyandu yang tersebar di berbagai kecamatan [10].

Salah satu tujuan pemerintah Kabupaten Jember di tahun 2021–2026 adalah “Meningkatnya Kualitas Kesehatan Masyarakat”. Hal tersebut secara eksplisit tertuang dalam sebuah Renstra Kesehatan Kabupaten Jember 2021–2026 yang salah satu sarannya adalah Meningkatkan Kualitas Kesehatan Keluarga dan Gizi Masyarakat. Sasaran yang telah ditetapkan tersebut, akan dinilai berdasarkan tiga indikator yaitu: 1) Angka Kematian Ibu (AKI); 2) Angka Kematian Bayi (AKB); dan 3) Persentase Balita Stunting. Berdasarkan sasaran yang telah ditetapkan maka upaya yang dilakukan pemerintah adalah dengan melakukan peningkatan upaya promosi kesehatan keluarga yang melibatkan berbagai elemen masyarakat. Peningkatan deteksi dini terkait dengan permasalahan kesehatan keluarga dengan memanfaatkan teknologi informasi, pemanfaatan teknologi informasi untuk mengelola data kesehatan dan mengurangi tingkat *stunting* [11].

Pemerintah Kabupaten Jember melalui Dinas Kesehatan Kabupaten Jember mempunyai peranan utama dalam proses pengambilan kebijakan terkait dengan kebijakan kesehatan. Untuk membuat sebuah kebijakan perlu adanya kajian yang mendalam dari berbagai aspek dimana sumber utamanya adalah data. Data mempunyai peranan yang sangat signifikan dalam proses pengambilan keputusan [12]. Data mentah lebih sulit untuk diinterpretasikan sehingga dalam hal ini diperlukan analisis untuk mengetahui pola datanya. Jika data sudah diketahui polanya maka data akan lebih informatif untuk dasar pengambilan kebijakan [3].

Perkembangan penggunaan *data mining* di Indonesia saat ini masih tergolong cukup rendah untuk analisis data kesehatan. Di Indonesia saat ini pemanfaatan teknologi informasi dalam bidang kesehatan masih sebatas pada pengembangan sistem informasi atau aplikasi. Kondisi sebaliknya, di negara-negara maju sudah mulai menerapkan *data mining* untuk meningkatkan pelayanan kesehatan [13]. Kondisi ini menjadi sebuah peluang bagi Indonesia untuk mulai memanfaatkan *data mining* dalam analisis data kesehatan.

Clustering merupakan metode *data mining* yang dapat digunakan untuk menganalisis data kesehatan. Beberapa penelitian terkait dengan *clustering* data kesehatan telah dilakukan. Abbas dkk telah melakukan analisis klusterisasi terhadap data kelahiran bayi [14]. Teknik *clustering* juga digunakan untuk mengelompokkan perkembangan dari penyakit pasien yang terdapat di rumah sakit [15]. Milka dkk telah melakukan analisis terhadap data fasilitas kesehatan berdasarkan performa yang dimiliki [16]. Proses *clustering* juga dapat dilakukan untuk mengelompokkan rumah sakit berdasarkan tingkat fasilitas yang dimiliki [17].

Metode *clustering* mempunyai beragam jenis, salah satunya adalah metode *K-Means* [18]. Metode *K-Means* termasuk dalam kategori *partition clustering*, dimana keunggulannya adalah dapat digunakan untuk menyelesaikan permasalahan dengan data besar [19]. Dalam penelitian ini dibahas penggunaan *K-Means* yang dapat digunakan untuk klusterisasi data kesehatan. Data bersumber pada data angka kematian ibu dan anak, serta data prevalensi *stunting* di Kabupaten Jember. Topik penelitian ini diambil karena merupakan bagian dari tantangan dihadapi oleh Dinas Kabupaten Jember saat ini, dimana salah satunya adalah meningkatnya kualitas kesehatan keluarga dan gizi masyarakat. Adapun tujuan dari proses klusterisasi ini adalah mengelompokkan berdasarkan tingkat kemiripan antara anggota. Selanjutnya data hasil kluster dapat digunakan sebagai referensi Dinas Kesehatan Kabupaten Jember dalam mengambil sebuah keputusan sebagai dasar untuk membuat kebijakan dalam hal peningkatan Kualitas Kesehatan Keluarga dan Gizi Masyarakat.

2. Tinjauan Pustaka

2.1 K-Means Clustering

Clustering merupakan sebuah metode *data mining* yang termasuk dalam kategori *unsupervised learning*. Proses pembelajarannya tidak membutuhkan kelas. Metode *K-Means* banyak digunakan untuk menyelesaikan kasus klusterisasi. Adapun algoritma *K-Means* dapat dijelaskan pada Kode 1 sebagai berikut[3].

Kode 1. Pseudocode K-Means

Input:

- *k*: jumlah klaster
- *D*: Data set yang berisi objek sebanyak *n*

Output: Klaster-klaster yang berisikan objek-objek

Method:

- 1) Memilih titik secara random sebanyak *k* buah dari data *D* yang dipergunakan untuk pusat dari masing-masing kelompok
- 2) **Repeat**
- 3) Tempatkan objek dikuster berdasarkan kedekatan dengan pusat cluster. Pusat cluster dihitung dari rata-rata object anggota cluster
- 4) Lakukan perhitungan rata rata untuk masing-masing cluster
- 5) **Until** tidak ada perubahan rata

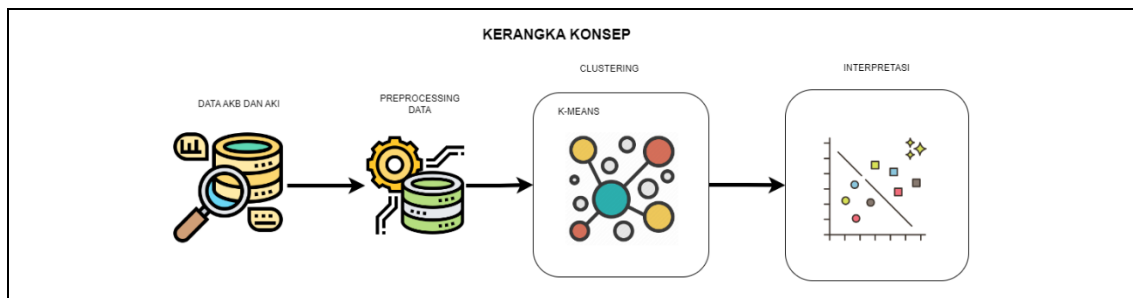
2.2 Preprocessing data

Tahapan *preprocessing* data merupakan tahapan yang penting didalam proses klusterisasi karena menentukan hasil klaster. Proses ini diawali dari data mentah yang di normalisasi dengan teknik *min-max*. Pada penelitian ini akan digunakan teknik normalisasi data dengan skala 0 s/d 1. Secara matematis rumus untuk normalisasi data dapat dilihat pada Persamaan 1 [20].

$$\frac{x - \min(x)}{\max(x) - \min(x)} \tag{1}$$

3. Metodologi Penelitian

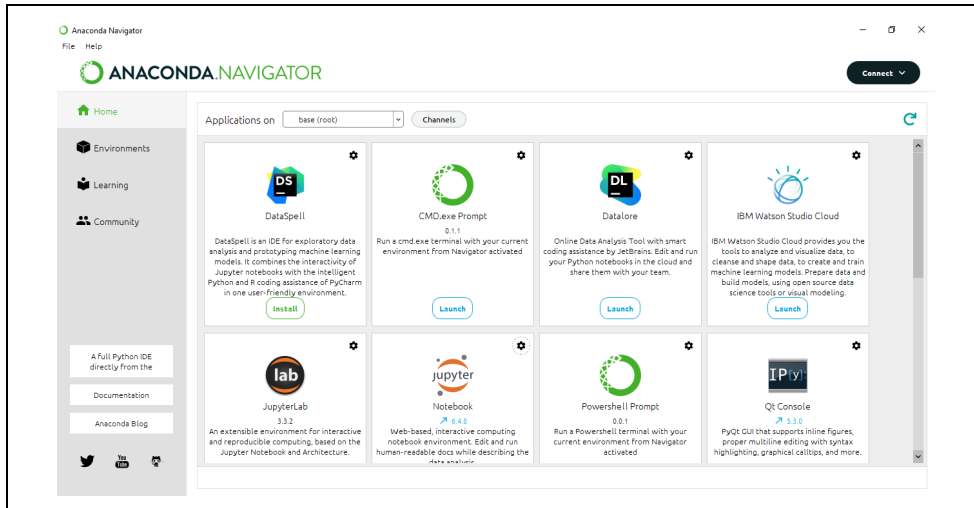
Penelitian ini bertujuan untuk menemukan sebuah pola data dalam bentuk klaster. Hasil klaster selanjutnya akan diinterpretasikan hasil dalam bentuk gambar dan tabel. Adapun kerangka konsep dalam penelitian ini dijelaskan pada Gambar 1.



Gambar 1. Kerangka Kosep Penelitian

Sebelum proses klusterisasi, data yang didapatkan terlebih dahulu dilakukan *preprocessing* untuk merduksi dimensi dan dilakukan normalisasi. Proses *clustering* dilakukan dengan menggunakan metode *K-Means*. Hasil klusterisasi akan dinterpretasikan dalam bentuk gambar *cluster* dan tabel, dimana hasil nya dapat digunakan sebagai referensi Dinas Kesehatan Kabupaten Jember dalam mengambil kebijakan.

Data penelitian berfokus pada permasalahan Kualitas Kesehatan Keluarga dan Gizi Masyarakat. Secara spesifik data penelitian yang diambil berupa Angka Kematian Ibu (AKI), Angka Kematian Bayi (AKB), dan *prevalensi stunting* pada tiap puskesmas di Kabupaten Jember. Data diambil dalam rentang tahun 2021 yang bersumber pada dokumen profil kesehatan Kabupaten Jember tahun 2021 yang terdapat pada laman Dinas Kesehatan Kabupaten Jember.



Gambar 2. Aplikasi Ananconda

Proses analisis klasterisasi dilakukan dengan menggunakan bahasa pemrograman *python* yang ter-install pada aplikasi *anaconda*. Seperti yang terlihat pada Gambar 2, aplikasi *anaconda* berisi tools yang dipergunakan untuk analisis *data mining*. Adapun didalamnya berisi *tools* seperti *jupyter notebook* dan beberapa *library python* seperti *sklearn*, *numpy*, *pandas*, dan *matplotlib*.

4. Hasil dan Diskusi

Pada bagian hasil dan diskusi dibahas analisis data. Terdiri dari beberapa sub bab, yaitu *preprocessing*, optimalisasi nilai *k* pada *K-Means*, dan analisis *K-Means*.

4.1 Preprocessing

Data yang didapatkan dari laporan Dinas Kesehatan Kabupaten Jember terdiri dari tiga indikator utama yaitu Angka Kematian Ibu (AKI), Angka Kematian Bayi (AKB), dan presentase balita *stunting*. Pada data Angka Kematian Ibu (AKI) terdapat tiga kolom utama yaitu kematian ibu hamil, kematian ibu bersalin, dan kematian ibu nifas. Masing-masing kolom dikategorikan lagi berdasarkan tingkat umur yaitu kategori umur kurang dari 20 tahun umur antara 20 sampai 34 tahun, dan umur lebih besar dari sama dengan 35 tahun. Detail data tersaji di dalam Tabel 1 dimana data terdiri dari 50 data puskesmas.

Tabel 1. Data Jumlah Kematian Ibu

No	Puskesmas	Angka Kematian Ibu (AKI)								
		Ibu Hamil			Ibu Bersalin			Ibu Nifas		
		< 20 tahun	20-34 tahun	>= 35 tahun	< 20 tahun	20-35 tahun	>= 35 tahun	< 20 tahun	20-35 tahun	>= 35 tahun
1	Kencong	0	0	0	0	0	0	0	0	0
2	Cakru	0	2	1	0	0	0	0	0	0
3	GumukMas	0	1	1	0	0	0	0	0	0
.....										
48	Gladak Pakem	0	0	1	0	0	0	0	2	0
49	Patrang	0	0	1	0	0	0	0	2	0
50	Banjarsengon	0	0	1	0	0	0	0	1	1

Sedangkan untuk Angka Kematian Anak (AKI) terbagi menjadi dua kategori utama, yaitu laki-laki dan perempuan. Masing-masing kategori terbagi lagi menjadi bayi dan balita. Kolom

Prevalensi Stunting merupakan nilai persentase terjadinya *stunting* untuk masing-masing puskesmas. Detail data tersaji pada Tabel 2.

Tabel 2. Data Jumlah Kematian Bayi dan Prevalensi *Stunting*

No	Puskesmas	Angka Kematian Bayi (AKB)		Prevalensi <i>Stunting</i>
		Laki-Laki	Perempuan	
1	Kencong	0	0	4,9
2	Cakru	2	1	5,7
3	GumukMas	2	2	7,8
.....				
48	Gladak Pakem	1	0	2,4
49	Patrang	6	4	1,7
50	Banjarsengon	6	3	6,3

Berdasarkan data yang didapatkan, terdapat banyak atribut sehingga dimensi data cukup besar. Data yang berdimensi besar akan berpengaruh terhadap hasil dari proses klusterisasi, dimana hasil yang didapatkan kurang begitu akurat [21]. Oleh karena itu perlu dilakukan reduksi dimensi data. Proses reduksi dilakukan dengan menggabungkan beberapa atribut yang terkait ke dalam satu buah atribut. Sesuai dengan tiga indikator utama yang telah disebutkan dalam pembahasan sebelumnya untuk atribut angka kematian ibu hamil, ibu bersalin, ibu nifas akan digabung menjadi satu buah atribut yaitu Angka Kematian Ibu (AKI). Begitu juga untuk angka kematian bayi laki-laki maupun perempuan akan digabung menjadi satu buah atribut yaitu Angka Kematian Bayi (AKB), sedangkan untuk atribut prevalensi stunting tidak ada penggabungan atribut. Pada akhirnya terdapat tiga buah atribut utama yang dipergunakan untuk perhitungan yaitu: Angka Kematian Ibu (AKI), Angka Kematian Bayi (AKB), dan Prevalensi *Stunting*. Data hasil penggabungan dapat dilihat pada Tabel 3.

Tabel 3. Data Setelah Proses Reduksi Atribut

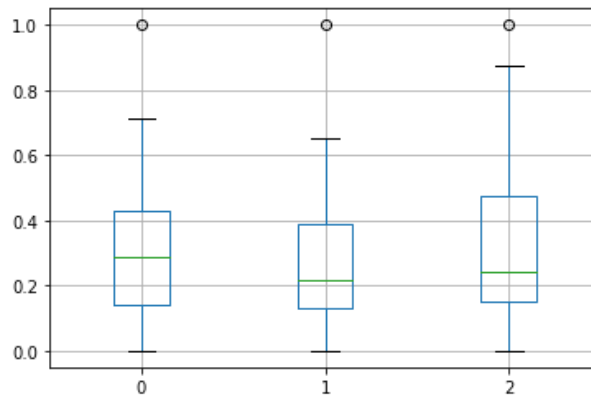
No	Puskesmas	AKI	AKB	Prevalensi <i>Stunting</i> (%)
1	Kencong	0	0	4,9
2	Cakru	3	3	5,7
3	GumukMas	2	4	7,8
.....				
48	Gladak Pakem	3	1	2,4
49	Patrang	3	10	1,7
50	Banjarsengon	3	9	6,3

Tahapan selanjutnya yaitu melakukan *preprocessing* data terhadap data Tabel 3. Tahapan ini merupakan tahapan yang penting, karena pada akhirnya akan berpengaruh terhadap hasil *clustering*. *Preprocessing* dilakukan dengan menggunakan teknik *normalization*. Hasil perhitungan *normalization* terhadap data di Tabel 3 dapat dilihat di Tabel 4.

Tabel 4. Data Hasil *Preprocessing*

No	Puskesmas	AKI	AKB	Prevalensi <i>Stunting</i>
1	Kencong	0,25	0,17	0,22
2	Cakru	0,25	0,26	0,22
3	GumukMas	0,00	0,22	0,26
.....				
48	Gladak Pakem	0,00	0,13	0,24
49	Patrang	0,25	0,26	0,00
50	Banjarsengon	0,00	0,30	0,11

Setelah melalui tahapan *preprocessing* data, proses selanjutnya yaitu mendeskripsikan data dalam bentuk *box-plot*. Hal ini dilakukan untuk mengetahui jumlah *outlier* pada data. Hasil *box-plot* terhadap data dapat dilihat pada Gambar 3.



Gambar 3. Box-plot Data AKI, AKB, dan Prevalensi Stunting

Terlihat bahwa dari hasil *ploting* dengan menggunakan *box-plot* terlihat beberapa *outlier* dari ketiga atribut yaitu AKI, AKB, dan prevalensi *stunting*. Dimana dari ketiga atribut tersebut terdapat titik yang diluar nilai minimum maupun maksimum. Dari Gambar 3 diketahui bahwa terdapat *outlier* pada masing-masing atribut. Analisis *outlier* ini perlu dilakukan untuk membantu analisis klusterisasi. Karena jika terdapat *outlier* pada data maka dapat berpengaruh pada hasil klusterisasi [19].

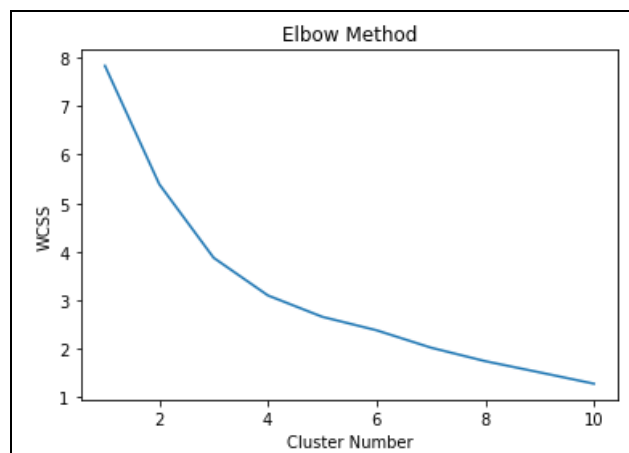
3.2 Optimalisasi K-Means

Hasil klusterisasi dengan menggunakan *K-Means* sangat dipengaruhi oleh nilai *k*, dimana nilai *k* merupakan jumlah klaster. Untuk mengetahui nilai *k* paling optimal maka dilakukan perhitungan dengan metrik *WCSS* (*Whitin Cluster Sum of Square*) dengan Persamaan 2 dibawah ini.

$$CSS = \sum_{P_i \text{ in cluster}_1} dist(P_i C_1)^2 + \sum_{P_i \text{ in cluster}_2} dist(P_i C_2)^2 + \dots + \sum_{P_i \text{ in cluster}_n} dist(P_i C_n)^2 \quad (2)$$

Dimana pada persamaan diatas akan dihitung kuadrat jarak *Pi* dengan Pusat klaster satu ditambah dengan jarak *Pi* dengan klaster dua, dan seterusnya sampai *n* klaster. Nilai klaster yang paling optimal didapatkan pada perhitungan dari *WCSS* yang digambarkan pada grafik sumbu *x* dan *y*, dimana *k* yang paling maksimal diperoleh dengan melihat nilai *WCSS* yang mempunyai sudut siku signifikan.

Uji coba dilakukan dengan mengambil nilai sampel nilai *k* =1 sampai dengan *k*=10. Adapun hasil perhitungan dari *WCSS* dapat dilihat pada Gambar 4 dibawah ini.



Gambar 4. Grafik Perhitungan *WCSS* (*Whitin Cluster Sum of Square*)

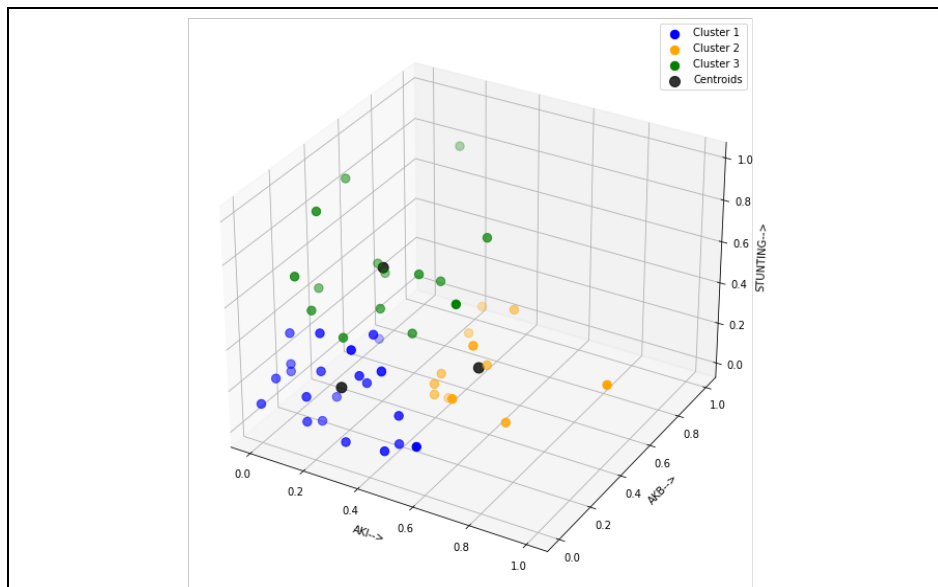
Berdasarkan Gambar 4 dapat dianalisis bahwa terdapat satu buah titik yang memiliki sudut siku yang cukup signifikan tingkat sudutnya dibandingkan dengan yang lainnya, yaitu terletak pada titik *cluster number* 3. Dengan demikian dapat disimpulkan bahwa nilai *k* yang paling optimal untuk digunakan pembagian kluster adalah tiga.

3.2 Proses Klasterisasi

Proses klasterisasi dilakukan dengan metode *K-Means*. Dari langkah sebelumnya diperoleh nilai *k* maksimal sebesar tiga. Konsep dari metode dari *K-Means* ini adalah mencari tingkat kemiripan dengan menghitung jarak titik ke pusat kluster. Jarak antara titik ke pusat cluster dihitung dengan menggunakan persamaan *ecludian distance*, secara matematis dapat dihitung melalui Persamaan 3 [22].

$$dist(C, X) = \sqrt{\sum_{i=0}^n (C_i - X_i)^2} \quad (3)$$

Adapun secara visual hasil klasterisasi data AKB, AKI, dan Prevalensi *Stunting* dapat dilihat pada Gambar 5. Pada Gambar 5 titik hitam menunjukkan pusat kluster (*centroids*), sedangkan warna biru, merah dan hijau merupakan anggota kluster satu, dua, dan tiga.



Gambar 5. Hasil Ploting Klasterisasi terhadap Data

Berdasarkan hasil analisis *WCSS* pada pembahasan sebelumnya, diperoleh jumlah kluster yang paling optimal digunakan adalah sebanyak tiga kluster. Hasil klasterisasi menunjukkan bahwa puskesmas yang terdapat di Kabupaten Jember terbagi menjadi tiga buah kluster. Kluster pertama merupakan daerah berwarna biru. Daerah ini mempunyai ciri khas untuk nilai AKB tergolong rendah dan tingkat prevalensi stunting cukup rendah. Pada kluster kedua yang ditandai warna kuning menunjukkan bahwa tingkat AKB yang tergolong cukup tinggi dan tingkat prevalensi stunting cukup rendah. Sedangkan untuk kluster terakhir yaitu kluster yang berwarna hijau didominasi dengan nilai prevalensi stunting yang cukup tinggi. Dari gambar hasil klasterisasi terlihat bahwa pola kluster dapat terbentuk dengan baik, meskipun terdapat *outlier* pada masing-masing atribut. Hasil klasterisasi puskesmas secara detail dapat dilihat pada Tabel 6.

Tabel 6. Hasil Klasterisasi Puskesmas

Jenis Klaster	Nama Puskesmas
Klaster 1	Kencong, Gladak Pakem, Jember Kidul, Kaliwates, Sukorambi, Panti, Sukorejo, Klatakan, Tanggul, Sumberbaru, Jombang, Semboro, Karangduren, Ajung, Mayang, Umbulsari, Cakru, Silo 1, GumukMas, Tembok Rejo, Puger, Kasiyan, Ambulu
Klaster 2	Mumbulsari, Summersari, Sukowono, Wuluhan, Lojejer, Banjarsengon, Andongsari, Tempurejo, Jenggawah, Patrang, Sabrang, K Sari Kidul
Klaster 3	Patrang, Sabrang, Rowotengah, Rambipuji, Mangli, Nogosari, Jelbuk, Balung, Ledokombo, Kalisat, Pakusari, Arjasa, Curahnongko, Paleran, Silo, Sumberjambe, Bangsalsari

5. Kesimpulan dan Saran

Data yang didapatkan mempunyai dimensi yang cukup tinggi perlu dilakukan reduksi dimensi agar proses klusterisasi *K-Means* menghasilkan pola klaster yang maksimal. Dari hasil reduksi terdapat tiga buah atribut utama yaitu AKI, AKB, dan Prevalensi *Stunting*. Berdasarkan perhitungan dengan metode *WCCS*, didapatkan jumlah klaster yang paling optimal sebanyak tiga buah klaster. Meskipun pada masing-masing atribut terdapat *outlier* pola klusterisasi masih dapat terbentuk dengan baik. Hasil dari klusterisasi puskesmas ini dapat digunakan sebagai referensi Dinas Kabupaten Jember dalam mengambil kebijakan khususnya tentang kebijakan terkait kualitas kesehatan keluarga dan gizi masyarakat. Sebagai pembandingan dari hasil klusterisasi, untuk penelitian selanjutnya dapat menggunakan metode *clustering* yang lain seperti *metode hierarchy* atau *density clustering* yang mana dari metode tersebut dapat digunakan untuk data yang mempunyai *outlier*.

6. Ucapan Terima Kasih

Di dalam kegiatan penelitian ini penulis mengucapkan terimakasih kepada Politeknik Negeri Jember yang telah memberikan kesempatan bagi penulis melakukan kegiatan penelitian secara mandiri sehingga penelitian ini dapat selesai. Ucapan terima kasih juga ditujukan kepada Dinas Kesehatan Kabupaten Jember yang telah memberikan data untuk dijadikan sebagai objek dalam penelitian ini.

Referensi

- [1] L. Hong, M. Luo, R. Wang, P. Lu, W. Lu, and L. Lu, "Big Data in Health Care: Applications and Challenges," *Data Inf. Manag.*, vol. 2, no. 3, pp. 175–197, 2018, doi: 10.2478/dim-2018-0014.
- [2] M. Alloghani, D. Al-Jumeily, J. Mustafina, A. Hussain, and A. J. Aljaaf, "A Systematic Review on Supervised and Unsupervised Machine Learning Algorithms for Data Science," in *Supervised and Unsupervised Learning for Data Science*, M. W. Berry, A. Mohamed, and B. W. Yap, Eds. Cham: Springer International Publishing, 2020, pp. 3–21. doi: 10.1007/978-3-030-22475-2_1.
- [3] M. Bramer, *Principles of Data Mining*, vol. 30, no. 7. London: Springer London, 2020. doi: 10.1007/978-1-4471-7493-6.
- [4] D. Mining, B. Cancer, and B. Cancer, "Application of Data Mining Techniques to Predict Breast Cancer," *Procedia Comput. Sci.*, vol. 163, pp. 11–18, 2019, doi: 10.1016/j.procs.2019.12.080.
- [5] F. G. Woldemichael and S. Menaria, "Prediction of Diabetes Using Data Mining Techniques," *Proc. 2nd Int. Conf. Trends Electron. Informatics, ICOEI 2018*, no. Icoei, pp. 414–418, 2018, doi: 10.1109/ICOEI.2018.8553959.
- [6] M. Singh, "Classification system via data mining algorithm: New tool to diagnose Alzheimer's disease," *J. Neurol. Sci.*, vol. 405, no. 2019, pp. 161–162, 2019, doi: 10.1016/j.jns.2019.10.1084.
- [7] M. M. Ali, B. K. Paul, K. Ahmed, F. M. Bui, J. M. W. Quinn, and M. A. Moni, "Heart disease prediction using supervised machine learning algorithms: Performance analysis and comparison," *Comput. Biol. Med.*, vol. 136, no. July, p. 104672, 2021, doi: 10.1016/j.combiomed.2021.104672.

- [8] M. L. Kolling *et al.*, “Data mining in healthcare: Applying strategic intelligence techniques to depict 25 years of research development,” *Int. J. Environ. Res. Public Health*, vol. 18, no. 6, pp. 1–21, 2021, doi: 10.3390/ijerph18063099.
- [9] BPS (Badan Pusat Statistik), “Kabupaten Jember dalam Angka 2021 (Jember Regency in Figures),” pp. 1–630, 2020.
- [10] Dinkes Jember, “Profil Kesehatan Kabupaten Jember 2020,” *Profil Kesehat. Kabupaten Jember 2020*, pp. 1–12, 2020.
- [11] K. Jember, “Rencana Strategis Dinas Kesehatan Kabupaten Jember Tahun 2021 – 2026,” Jember, 2019.
- [12] K. Buse, N. Mays, and G. Walt, *Making Health Policy*, Second Edi. New York, NY: McGraw-Hill, 2012. doi: 10.1163/9789004333109.
- [13] M. Madjido, A. Espressivo, A. W. Maula, A. Fuad, and M. Hasanbasri, “Health information system research situation in Indonesia: A bibliometric analysis,” *Procedia Comput. Sci.*, vol. 161, pp. 781–787, 2019, doi: 10.1016/j.procs.2019.11.183.
- [14] S. A. Abbas, A. Aslam, A. U. Rehman, W. A. Abbasi, S. Arif, and S. Z. H. Kazmi, “K-Means and K-Medoids: Cluster Analysis on Birth Data Collected in City Muzaffarabad, Kashmir,” *IEEE Access*, vol. 8, pp. 151847–151855, 2020, doi: 10.1109/ACCESS.2020.3014021.
- [15] D. T. A. Luong and V. Chandola, “A K-Means Approach to Clustering Disease Progressions,” *Proc. - 2017 IEEE Int. Conf. Healthc. Informatics, ICHI 2017*, pp. 268–274, 2017, doi: 10.1109/ICHI.2017.18.
- [16] M. B. Gesicho, M. C. Were, and A. Babic, “Evaluating performance of health care facilities at meeting HIV-indicator reporting requirements in Kenya: an application of K-means clustering algorithm,” *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, pp. 1–18, 2021, doi: 10.1186/s12911-020-01367-9.
- [17] S. Cinaroglu, “Integrated k-means clustering with data envelopment analysis of public hospital efficiency,” *Health Care Manag. Sci.*, vol. 23, no. 3, pp. 325–338, Sep. 2020, doi: 10.1007/s10729-019-09491-3.
- [18] H. Santoso, H. Magdalena, H. Wardhana, and I. Artikel, “Aplikasi Dynamic Cluster pada K-Means Berbasis Web untuk Klasifikasi Data Industri Rumahan Web-based Application of Dynamic Cluster on K-Means for Classification of Home Industry Data,” *J. Manaj. , Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 3, pp. 541–554, 2022, doi: 10.30812/matrik.v21i3.1720.
- [19] A. Saxena *et al.*, “A review of clustering techniques and developments,” *Neurocomputing*, vol. 267, pp. 664–681, 2017, doi: 10.1016/j.neucom.2017.06.053.
- [20] I. H. Witten, *Data Mining- Practical Machine Learning Tools and Techniques*, Fourth Edi. Cambridge: Morgan Kaufman, 2017.
- [21] P. Zong, J. Jiang, and J. Qin, “Study of high-dimensional data analysis based on clustering algorithm,” *15th Int. Conf. Comput. Sci. Educ. ICCSE 2020*, no. Iccse, pp. 638–641, 2020, doi: 10.1109/ICCSE49874.2020.9201656.
- [22] B. Subramanian, A. Paul, J. Kim, and K. W. A. Chee, “Metrics Space and Norm: Taxonomy to Distance Metrics,” *Sci. Program.*, vol. 2022, 2022, doi: 10.1155/2022/1911345.