

OPTIMALISASI PENGELOMPOKAN DATA TINGKAT HUNIAN HOTEL DENGAN ALGORITMA K-MEDOID

Dewa Ayu Indah Cahya Dewi¹⁾ Dewa Ayu Kadek Pramita²⁾

Program Studi Teknik Otomasi, Jurusan Teknik Elektro, Politeknik Negeri Bali¹⁾

Program Studi Sistem Komputer, Institut Bisnis dan Teknologi Indonesia, Denpasar, Bali ²⁾
ayuindahcahyadewi@pnb.ac.id¹⁾, pramita.wayu@instiki.ac.id²⁾

ABSTRACT

Hotel occupancy rates have decreased significantly during the Covid-19 pandemic. For the sake of increasing the economy in the hotel industry sector, an analysis of hotel occupancy data grouping is needed so that it can provide the right problem solving to increase hotel occupancy rates through promotions and other vouchers based on group characteristics. This research is in the form of grouping data on hotel occupancy rates in Indonesia using the K-Medoid clustering method and the Davies Bouldin Index (DBI). The selection of the K-Medoid clustering method is based on the K-medoid method which is not affected by noise and data outliers. Grouping the data with 9 clusters resulted in a better cluster quality seen from the minimum DBI value obtained of 1,003. The results of clustering are determined by the characteristics of the data in the cluster with a small dispersion/dispersion size because the standard deviation value is smaller than the mean value. The data used for clustering has a small variation. The cluster group with the lowest occupancy rate is located in the center of the cluster, namely January with an occupancy rate of 11.86%, February 21.66%, March 31.16%, April 17.12%, May 33.73% and June 26.88%. Areas classified as low occupancy groups are Aceh, West Sulawesi and Bali based on data from January 2022 to June 2022. These areas need to increase the number of hotel occupancy with various promotions.

Keywords: K-medoid, Davies-Bouldin Index, Occupancy, Hotel.

ABSTRAK

Tingkat hunian hotel mengalami penurunan signifikan selama pandemi Covid-19. Demi peningkatan perekonomian pada sektor industri perhotelan, diperlukan suatu analisis pengelompokan data tingkat hunian hotel sehingga dapat memberikan pemecahan masalah yang tepat untuk meningkatkan tingkat hunian hotel melalui promosi dan voucher lainnya berdasarkan karakteristik kelompok. Penelitian ini berupa pengelompokan data tingkat hunian hotel di Indonesia menggunakan metode clustering K-Medoid dan Davies Bouldin Index (DBI). Pemilihan metode clustering K-Medoid didasarkan pada metode K-medoid tidak terpengaruh *noise* dan pencilaan data (*outlier*). Pengelompokan data dengan 9 cluster menghasilkan kualitas cluster yang lebih baik dilihat dari nilai DBI minimum yang diperoleh sebesar 1.003. Hasil clustering ditentukan oleh karakteristik data yang di clustering dengan ukuran penyebaran/dispersi yang kecil karena nilai standar deviasi yang lebih kecil dari nilai mean. Data yang digunakan untuk clustering memiliki variasi yang kecil. Kelompok cluster yang memiliki tingkat hunian paling rendah terletak pada pusat cluster yaitu bulan januari dengan tingkat hunian 11.86%, Februari 21.66%, Maret 31.16%, april 17.12%, mei 33.73% dan juni 26.88%. Daerah yang tergolong kelompok hunian rendah yaitu Aceh, Sulawesi Barat dan Bali berdasarkan data dari Januari 2022 hingga Juni 2022 Daerah-daerah tersebut perlu dilakukan peningkatan jumlah hunian hotel dengan berbagai promosi.

Kata Kunci : K-medoid, Davies-Bouldin Index, Hunian, Hotel.

PENDAHULUAN

Penyebaran Covid-19 di Indonesia tahun 2020 mengakibatkan Pemerintah membuat kebijakan untuk menekan penyebaran Covid-19 dengan *lockdown* di beberapa wilayah Indonesia. Penerapan kebijakan tersebut melalui penerapan belajar dari rumah (BDR), penutupan tempat wisata/hiburan dan perbelanjaan serta penerapan kebijakan *work from home* dan pembatasan lainnya untuk mengurangi penyebaran Covid-19. Kebijakan pembatasan tersebut berdampak signifikan pada sektor industri perhotelan. Tingkat hunian hotel mengalami penurunan signifikan selama pandemi Covid-19. Demi peningkatan perekonomian pada sektor industri perhotelan, diperlukan suatu analisis pengelompokan data tingkat hunian hotel sehingga dapat memberikan pemecahan masalah yang tepat untuk meningkatkan tingkat hunian hotel melalui promosi dan voucher lainnya berdasarkan karakteristik kelompok. Penelitian ini berupa pengelompokan data tingkat hunian hotel di Indonesia menggunakan metode clustering K-Medoid dan Davies Bouldin Index (DBI). Pemilihan metode clustering K-Medoid didasarkan pada metode K-medoid tidak terpengaruh *noise* dan pencilaan data (*outlier*), selain itu metode K-Medoid tidak bergantung pada urutan masuk dataset (Sarle, Kaufman, & Rousseeuw, 1991). Metode validitas cluster yang digunakan untuk mengetahui kualitas hasil cluster, yaitu silhouette index, davies-bouldin index, dunn index dan lainnya. Penentuan jumlah cluster terbaik dapat mengoptimalkan hasil clustering pada suatu data. Kualitas hasil cluster terbaik ditunjukkan dengan nilai DBI yang semakin rendah (Davies & Bouldin, 1979).

Beberapa penelitian terdahulu yang menjadi dasar penelitian penulis adalah sebagai berikut.

Penelitian yang berjudul “Clustering Untuk Menentukan Strategi Promosi Universitas Muhammadiyah Jember Dengan Algoritma K-Medoids” tahun 2022 yang dilakukan oleh Radiyanto Dekaprasetya membahas pengelompokan data mahasiswa untuk mendapatkan informasi profil mahasiswa sebagai pendukung keputusan untuk strategi promosi yang tepat. Hasil penelitian yaitu

mendapatkan nilai DBI yang rendah pada cluster $k = 3$ yang dijadikan sebagai cluster optimal (Dekaprasetya, Anisatur, & Muharom, 2022). Penelitian tahun 2019 yang dilakukan Frenda berjudul “Perbandingan Algoritma K-Means dan K-Medoids dalam Klasterisasi Produk Asuransi Perusahaan Nasional” membahas perbandingan metode K-Means dan K-Medoids pada produk asuransi perusahaan nasional. Hasil penelitian menunjukkan metode K-Means lebih optimal didasarkan nilai Davies Bouldin Index (DBI) terendah 0,018 pada jumlah cluster $k=5$ (Farahdinna, Nurdiansyah, Suryani, & Wibowo, 2019). Penelitian yang dilakukan Ansita tahun 2019 berjudul “Analisis Kesiapan Modernisasi Daerah Irigasi Kedung Putri pada Tingkat Sekunder Menggunakan Metode K-Medoids Clustering” membahas tentang klasterisasi saluran sekunder daerah irigasi Kedung Putri. Hasil penelitian menunjukkan bahwa jumlah cluster paling optimal adalah 4 cluster, dengan nilai Davies Bouldin Index (DBI) -1,959 (Pradipta, Pratyasta, & Arif, 2019).

TINJAUAN PUSTAKA

Algoritma K-Medoid

Algoritma K-Medoids merupakan algoritma yang digunakan untuk clustering menggunakan objek yang mewakili cluster yang biasa disebut medoids. Algoritma k-medoids menghitung kedekatan yang dimiliki antara medoid dengan objek non medoids serta meminimalisir ketidaksamaan objek- objek dalam cluster menggunakan nilai absolute error (E).

$$E = \sum_{c=1}^k \sum_{i=1}^{n_c} |p_{ic} - o_c| \quad (1)$$

Keterangan :

n_c = banyaknya objek dalam cluster ke-c

p_{ic} = objek non medoids i dalam cluster ke-c

O_c = nilai medoids di cluster ke-c

Metode Davies Bouldin Index

Davies Bouldin Index merupakan metode untuk mengetahui kualitas clustering dengan pendekatan nilai kohesi dan separasi. Kohesi adalah jarak objek data terhadap pusat clusternya. Separasi adalah jarak objek data terhadap pusat cluster lainnya. Cluster yang baik adalah cluster yang memiliki nilai kohesi yang rendah dan nilai separasi yang tinggi (Wirawan, Fahmi, & Suprpto, 2017). Nilai *Davies Bouldin Index (DBI)* yang semakin mendekati nilai 0 menandakan semakin baik cluster yang diperoleh. Semakin kecil nilai DBI menunjukkan hasil skema cluster yang optimal (Davies & Bouldin, 1979).

Sum of square within cluster (SSW) digunakan untuk mengetahui matrik kohesi dalam sebuah cluster ke- i yang dapat dilihat pada persamaan berikut.

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (2)$$

Keterangan :

m_i = jumlah data dalam cluster ke- i

c_i = centroid cluster ke- i

$d(x_j, c_i)$ = jarak euclidean setiap data terhadap pusat cluster/centroid

Sum of square between cluster (SSB) digunakan untuk mengetahui nilai separasi antar cluster yang dapat dilihat pada persamaan berikut.

$$SSB_{i,j} = d(c_i, c_j) \quad (3)$$

Keterangan :

$d(c_i, c_j)$ = jarak antar centroid

Melalui nilai kohesi dan separasi selanjutnya dilakukan pengukuran rasio (R_{ij}) untuk mengetahui nilai perbandingan antara cluster ke- i dan cluster ke- j , berikut ini adalah persamaannya.

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (4)$$

Persamaan untuk menghitung nilai Davies Bouldin Index (DBI) sebagai berikut.

$$DBI = \frac{1}{K} \sum_{i=1}^K \max_{i \neq j} (R_{i,j}) \quad (5)$$

Keterangan :

K = jumlah cluster yang digunakan

METODE PENELITIAN

Preprocessing Data

Tahap preprocessing data merupakan tahap persiapan data sebelum proses clustering, pada penelitian ini melakukan proses transformasi data dalam format number dan reduksi data missing value sehingga data yang didapatkan bersih dan bebas dari noise. Variabel data sekunder yang digunakan dari Badan Pusat Statistik berbentuk excel terdiri dari nama provinsi, tingkat hunian hotel bulan januari hingga bulan juni 2022. Data yang digunakan dalam penelitian menggunakan data dari Badan Pusat Statistik tahun 2022, berjumlah 34 buah provinsi di Indonesia. Berikut ini adalah variabel-variabel data dari Badan Pusat Statistik:

Tabel 1 Karakteristik data clustering

Provinsi	Januari	Februari	Maret	April	Mei	Juni
Aceh	19.09	29.70	37.72	22.77	33.71	37.33
Sumatera Utara	42.62	38.65	42.05	33.83	49.09	44.18
Sumatera Barat	45.19	45.84	46.37	25.13	57.85	51.94
Riau	40.84	42.13	46.52	34.64	44.59	45.58
Jambi	37.88	36.38	43.63	38.09	48.28	50.95
Sumatera Selatan	49.94	48.69	54.77	42.31	57.18	55.70
Bengkulu	35.43	32.82	39.01	24.32	41.59	39.67
Lampung	51.19	48.48	56.59	42.20	66.04	58.87
Kep. Bangka Belitung	26.11	25.90	34.12	23.80	39.32	37.97
Kep. Riau	26.39	30.00	35.72	27.21	37.11	45.21
Dki Jakarta	52.26	46.05	52.70	45.83	51.79	54.32

Jawa Barat	46.99	39.04	50.44	35.35	54.13	52.82
Jawa Tengah	39.38	37.59	41.90	32.13	49.49	46.74
Di Yogyakarta	59.90	45.00	51.65	27.62	67.65	66.45
Jawa Timur	44.71	46.56	51.68	38.70	57.46	54.25
Banten	41.64	41.84	51.22	41.91	52.17	57.00
Bali	20.71	14.86	21.90	18.98	37.35	38.77
Nusa Tenggara Barat	31.35	29.31	38.43	15.27	32.35	38.73
Nusa Tenggara Timur	29.85	28.46	40.80	29.64	41.46	48.32
Kalimantan Barat	45.27	46.24	50.07	41.98	51.48	51.97
Kalimantan Tengah	43.20	44.97	49.56	42.64	53.56	54.71
Kalimantan Selatan	41.08	42.59	46.85	34.85	52.72	53.05
Kalimantan Timur	60.78	58.90	63.82	51.30	62.54	64.60
Kalimantan Utara	50.43	48.94	54.91	35.81	44.47	49.47
Sulawesi Utara	40.11	37.71	39.24	35.06	43.08	41.52
Sulawesi Tengah	44.40	45.97	54.80	33.57	49.39	51.20
Sulawesi Selatan	38.90	40.27	51.16	34.84	46.98	53.58
Sulawesi Tenggara	38.29	41.04	40.07	27.36	32.85	41.29
Gorontalo	29.60	36.75	47.22	28.40	42.44	44.03
Sulawesi Barat	11.86	21.66	31.16	17.12	33.73	26.88
Maluku	40.36	34.44	35.83	25.82	31.54	32.34
Maluku Utara	38.87	47.72	43.28	36.75	50.77	50.27
Papua Barat	44.07	47.67	49.14	39.98	38.62	54.23
Papua	39.54	37.66	38.18	38.62	32.98	34.94
Standar Deviasi	10.55	8.90	8.40	8.35	9.73	8.78
Mean	39.65	39.41	45.07	33.05	46.64	47.91

Teknik Pengumpulan Data

Teknik pengumpulan data digunakan untuk mendapatkan data, fakta maupun informasi yang berguna dan terkait dengan penelitian. Teknik pengumpulan data yang penulis lakukan pada penelitian ini adalah sebagai berikut.

1. Studi Literatur merupakan pengumpulan data sekunder dan informasi melalui berbagai sumber pustaka meliputi artikel, jurnal, internet dan lain sebagainya.
2. Dokumen merupakan pengumpulan data sekunder melalui berbagai data yang telah dikumpulkan dan dimiliki oleh pihak

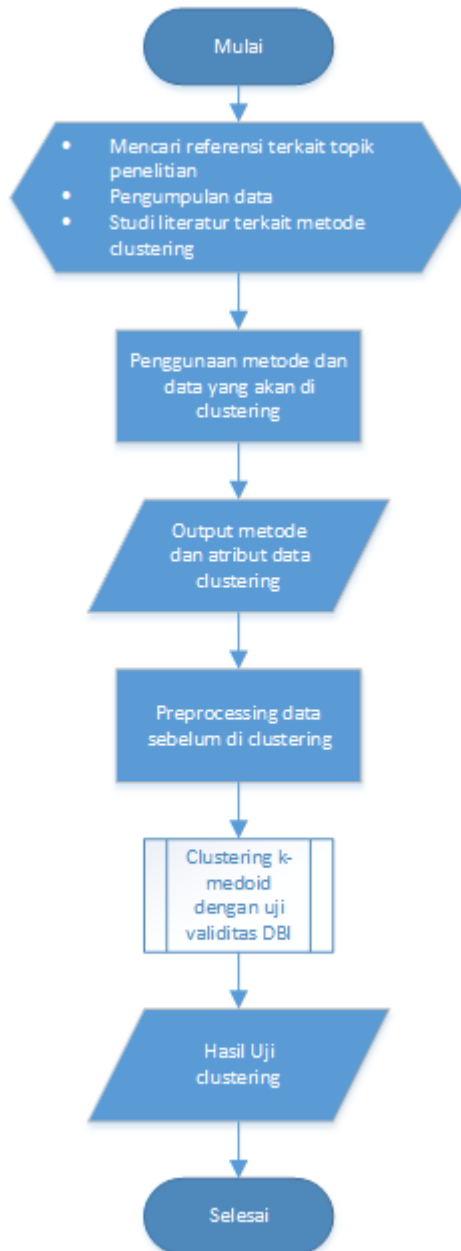
tertentu. Pada penelitian ini penulis menggunakan data tingkat hunian hotel

di Indonesia dari Badan Pusat Statistik yang berbentuk file excel.

3. Observasi dilakukan dengan melakukan pengamatan secara langsung. Pada penelitian ini observasi dilakukan pada data kuantitatif hasil pengelompokkan data dengan pengujian validitas cluster dengan metode DBI melalui Rapidminer.

Alur Penelitian

Tahapan penelitian adalah sebagai berikut.



Gambar 1 Alur Penelitian

Gambaran Umum Penelitian

Gambar 2 menjelaskan tentang gambaran umum dari clustering pada penelitian ini.



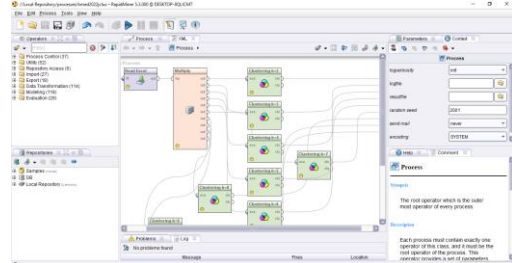
Gambar 2 Gambaran Umum Penelitian

Gambaran umum penelitian tahap pertama yang dilakukan adalah mengumpulkan data dari Badan Pusat Statistik, setelah data didapatkan dilanjutkan ke langkah kedua yaitu data preprocessing. Langkah ketiga melakukan uji kualitas clustering dengan metode DBI pada Rapidminer, lalu melakukan analisis hasil cluster sehingga didapatkan jumlah cluster terbaik pada proses clustering k-medoid.

HASIL DAN PEMBAHASAN

Pengelompokkan data dengan metode K-Medoid

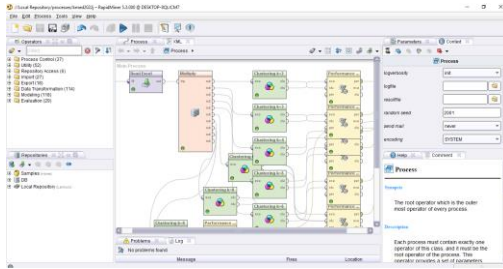
Pengelompokkan data dengan metode clustering k-medoid menggunakan aplikasi Rapid Miner dengan pengujian validitas cluster menggunakan nilai DBI. Pengelompokkan data dengan k-medoid dimulai dari pengelompokkan 2 cluster hingga 10 cluster.



Gambar 3 Tampilan Clustering

Pengujian Validitas Cluster

Pada penelitian ini metode yang digunakan untuk menguji hasil cluster adalah metode Davies Bouldin Index (DBI). Dengan menggunakan DBI suatu cluster akan dianggap memiliki hasil clustering yang optimal jika memiliki DBI minimal (Kamila, Khairunnisa, & Mustakim, 2019).



Gambar 4 Tampilan Uji Validitas Clustering

Berdasarkan clustering k-medoid menggunakan uji validitas cluster davies bouldin index pada masing-masing jumlah cluster didapatkan tabel sebagai berikut.

Tabel 2 Nilai DBI uji validitas cluster

Jumlah Cluster	Nilai DBI
2 cluster	1.588
3 cluster	1.165
4 cluster	1.517
5 cluster	1.085
6 cluster	1.093
7 cluster	1.2
8 cluster	1.168
9 cluster	1.003
10 cluster	1.011

Pada tabel 2 terlihat jumlah cluster yang memiliki nilai DBI minimum adalah jumlah cluster sebanyak 9 dengan nilai DBI 1.003, hal ini menandakan kualitas cluster yang paling optimal.

Cluster Model

```
Cluster 0: 2 items
Cluster 1: 5 items
Cluster 2: 8 items
Cluster 3: 1 items
Cluster 4: 3 items
Cluster 5: 1 items
Cluster 6: 4 items
Cluster 7: 6 items
Cluster 8: 4 items
Total number of items: 34
```

Gambar 5 Cluster model

Kelompok cluster yang memiliki tingkat hunian paling rendah terletak pada pusat cluster yaitu bulan januari dengan tingkat hunian 11.86%, Februari 21.66%, Maret 31.16%, april 17.12%, mei 33.73% dan juni 26.88%. Daerah yang tergolong kelompok hunian rendah yaitu Aceh, Sulawesi Barat dan Bali berdasarkan data dari Januari 2022 hingga Juni 2022.

SIMPULAN

Berdasarkan hasil pembahasan Optimalisasi Pengelompokkan Data Tingkat Hunian Hotel dengan Algoritma K-Medoid didapatkan simpulan sebagai berikut.

Pengelompokkan data dengan 9 cluster menghasilkan kualitas cluster yang lebih baik dilihat dari nilai DBI minimum yang diperoleh. Hasil clustering ditentukan oleh karakteristik data yang di clustering dengan ukuran penyebaran/dispersi yang kecil karena nilai standar deviasi yang lebih kecil dari nilai mean. Data yang digunakan untuk clustering memiliki variasi yang kecil. Daerah yang tergolong kelompok hunian rendah yaitu Aceh, Sulawesi Barat dan Bali. Daerah-daerah tersebut perlu dilakukan peningkatan jumlah hunian hotel dengan berbagai promosi.

DAFTAR PUSTAKA

1. Davies, D. L., & Bouldin, D. W. (1979). A Cluster Separation Measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, PAMI-1(2). <https://doi.org/10.1109/TPAMI.1979.4766909>
2. Dekaprasetya, R., Anisatur, U., & Muharom, L. A. (2022). Clustering Untuk Menentukan Strategi Promosi Universitas Muhammadiyah Jember Dengan Algoritma K-Medoids Clustering To Determine The University Of Muhammadiyah Jember Promotion Strategy With K-Medoids Algorithm. *Jurnal Smart Teknologi*, 3(3), 312–320.
3. Farahdinna, F., Nurdiansyah, I., Suryani, A., & Wibowo, A. (2019). PERBANDINGAN ALGORITMA K-MEANS DAN K-MEDOIDS

DALAM KLASTERISASI
PRODUK ASURANSI
PERUSAHAAN NASIONAL. *Jurnal
Ilmiah FIFO*, 11(2).
[https://doi.org/10.22441/fifo.2019.v1
1i2.010](https://doi.org/10.22441/fifo.2019.v11i2.010)

4. Kamila, I., Khairunnisa, U., & Mustakim. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119–125.
5. Pradipta, A. G., Pratyasta, A. S., & Arif, S. S. (2019). Analisis Kesiapan Modernisasi Daerah Irigasi Kedung Putri Pada Tingkat Sekunder Menggunakan Metode K-Medoids Clustering. *AgriTECH*, 39(1).
[https://doi.org/10.22146/agritech.410
06](https://doi.org/10.22146/agritech.41006)
6. Sarle, W. S., Kaufman, L., & Rousseeuw, P. J. (1991). Finding Groups in Data: An Introduction to Cluster Analysis. *Journal of the American Statistical Association*.
<https://doi.org/10.2307/2290430>
7. Wirawan, Fahmi, M. F., & Suprpto, Y. K. (2017). Segmentation and distribution of watershed using K-Modes clustering algorithm and davies-bouldin Index based on geographic information system (GIS). *Proceedings - 2016 International Seminar on Application of Technology for Information and Communication, ISEMANTIC 2016*.
[https://doi.org/10.1109/ISEMANTIC.
2016.7873844](https://doi.org/10.1109/ISEMANTIC.2016.7873844)