

# Optimasi *Cluster* Pada Data *Stunting*: Teknik Evaluasi *Cluster Sum of Square Error* dan *Davies Bouldin Index*

Deny Jollyta<sup>1</sup>, Syahril Efendi<sup>2</sup>, Muhammad Zarlis<sup>2</sup>, Herman Mawengkang<sup>3</sup>

<sup>1</sup>Program Doktorat Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara

<sup>2</sup>Fakultas Ilmu Komputer dan Teknologi Informasi, Universitas Sumatera Utara

<sup>3</sup>Fakultas Matematika, Universitas Sumatera Utara

Jl. Universitas No. 9A Kampus USU, Medan

deny.jollyta@students.usu.ac.id

**Abstract** -The clusters number optimization problem is a problem that still requires continuous research so that the information produced can be a consideration. Cluster evaluation techniques with Sum of Square Error (SSE) and Davies Bouldin Index (DBI) are techniques that can evaluate the number of clusters from a data test. Research with these two techniques utilizes Stunting data from a number of regions in Indonesia. The result is information on stunting data which is formed from the optimal number of clusters where the largest SSE is formed at  $k = 5$  and the smallest DBI is formed at  $k = 5$ , with values of 23.403 and 1,178 respectively. Changes in the number of clusters also influence the information produced and DBI is proven to produce optimal number of clusters that contain information with a better pattern because it has a small intra-cluster value. It is expected that the results of this study can show the performance of the two evaluation techniques in producing the optimal number of clusters so that grouping information is in accordance with the expected pattern.

**Keywords:** Cluster, Optimization, SSE, DBI

**Abstrak**-Permasalahan optimasi jumlah cluster merupakan masalah yang masih membutuhkan penelitian berkelanjutan agar informasi yang dihasilkan dapat menjadi pertimbangan. Teknik evaluasi cluster dengan Sum of Square Error (SSE) dan Davies Bouldin Index (DBI) merupakan teknik yang dapat melakukan evaluasi terhadap jumlah cluster dari sebuah pengujian data. Penelitian dengan kedua teknik ini memanfaatkan data Stunting dari sejumlah wilayah di Indonesia. Hasilnya merupakan informasi data stunting yang terbentuk dari jumlah cluster optimal dimana SSE terbesar dibentuk pada  $k=5$  dan DBI terkecil dibentuk pada  $k=5$  dengan nilai masing-masing adalah 23.403 dan 1.178. Perubahan jumlah cluster turut mempengaruhi informasi yang dihasilkan dan DBI terbukti menghasilkan jumlah cluster optimal yang berisikan informasi dengan pola yang lebih baik karena memiliki nilai intra cluster kecil. Diharapkan hasil penelitian ini dapat memperlihatkan kinerja kedua teknik evaluasi dalam menghasilkan jumlah cluster optimal sehingga informasi pengelompokan sesuai dengan pola yang diharapkan.

**Kata kunci:** Cluster, Optimasi, SSE, DBI

## 1. PENDAHULUAN

*Mining* adalah penambangan yang mencirikan proses dalam menemukan informasi berharga dalam *database* yang sangat besar [1]. Berbagai metode dalam data *mining* digunakan untuk menganalisa data demi mendapatkan informasi berharga. *Clustering* merupakan salah satu cara untuk menganalisa data melalui

pengelompokan objek data yang mirip satu sama lain dalam *cluster* yang sama dan berbeda dengan objek di *cluster* lain [1]. Salah satu metode *clustering* yang terkenal adalah *K-Means*. Permasalahan yang sering terjadi pada *clustering* yang menggunakan *K-Means* adalah belum adanya ketentuan jumlah *cluster* yang tepat dari sejumlah  $k$  yang diujikan pada data. Tidak ada ukuran terbaik untuk pengelompokan data [1]. Hal ini menyebabkan informasi dari pengelompokan terkadang tidak sesuai dengan pola yang ada atau yang diharapkan. Saat ini telah banyak metode atau algoritma yang dihasilkan untuk menentukan jumlah *cluster* terbaik. Berbagai metode seperti *Elbow*, *Dynamic Cluster Algorithm*, *Hamming Distance* dan *Xie Beni Index*, merupakan cara yang ditemukan untuk menghasilkan jumlah *cluster* yang optimal. Metode-metode tersebut tidak berdiri sendiri dalam menentukan jumlah *cluster* yang tepat. Terdapat pula teknik untuk mengevaluasi *cluster* yang dihasilkan berdasarkan *error* tertinggi dan *index* terkecil yang dihasilkan, seperti *Sum of Square Error (SSE)* dan *Davies Bouldin Index (DBI)*. Kedua teknik ini sangat membantu dalam memastikan jumlah *cluster* yang paling optimal dari sejumlah  $k$  yang diujikan pada data.

Pada penelitian [2], *SSE* menghasilkan jarak terbesar dari sejumlah pengujian  $k$  sesuai dengan konsep metode *Elbow* yang diterapkan pada pengelompokan konsumen terbaik menggunakan *K-Means*. Dalam [3], *SSE* melakukan optimasi jumlah *cluster* pada klasterisasi energi WSN. Pengujian dilakukan menggunakan aplikasi *Matlab* dengan beberapa variasi parameter *performance* yang ada pada *Matlab* untuk memperoleh jumlah *cluster* yang optimal. Selain itu evaluasi *cluster* menggunakan *SSE* dan metode *Elbow*, juga dikolaborasikan dengan *Adjusted Rand Index (ARI)* dan *Silhouette Coefficient* untuk menemukan *cluster* paling optimal melalui pengujian yang dilakukan dengan metode *Ek-NNclus* [4]. *DBI* merupakan teknik evaluasi *cluster* yang digunakan pada segmentasi nasabah potensial hasil pengujian algoritma *K-Means* [5]. *DBI* diperlukan untuk menyempurnakan kinerja sebuah metode atau algoritma optimasi demi mendapatkan jumlah *cluster* yang terbaik. Pada penelitian [6], *DBI* digunakan untuk seleksi atribut berdasarkan koefisien variansi dan korelasi pada pemetaan e-government tahun 2016. *DBI* dikombinasikan dengan algoritma *erisoglu*, *calis*, dan *sakallioglu* untuk meningkatkan kinerja dalam menghasilkan *cluster* terbaik. *DBI* juga berhasil menentukan jumlah *cluster* optimal dalam melacak target dengan *celerity variable* yang tinggi [7]. Pada penelitian yang dibantu dengan aplikasi *RapidMiner* ini bertujuan untuk menghasilkan jumlah *cluster* yang optimal melalui komparasi performa teknik evaluasi *cluster SSE* dan *DBI*. Performa *SSE* dan *DBI* juga diuraikan dalam bentuk informasi yang dihasilkan dari setiap *cluster* dari  $k$  yang paling baik sehingga diharapkan informasi tersebut dapat menjadi acuan kebijakan karena sudah terpola dan sesuai dengan yang diharapkan.

## 2. MATERIAL DAN METODOLOGI PENELITIAN

### 2.1 Metode *Elbow* dan *Sum of Square Error (SSE)*

Metode *Elbow* bekerja membandingkan nilai atau persentase dari sejumlah  $k$  yang telah diujikan dan membentuk siku pada suatu titik [3]. Nilai  $k$  pada kombinasi siku dengan *K-Means* adalah grafik hubungan *cluster* dengan penurunan *error*. Peningkatan nilai  $k$  menyebabkan grafik menurun perlahan sampai hasil

nilai  $k$  stabil [2]. Jumlah *cluster*  $k$  yang dihasilkan dari pengujian dengan *K-Means*, dievaluasi menggunakan teknik *SSE*. Semakin banyak jumlah  $k$  yang diujikan, maka nilai *SSE* semakin kecil [8]. *SSE* merupakan cara dalam melakukan validasi *cluster* melalui jumlah kuadrat setiap anggota *cluster* menuju pusatnya [9]. Semakin jauh jarak yang membentuk titik siku, maka jumlah *cluster* tersebut menjadi yang paling optimal. Rumus *SSE* adalah sebagai berikut [10]:

$$SSE = \sum_{K=1}^K \sum_{x_i \in S_k} \|x_i - c_k\|_2^2 \quad (1)$$

## 2.2 Davies Bouldin Index (DBI)

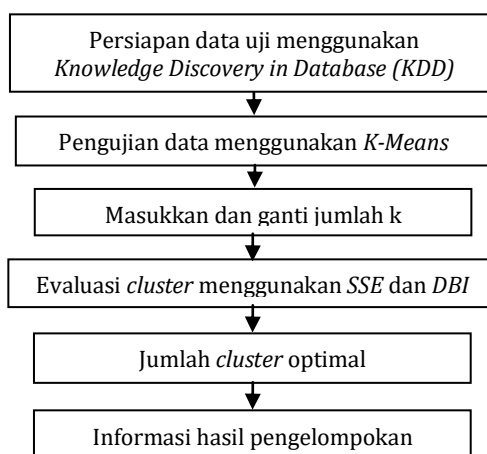
Metode *DBI* ini pertama kali diusulkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Evaluasi menggunakan *DBI* ini memiliki skema evaluasi *cluster* internal, dimana baik atau tidaknya hasil *cluster* dilihat dari kuantitas dan kedekatan antar data hasil *cluster*. Dalam penelitian [5], dijelaskan bahwa *DBI* adalah fungsi rasio dari jumlah antara *cluster scatter* sampai dengan *cluster separation*. Pendekatan pengukuran *DBI* yaitu memaksimalkan jarak *inter cluster* serta meminimalkan jarak *intra cluster*. Semakin kecil nilai *DBI* menunjukkan skema *cluster* yang paling optimal [5]. Rumus *DBI* adalah [11]:

$$R_{j,k} = \frac{MAE_j + MAE_k}{d(c_j, c_k)} \quad (2)$$

$$DBI = \frac{1}{M} \sum_{j=1}^M \max_{j \neq k} R_{j,k} \quad (3)$$

## 2.3 Kerangka Kerja Penelitian

Untuk mendapatkan hasil pengelompokan dengan nilai  $k$  yang optimal, dilakukan dengan beberapa tahapan seperti yang ditampilkan pada Gambar 1.



Gambar 1. Kerangka Kerja Penelitian

Penelitian diawali dengan mempersiapkan data pengujian yakni data *Stunting*. Persiapan data mengikuti langkah yang terdapat dalam *Knowledge*

*Discovery in Database (KDD)*, dimulai dengan seleksi data, pemilihan data dan transformasi data, hingga data siap untuk diujikan.

Atribut-atribut tersebut dipilih untuk mendapatkan informasi berdasarkan pengelompokan yang terbentuk setelah proses pengujian selesai dilakukan. Data diujikan dengan algoritma *K-Means* dengan memasukkan sejumlah  $k$ , mulai dari  $k=3$  hingga  $k=10$  dengan jumlah data 150. Hasil pengujian berupa nilai dari beberapa jumlah  $k$  yang selanjutnya dievaluasi menggunakan teknik evaluasi *cluster SSE* dan *DBI*. Hasil evaluasi *SSE* dan *DBI* merupakan jumlah *cluster* yang paling optimal dari pengelompokan yang terbentuk.

## 2.4 Data Uji

*Stunting* merupakan masalah gizi kronis pada balita yang ditandai dengan tinggi badan yang lebih pendek dari anak seusianya [12]. Istilah *Stunting* sangat lekat dengan sebutan kerdil bagi anak yang tumbuh dengan kurang sempurna. Kondisi *Stunting* disebabkan oleh beberapa atribut yakni kondisi calon ibu, usia, jarak kehamilan, postur calon ibu, kondisi ekonomi, kondisi sosial, gizi bayi dan kesakitan pada bayi [12]. Data disusun berdasarkan atribut-atribut tersebut untuk dikelompokkan. Informasi terkait kondisi bayi atau anak yang *stunting* dapat terbentuk sesuai kondisi yang ada agar informasi yang diperoleh dapat dipertanggungjawabkan. Deskripsi data yang diujikan dapat dilihat pada Tabel 1.

**Tabel 1.** Deskripsi Atribut Data *Stunting* Wilayah Indonesia

No	Atribut	Deskripsi Atribut
1	Kondisi calon ibu	Terdiri dari kecukupan gizi calon ibu, postur calon ibu, jarak kehamilan, usia
2	Usia	Terdiri dari usia terlalu tua, terlalu muda dan cukup umur
3	Jarak kehamilan	Terdiri dari calon ibu yang setiap tahun melahirkan dan calon ibu yang melahirkan dengan jarak antar kelahiran 2 tahun atau lebih
4	Postur calon ibu	Terdiri dari tinggi badan calon ibu yang ideal dan pendek
5	Kondisi ekonomi	Terdiri dari keadaan tidak mampu, cukup dan ideal
6	Kondisi social	Terdiri dari keadaan sanitasi dan keamanan pangan
7	Gizi Bayi	Terdiri dari inisiasi menyusui dini (IMD), gagal ASI eksklusif, penyapihan dini dan makanan pendamping ASI (MP-ASI)
8	Kesakitan pada bayi	Terdiri dari gangguan kognitif, motorik dan gangguan verbal

Sumber: (Buletin *Stunting*, 2018)

Kedelapan atribut data pada Tabel 1 merupakan deskripsi dari data yang dikelompokkan menggunakan algoritma *K-Means* dan dipersiapkan melalui tahapan *KDD* untuk selanjutnya dievaluasi jumlah *cluster* yang terbentuk dengan *SSE* dan *DBI*.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Hasil Pengelompokan

Data *Stunting* diujikan melalui algoritma *K-Means* dengan jumlah *cluster* uji mulai dari  $k = 3$  hingga  $k = 10$  menggunakan aplikasi *RapidMiner*. Hasil pengujian pertama merupakan hasil uji yang menggunakan evaluasi *cluster SSE* dengan parameter *similarity* dan *performance* dalam aplikasi *RapidMiner*. Jumlah anggota tiap pengelompokan ditampilkan pada Tabel 2 berikut.

**Tabel 2.** Pengelompokan dengan Evaluasi *SSE*

Jumlah k	Jarak antar Cluster	Jumlah Anggota Cluster
3	88.963	cluster_2 (64), cluster_1 (34), cluster_0 (52)
4	59.995	cluster_2 (50), cluster_3 (23), cluster_1 (31), cluster_0 (46)
5	54.43	cluster_0 (50), cluster_3 (24), cluster_2 (25), cluster_4 (45), cluster_1 (6)
6	39.178	cluster_5 (39), cluster_0 (22), cluster_3 (25), cluster_2 (39), cluster_1 (19), cluster_4 (6)
7	36.978	cluster_0 (37), cluster_2 (23), cluster_1 (23), cluster_6 (20), cluster_5 (39), cluster_4 (6), cluster_3 (2)
8	29.907	cluster_7 (15), cluster_6 (22), cluster_1 (23), cluster_0 (30), cluster_2 (17), cluster_4 (6), cluster_5 (35), cluster_3 (2)
9	29.261	cluster_0 (15), cluster_4 (21), cluster_1 (23), cluster_8 (19), cluster_7 (39), cluster_3 (4), cluster_6 (24), cluster_2 (2), cluster_5 (3)
10	22.771	cluster_7 (20), cluster_2 (21), cluster_0 (23), cluster_4 (17), cluster_5 (28), cluster_6 (3), cluster_9 (12), cluster_3 (2), cluster_8 (20), cluster_1 (4)

Jarak antar anggota *cluster* dari setiap pengujian  $k$  membentuk jumlah anggota dalam setiap kelompok yang terbentuk. Semakin banyak jumlah  $k$  yang diujikan, maka jarak antar *cluster* semakin kecil.

Selanjutnya hasil pengujian ditampilkan untuk pengujian yang dievaluasi dengan *DBI*. Parameter *RapidMiner* yang digunakan untuk menghasilkan pengelompokan ini adalah *cluster distance performance*. Jumlah anggota tiap pengelompokan ditampilkan pada Tabel 3 berikut.

**Tabel 3.** Pengelompokan dengan Evaluasi *DBI*

Jumlah k	Jarak antar Cluster	Jumlah Anggota Cluster
3	1.622	cluster_1 (61), cluster_0 (50), cluster_2 (39)
4	1.336	cluster_2 (49), cluster_1 (24), cluster_0 (31), cluster_3 (46)
5	1.206	cluster_1 (50), cluster_4 (23), cluster_3 (25), cluster_2 (42), cluster_0 (10)
6	1.071	cluster_4 (50), cluster_0 (22), cluster_3 (25), cluster_5 (19), cluster_1 (28), cluster_2 (6)
7	1.003	cluster_6 (30), cluster_0 (22), cluster_5 (25), cluster_2 (12), cluster_3 (11), cluster_4 (44), cluster_1 (6)
8	0.925	cluster_3 (15), cluster_6 (21), cluster_1 (25), cluster_5 (28), cluster_2 (18), cluster_7 (6), cluster_4 (18), cluster_0 (19)
9	0.909	cluster_4 (12), cluster_1 (15), cluster_7 (13), cluster_8 (19), cluster_2 (28), cluster_5 (17), cluster_0 (26), cluster_6 (16), cluster_3 (4)
10	0.836	cluster_0 (18), cluster_2 (22), cluster_8 (8), cluster_6 (10), cluster_3

Jumlah k	Jarak antar Cluster	Jumlah Anggota Cluster
		(11), cluster_5 (12), cluster_9 (6), cluster_4 (17), cluster_7 (17), cluster_1 (29)

Nilai jarak yang terbentuk dari hasil pengelompokan menggunakan parameter *DBI* pada Tabel 3 jauh lebih kecil, baik dari jarak antar *cluster* maupun jarak antar anggota *cluster*. Hasil perhitungan jarak membentuk keanggotaan yang hampir merata di setiap *cluster*. Hal ini menunjukkan bahwa kemiripan setiap anggota *cluster* sangat kuat sehingga membentuk informasi yang sesuai dengan keadaan yang ada.

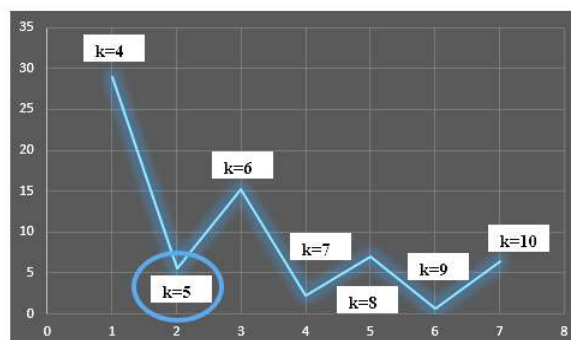
### 3.2 Evaluasi Cluster SSE dan DBI untuk Penentuan Cluster Optimal

Penelitian ini menggunakan 8 kali perubahan jumlah *cluster*. Untuk menentukan jumlah *cluster* yang paling optimal, dilakukan evaluasi *cluster* dengan *Sum of Square Error (SSE)* dan *Davies Bouldin Index (DBI)*. Setiap pengujian menghasilkan nilai jarak yang saling dibandingkan sesuai ketentuan. Adapun nilai kedua teknik evaluasi ditampilkan oleh Tabel 4.

**Tabel 4.** Nilai Evaluasi Cluster SSE dan DBI

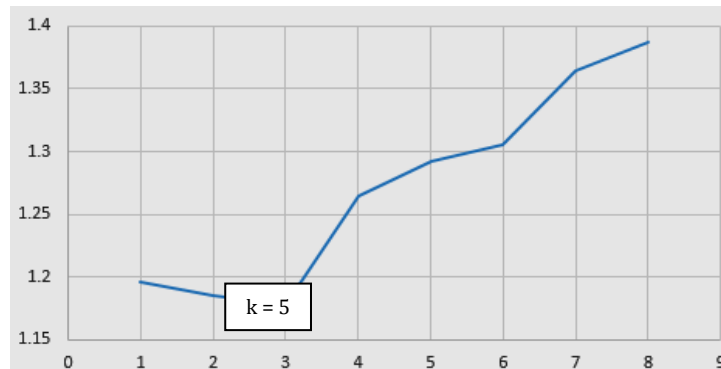
Jumlah k	Nilai SSE	Nilai DBI
k = 3	Inisiasi	1.196
k = 4	28.968	1.185
<b>k = 5</b>	5.565	1.178
k = 6	15.252	1.265
k = 7	2.2	1.292
k = 8	7.071	1.305
k = 9	0.646	1.365
k = 10	6.49	1.387

Nilai *SSE* didasarkan pada besarnya *error* yang dihasilkan untuk dapat membentuk siku dan menjadi jarak terbesar. Pada Tabel 4, jarak terbesar terdapat pada k = 5. Grafik hasil perhitungan *SSE* dapat dilihat pada Gambar 2.



**Gambar 2.** Cluster Optimal Teknik SSE

Sedangkan untuk teknik *DBI*, semakin kecil nilai indek yang terbentuk, semakin optimal *cluster* tersebut. *DBI* terkecil juga terdapat pada k = 5. Grafik *DBI* dapat ditampilkan seperti pada Gambar 3.



**Gambar 3.** Cluster Optimal Teknik DBI

Pengujian data *Stunting* menghasilkan jumlah *cluster* optimal yang sama melalui teknik evaluasi *SSE* dan *DBI*. Pada  $k = 5$ , perhitungan *SSE* dan *DBI* sampai pada perhitungan *cluster* yang paling baik. Namun demikian, walau *cluster* optimal sama-sama berada pada  $k = 5$ , informasi yang terbentuk berbeda. Hal ini terlihat dari keanggotaan tiap *cluster* yang tidak sama, yang artinya adalah informasi atau pengetahuan yang dapat dijadikan acuan juga menjadi tidak sama. Tabel 5 berikut merupakan informasi dari masing-masing evaluasi *cluster SSE* dan *DBI* pada  $k = 5$ .

**Tabel 5.** Informasi Hasil Pengelompokan *K-Means* Pada  $k=5$

Jumlah k ( <i>SSE</i> )	Informasi	Jumlah k ( <i>DBI</i> )	Informasi
0	<i>Cluster</i> ini berisikan 50 anggota yang menunjukkan bahwa bayi <i>stunting</i> lebih disebabkan oleh kondisi ekonomi yang parah. Akibatnya calon ibu kurang gizi, pertumbuhan tidak ideal dan tidak mampu menghasilkan ASI	1	Merupakan kelompok bayi <i>stunting</i> yang berjumlah 50 dari calon ibu ekonomi bawah, kurang gizi, berpostur pendek dan lingkungan yang kurang sanitasi. Akibatnya calon ibu tidak mampu memproduksi ASI sehingga bayi juga ada yang mengalami gangguan motorik, kognitif dan verbal
3	Anggota <i>cluster</i> yang berjumlah 24 anggota ini memberikan informasi bahwa bayi <i>stunting</i> berasal dari calon ibu ekonomi menengah namun gagal ASI. Calon ibu juga sudah terlalu tua untuk melahirkan dan memiliki riwayat kehamilan yang berdekatan sehingga selain <i>stunting</i> , bayi juga mengalami gangguan motorik	4	<i>Cluster</i> ini memiliki 23 bayi <i>stunting</i> yang berasal dari keluarga mampu. Namun calon ibu memiliki postur yang pendek dan gagal ASI. Akibatnya bayi mengalami gangguan motorik dan kognitif yang semakin menghambat pertumbuhannya
2	Informasi dari kelompok dengan 25 anggota ini adalah bahwa gangguan motorik pada bayi <i>stunting</i> juga dapat dialami oleh calon ibu yang	3	Ke-25 bayi dalam <i>cluster</i> ini sebagian besar mengalami gangguan motorik. Walaupun calon ibu berasal dari keluar menengah ke atas, namun postur

Jumlah k (SSE)	Informasi	Jumlah k (DBI)	Informasi
	berpostur pendek dengan riwayat ekonomi dan sosial yang cukup		tubuh calon ibu yang pendek turut menentukan pembentukan bayi <i>stunting</i>
4	45 anggota <i>cluster</i> ini adalah bayi <i>stunting</i> yang dilahirkan dari ibu belum cukup umur, berekonomi-sosial menengah ke bawah, dan gagal ASI. Selain <i>stunting</i> , 42% bayi pada <i>cluster</i> ini mengalami gangguan kognitif, 47% gangguan motorik dan sisanya mengalami gangguan verbal	2	<i>Cluster</i> ini berisikan 42 bayi <i>stunting</i> yang dilahirkan oleh calon ibu dengan usia masih sangat muda, kondisi ekonomi sosial yang sangat tidak memadai, postur tubuh pendek dan riwayat kehamilan yang rapat. Kekurangan gizi dan gagal ASI mengakibatkan bayi mengalami gangguan motorik, kognitif dan verbal
1	<i>Cluster</i> yang beranggotakan 6 bayi <i>stunting</i> ini dilahirkan oleh ibu yang masih terlalu muda namun mempunyai riwayat kehamilan yang sering. Akibatnya bayi juga mengalami gangguan kognitif yang perlu mendapat perhatian.	0	Terdapat 10 bayi <i>stunting</i> yang dilahirkan oleh calon ibu dengan usia sangat muda dan telah sering melahirkan. Walau dapat memberikan ASI, namun kualitasnya sangat menurun sehingga bayi mengalami gangguan motorik, kognitif dan verbal

Berdasarkan Tabel 5, terdapat kesamaan jumlah anggota *cluster* pada  $k = 0$  untuk SSE dan  $k = 1$  untuk DBI. Namun informasi yang dihasilkan berbeda. Perbedaan terletak pada atribut yang memiliki jarak lebih dekat pada pengujian dengan DBI. Setiap atribut saling terkait erat membentuk informasi dengan lengkap. Kedekatan jarak *intra cluster* ini menjadikan kinerja DBI lebih baik, tidak hanya dalam penentuan jumlah *cluster* yang optimal, namun juga dalam menghasilkan informasi yang lebih rinci dan terpola.

#### 4. KESIMPULAN

Teknik evaluasi *cluster SSE* dan *DBI* mampu menentukan jumlah *cluster* yang paling optimal dari serangkaian  $k$  yang diujikan pada data uji. Berdasarkan hasil pengujian, *SSE* dan *DBI* membentuk jumlah *cluster* optimal pada  $k=5$ . Namun performa *DBI* lebih baik dalam meminimalkan jarak *intra cluster* sehingga informasi yang terbentuk lebih lengkap dan terpola. Penelitian ini berbeda karena komparasi disertai dengan analisa hasil setiap  $k$  yang diujikan. Komparasi performa kedua teknik masih dapat dikembangkan melalui sejumlah data uji yang dibedakan dari segi jumlah maupun penggunaan pengukuran jarak dalam algoritma *K-Means*. Hasil penelitian ini diharapkan dapat memberikan pandangan baru dalam penggunaan teknik evaluasi *cluster* sehingga mempermudah pengguna untuk menentukan teknik yang sesuai dengan data yang ada dalam menghasilkan jumlah *cluster* yang paling optimal dan informasi yang benar.

#### DAFTAR PUSTAKA

- [1] Sugiyamta, "Sistem Deteksi Kemiripan Dokumen Dengan Algoritma Cosine Similarity dan Single Pass Clustering," *Din. Inform.*, vol. 7, no. 2, pp. 85-91, 2015.
- [2] M. A. Syakur, B. K. Khotimah, E. M. S. Rochman, and B. D. Satoto, "Integration K-Means



- Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster,” in *IOP Conference Series: Materials Science and Engineering*, 2017, vol. 336, no. 1, p. 7.
- [3] P. Bholowalia and A. Kumar, “EBK-Means: A Clustering Technique based on Elbow Method and K-Means in WSN,” *Int. J. Comput. Appl.*, vol. 105, no. 9, pp. 17–24, 2014.
- [4] Y. Zhang, T. Bouadi, and A. Martin, “An empirical study to determine the optimal k in Ek-NNclus method,” in *5th International Conference on Beleaf Functions (BELIEF2018)*, 2018, vol. 9, pp. 260–268.
- [5] Widiarina, “Algoritma Cluster Dinamik Untuk Optimasi Cluster Pada Algoritma K-Means Dalam Pemetaan Nasabah Potensial,” *J. Intell. Syst.*, vol. 1, no. 1, pp. 33–36, 2015.
- [6] I. D. Apriliyaningsih and D. Istiawan, “Penerapan Seleksi Atribut Berdasarkan Koefisien Variansi dan Korelasi untuk Inisialisasi Pusat Awal Klaster pada Algoritma K- Means dalam Pemetaan E-Government Tahun 2016,” in *Jurnal University Research Colloquium 2017 Universitas Muhammadiyah Magelang*, 2017, no. September, pp. 245–250.
- [7] M. Toumi, A. Maizate, M. Ouzzif, and M. S. Salah, “Dynamic clustering algorithm for tracking target with high and variable celerity (ATHVC),” *J. Comput. Networks Commun.*, no. 10, pp. 475–480, 2016.
- [8] N. P. E. Merliana, Ernawati, and A. J. Santoso, “Analisa Penentuan Jumlah Cluster Terbaik Pada Metode K-Means Clustering,” in *Prosiding Seminar Nasional Multi Disiplin Ilmu&Call for Papers Unisbank (Sendi\_U)*, 2015, pp. 978–979.
- [9] H. P. Duvvada, G. D. R. Naidu, and V. D. Sri, “K-Means Cluster Analysis Of Cities Based On Their,” *Int. J. Eng. Dev. Res.*, vol. 5, no. 4, pp. 1356–1363, 2017.
- [10] Irwanto, Y. Purwananto, and R. Soelaiman, “Optimasi Kinerja Algoritma Klasterisasi K-Means untuk Kuantisasi Warna Citra,” *J. Tek. ITS*, vol. 1, no. 1, pp. A197–A202, 2012.
- [11] P. Fränti, “Introduction,” in *Cluster Validation*, 2017.
- [12] Kementerian Kesehatan RI, “Situasi Balita Pendek (stunting) di Indonesia,” *Kementerian Kesehatan RI*, vol. 1, pp. 1–43, 2018.