

# Perbandingan Euclidean dan Manhattan Untuk Optimasi Cluster Menggunakan *Davies Bouldin Index*: Status Covid-19 Wilayah Riau

Wen Gie<sup>\*1</sup>, Deny Jollyta<sup>2</sup>

<sup>1,2</sup>Fakultas Ilmu Komputer, Institut Bisnis dan Teknologi Pelita Indonesia  
Jl. Jendral Ahmad Yani No. 78-88, Pekanbaru  
deny.jollyta@lecturer.pelitaindonesia.ac.id

## Abstrak

Optimasi jumlah cluster diperlukan untuk memastikan kebijakan yang dapat diambil terkait hasil pengelompokan, termasuk memastikan kelompok wilayah dengan status ODP, PDP dan Positif Covid-19 di provinsi Riau. Pengelompokan berdasarkan status pasien perlu dilakukan untuk menentukan tindakan pencegahan yang mungkin dapat diambil pemerintah. *Davies Bouldin Index* (DBI) merupakan teknik evaluasi cluster yang dapat digunakan pada algoritma clustering dengan pengukuran jarak Euclidean dan Manhattan. Penelitian ini dimaksudkan untuk mengetahui kinerja terbaik DBI pada kedua pengukuran jarak tersebut melalui pengujian data sebaran Covid-19 wilayah Riau. Hasil penelitian menunjukkan bahwa DBI terendah terdapat pada  $k=8$  untuk Euclidean dan  $k=7$  untuk Manhattan dengan nilai masing-masing sebesar 0,394 dan 0,434. Selain itu, DBI bekerja lebih baik pada Euclidean dibandingkan Manhattan karena memiliki nilai DBI lebih rendah pada semua  $k$  uji.

**Kata kunci:** Cluster optimal, Covid-19, DBI, Euclidean, Manhattan

## 1. Pendahuluan

*Davies Bouldin Index* (DBI) diperlukan untuk menyempurnakan kinerja sebuah metode atau algoritma optimasi demi mendapatkan jumlah *cluster* yang terbaik. Sebelum proses DBI, objek yang diuji telah diukur kemiripan menggunakan pengukuran jarak, seperti *Euclidean* dan *Manhattan*. Kedua pengukuran ini merupakan pengukuran yang umum dipakai pada algoritma *k-means* untuk menghasilkan pengelompokan. Terdapat beberapa penelitian yang memperlihatkan hasil kerja DBI melalui perhitungan jarak *Euclidean* dan *Manhattan*, seperti [1] yang mengoptimasi nasabah potensial hasil pengujian algoritma *K-Mean*, mengoptimasi kelompok hasil tangkapan ikan di kepulauan Ternate [2], menghasilkan cluster terbaik melalui perbandingan dengan Sum of Square Error (SSE) [3] dan melacak target dengan *celerity variable* yang tinggi [4]. Selain itu [5] dan [6] berhasil memperoleh *cluster* optimal dari hasil pengelompokan kecamatan dengan indikator pendidikan dan hasil pengelompokan calon penerima Jamkesda. Untuk penerapan *Manhattan* terdapat pada penelitian [7] yang menghitung jarak terpendek dari bentuk matrik menggunakan algoritma  $A^*$ . *Manhattan* juga mampu menghitung kemiripan 32 bahasa dunia [8]. Berdasarkan penelitian yang dipaparkan, belum terdapat penelitian yang membandingkan hasil *cluster* optimal dari dua pengukuran jarak yang telah disebutkan sebelumnya. Penelitian ini bertujuan untuk mendapatkan kinerja terbaik DBI melalui perbandingan *cluster* optimal dari perhitungan *Euclidean* dan *Manhattan* menggunakan data *Covid-19* provinsi Riau. Diketahui bahwa Riau termasuk salah satu wilayah dengan sebaran *Covid-19* cukup tinggi dan masuk dalam zona merah. Informasi ini disampaikan oleh Gubernur Riau yang dikutip dari *Bisnis.com* pada tanggal 14 April 2020. Provinsi Riau terdiri dari 12 kabupaten/kota dengan sejumlah kecamatan.

Hampir semua wilayah telah terjangkit wabah *Covid-19* sehingga perlu dievaluasi untuk merancang tindakan pencegahan yang mungkin dapat dilakukan. Data sebaran digunakan sebagai data uji dengan sejumlah kriteria seperti orang dalam pemantauan

(ODP), pasien dalam pengawasan (PDP) dan positif agar ditemukan kesamaan dalam memetakan wilayah sebaran. DBI digunakan untuk memperoleh cluster optimal dengan k uji mulai dari k=2 hingga k=8. Optimasi cluster yang dihasilkan dari perhitungan jarak *Euclidean* dan *Manhattan* diharapkan dapat memberikan informasi tentang kinerja kedua teknik ukur dalam membentuk *cluster* terbaik dengan DBI. Kinerja DBI juga teruji untuk bekerja pada pengukuran jarak sehingga diperoleh *cluster* yang optimal.

## 2. Material dan metodologi penelitian

### 2.1. Davies Bouldin Index (DBI)

Prinsip pendekatan pengukuran DBI adalah memaksimalkan jarak *inter cluster* serta meminimalkan jarak *intra cluster*. Semakin kecil nilai DBI menunjukkan skema *cluster* yang paling optimal [9]. Formula DBI dijabarkan sebagai berikut [10]:

$$R_{j,k} = \frac{MAE_j + MAE_k}{d(c_j, c_k)} \quad (1)$$

$$DBI = \frac{1}{M} \sum_{j=1}^M \max_{j \neq k} R_{j,k} \quad (2)$$

### 2.2. Pengukuran Jarak

*Euclidean Distance* merupakan perhitungan jarak antar titik [11] dan memiliki formula sebagai berikut:

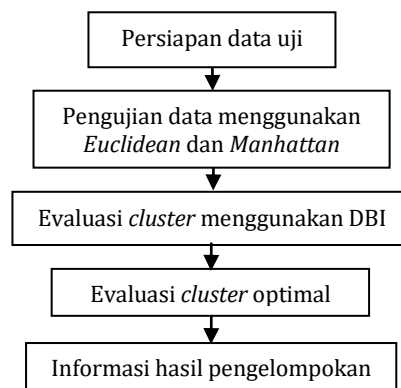
$$d_E = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_n - y_n)^2} \quad (3)$$

Pengukuran jarak *Manhattan* didasarkan pada selisih dua objek yang dijumlahkan dan hasil yang diperoleh dalam bentuk absolut [12]. Bentuk formula adalah sebagai berikut:

$$d(i, j) = |x_{i,1} - x_{j,1}| + |x_{i,2} - x_{j,2}| + \dots + |x_{i,n} - x_{j,n}| \quad (4)$$

### 2.3. Kerangka Penelitian

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



**Gambar 1. Kerangka Penelitian**

Langkah awal penelitian adalah melakukan persiapan data uji yang diambil dari situs resmi *Covid-19* Provinsi Riau. Selanjutnya kemiripan data dihitung dengan rumus jarak *Euclidean* dan *Manhattan*. Pengelompokan yang dihasilkan dari pengujian k=2 hingga k=8, dievaluasi menggunakan DBI.

### 3. Hasil dan pembahasan

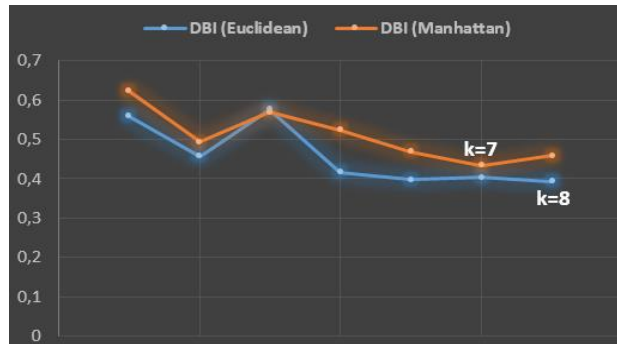
#### 3.1. Hasil Pengelompokan *Euclidean* dan *Manhattan* dengan *Cluster Optimal*

Berdasarkan hasil pengolahan 97 data status *Covid-19* di provinsi Riau, *cluster* optimal diperoleh dari perhitungan *Euclidean* dan DBI dengan menggunakan formula (3), (1) dan (2). Untuk *cluster* optimal dari perhitungan *Manhattan* digunakan formula (4), (1), (2). Hasil perhitungan membentuk pengelompokan sebagai berikut:

**Tabel 1. Pengelompokan *Euclidean-Manhattan* dan DBI**

K Uji	<i>Euclidean</i> (E)	DBI (E)	<i>Manhattan</i> (M)	DBI (M)
K=2	C1(88), C2(9)	0,558	C1(86), C2(11)	0,622
K=3	C1(74), C2(19), C3(4)	0,457	C1(74), C2(19), C3(4)	0,494
K=4	C1(42), C2(36), C3(15), C4(4)	0,576	C1(41), C2(37), C3(15), C4(4)	0,569
K=5	C1(41), C2(36), C3(16), C4(3), C5(1)	0,417	C1(36), C2(38), C3(14), C4(5), C5(4)	0,524
K=6	C1(36), C2(38), C3(13), C4(63), C5(3), C6(1)	0,397	C1(36), C2(38), C3(17), C4(7), C5(3), C6(1)	0,469
K=7	C1(20), C2(30), C3(24), C4(12), C5(7), C6(3), C7(1)	0,403	C1(20), C2(30), C3(24), C4(14), C5(5), C6(3), C7(1)	<b>0,434</b>
K=8	C1(20), C2(30), C3(24), C4(8), C5(6), C6(5), C7(3), C8(1)	<b>0,394</b>	C1(20), C2(30), C3(23), C4(9), C5(6), C6(5), C7(3), C8(1)	0,458

Hasil pengelompokan pada Tabel 1 menunjukkan bahwa *cluster* optimal setiap pengukuran terletak pada k uji yang berbeda. Berdasarkan k uji dari k=2 hingga k=8, k=8 dan k=7 terpilih sebagai *cluster* optimal untuk *Euclidean* dan *Manhattan*. K=8 dan k=7 memiliki DBI terendah dibandingkan k uji lainnya dengan nilai masing-masing sebesar 0,394 dan 0,434 sehingga membentuk pola sebagai berikut:



**Gambar 2. Pola *Cluster* Optimal**

Pola nilai DBI pengukuran jarak *Manhattan* pada Gambar 2 lebih tinggi dari pada nilai DBI pengukuran jarak *Euclidean* untuk setiap k uji. Ini membuktikan bahwa DBI mampu bekerja lebih baik pada pengukuran jarak *Euclidean* dibandingkan *Manhattan*.

#### 3.2. Evaluasi Anggota *Cluster* Hasil *Cluster* Optimal

*Cluster* optimal yang dihasilkan pada setiap pengukuran jarak tentu memberikan informasi yang berbeda. Berdasarkan hasil pengujian kedua pengukuran jarak, terdapat persamaan anggota *cluster* pada k=3. Selain itu, persamaan hanya pada beberapa *cluster* dalam setiap k uji. Berikut dijabarkan informasi yang diperoleh dari sebagian kelompok dari setiap *cluster* optimal.

**Tabel 2. Cluster Optimal Euclidean pada C=6**

Kabupaten/kota	Kecamatan	ODP	PDP	Positif
PEKANBARU	Tampan	960	124	18
ROKAN HULU	Tambusai Utara	974	5	0
ROKAN HILIR	Bangko Pusako	999	3	0
	Balai Jaya	950	2	0
	Bagan Sinembah	982	5	0

Tabel 2 memperlihatkan kelompok kabupaten/kota dengan kecamatan yang memiliki status ODP, PDP dan Positif tertinggi, yakni Pekanbaru, Rokan Hulu dan Rokan Hilir. C6 merupakan bagian *cluster* k=8 dari pengukuran jarak *Euclidean*. Status PDP dan Positif tertinggi berasal dari kota Pekanbaru.

**Tabel 3. Cluster Optimal Manhattan pada C=4**

Kabupaten/kota	Kecamatan	ODP	PDP	Positif
BENGKALIS	Mandau	612	12	2
SIAK	Tualang	692	22	0
	Dayun	812	6	1
	Sungai Apit	562	1	0
PELALAWAN	Pangkalan Kerinci	787	38	5
ROKAN HULU	Kunto Darussalam	554	7	0
	Rambah	743	13	1
	Kabun	597	2	0
	Rokan IV Koto	570	1	0
	Tandun	604	1	0
	Rambah Hilir	705	2	0
ROKAN HILIR	Rambah Samo	689	2	0
	Kubu	640	2	0
	Pasir Limau Kapas	555	1	0

Lima kabupaten/kota masuk dalam kelompok dengan status ODP akibat *Covid-19* di bawah angka 900 seperti yang diperlihatkan Tabel 3. Informasi ini diperoleh dari *cluster* optimal dengan pengukuran jarak *Manhattan* pada C4. Anggota *cluster* C4 juga menjadi sebaran *Covid-19* dengan jumlah PDP tertinggi di 5 kabupaten/kota. Jika dilihat dari posisi geografis, kelima kabupaten/kota ini saling berbatasan satu sama lain. Artinya, potensi penularan sangat besar karena jangkauan wilayah yang cukup dekat.

#### 4. Kesimpulan

Pengujian data sebaran *Covid-19* pada pengukuran jarak *Euclidean* dan *Manhattan* memiliki jumlah anggota *cluster* yang sama hanya pada k=3, namun dengan nilai DBI yang berbeda. *Euclidean* memiliki nilai DBI terendah pada k=8 dengan nilai 0,394. Nilai DBI terendah dari perhitungan *Manhattan* terdapat pada k=7 dengan nilai 0,434. Hasil *clustering* memperlihatkan bahwa DBI bekerja lebih baik pada *Euclidean*. Hal tersebut dimungkinkan karena formula *Euclidean* menggunakan akar kuadrat sehingga menghasilkan nilai DBI yang lebih kecil dibandingkan *Manhattan* yang perhitungannya menggunakan nilai mutlak. Kontribusi ini menjadi pembeda dengan penelitian [3] yang membandingkan hasil DBI dengan teknik evaluasi cluster lain, yakni SSE. Penentuan *cluster* yang optimal perlu dilakukan mengingat informasi tentang status *Covid-19* di masyarakat dari setiap k uji penting. Adanya *cluster* optimal memudahkan penentuan prioritas wilayah yang ditangani dari semua sisi. Pengukuran jarak *Euclidean* dan *Manhattan* dapat membantu kinerja DBI menjadi lebih mudah dalam menghasilkan cluster yang terbaik pada data *Covid-19* di provinsi Riau.

**Daftar pustaka**

- [1] 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.
- [2] R. Hablum, A. Khairan, and R. Rosihan, "Clustering Hasil Tangkap Ikan Di Pelabuhan Perikanan Nusantara (Ppn) Ternate Menggunakan Algoritma K-Means," *JIKO (Jurnal Inform. dan Komputer)*, vol. 2, no. 1, pp. 26–33, 2019, doi: 10.33387/jiko.v2i1.1053.
- [3] D. Jollyta, S. Efendi, M. Zarlis, and H. Mawengkang, "Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum of Square Error dan Davies Bouldin Index," in *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2019, vol. 1, no. September, pp. 918–926, doi: 10.30645/senaris.v1i0.100.
- [4] 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.*, vol. 2016, pp. 1–10, 2016, doi: 10.1109/MASCOTS.2016.10.
- [5] F. Irhamni, F. Damayanti, B. Khusnul, and A. Miftachul, "Optimalisasi pengelompokan kecamatan berdasarkan indikator pendidikan menggunakan metode clustering dan davies bouldin index," in *Seminar Nasional dan Teknologi UMJ*, 2014, no. 11, pp. 1–6.
- [6] M. N. V. Waworuntu and M. F. Amin, "Penerapan Metode K-Means Untuk Pemetaan Calon Penerima Jamkesda," *Klik - Kumpul. J. Ilmu Komput.*, vol. 5, no. 2, pp. 190–200, 2018, doi: 10.20527/klik.v5i2.157.
- [7] S. K. Sharma and S. Kumar, "Comparative Analysis of Manhattan and Euclidean Distance Metrics Using A \* Algorithm," *J. Res. Eng. Appl. Sci.*, vol. 1, no. 4, pp. 196–198, 2016.
- [8] T. Strauss and M. J. Von Maltitz, "Generalising ward's method for use with manhattan distances," *PLoS One*, vol. 12, no. 1, pp. 1–21, 2017, doi: 10.1371/journal.pone.0168288.
- [9] Sugiyamta, "Sistem Deteksi Kemiripan Dokumen Dengan Algoritma Cosine Similarity dan Single Pass Clustering," *Din. Inform.*, vol. 7, no. 2, pp. 85–91, 2015.
- [10] P. Fränti, "Introduction," in *Cluster Validation*, 2017, p. 110.
- [11] K. M. Cubukcu and T. Hatcher, "Are Euclidean Distance and Network Distance Related?," in *7th Asia-Pacific International Conference on Environment-Behaviour Studies*, 2016, vol. 1, no. 4, pp. 219–227, doi: 10.21834/e-bpj.v1i4.137.
- [12] D. Sinwar and R. Kaushik, "Study of Euclidean and Manhattan Distance Metrics using Simple K-Means Clustering," *Int. J. Res. Appl. Sci. Eng. Technol.*, vol. 2, no. V, pp. 270–274, 2014.