

Penerapan Metode *K-Medoids Cluster* Dalam Mengelompokkan Hotel di Jakarta Berdasarkan *Website tiket.com*

Athallah Widyatama Rafii, Raden Bagus Fajriya Hakim

Universitas Islam Indoneisa, Jalan Kaliurang KM 14.5, D.I. Yogyakarta, Indonesia 55584

*Corresponding author: athallah.rafi@students.uui.ac.id



E-ISSN: 2986-4178

Riwayat Artikel

Dikirim:

10 April 2023

Direvisi:

26 April 2023

Diterima:

27 Juni 2023

ABSTRAK

Jakarta merupakan ibu kota dan kota terbesar di Indonesia. Jakarta memiliki peran tersendiri dalam peningkatan sektor ekonomi di Indonesia. Oleh karena itu banyak perkantoran yang berada di Jakarta, selain itu dari sektor pariwisata juga menunjang perekonomian. Salah satunya hotel, juga menjadi akomodasi yang penting untuk wisatawan ataupun pebisnis. Dalam memilih hotel tentu calon konsumen memiliki kriteria tertentu seperti dari segi kelas, harga, bahkan dari penilaian orang (*review*) yang pernah singgah di hotel tersebut, calon konsumen dapat melihat kriteria ini melalui website perhotelan. Sehingga tujuan dari penelitian ini adalah melakukan pengelompokan hotel berdasarkan kelas (bintang), *review*, dan harga. Pengelompokan ini menggunakan metode *K-Medoids Cluster Analysis*. Dari hasil analisis *K-medoids* yang telah dilakukan, didapatkan 3 kelompok kluster. Klaster 1 dengan karakteristik hotel kelas bawah beranggotakan 47 hotel. Klaster 2 dengan karakteristik hotel kelas menengah beranggotakan 69 hotel. Dan untuk klaster 3 dengan karakteristik hotel kelas atas beranggotakan 14 hotel.

Kata Kunci: Jakarta, Hotel, *K-Medoids*, Cluster

ABSTRACT

*Jakarta is a capital city of Indonesia and the biggest city in Indonesia. Jakarta has a role in improving the economic sector in Indonesia. Therefore, there are many offices in Jakarta. Apart from that, the tourism sector also supports the economy. One of them is a hotel, a basic accommodation for tourists or business people. Of course, in choosing a hotel, potential customers have specific criteria, such as terms of class and price. Even from the reviews of people who have stopped at the hotel, potential customers can see these criteria through the hotel website. So the purpose of this research is to group hotels based on class (stars), reviews, and prices. This grouping uses the *K-Medoids Cluster Analysis* method. From the results of the *K-medoids* analysis that has been carried out, there are 3 cluster groups. Cluster 1 with lower-class hotel characteristics consists of 47 hotels. Cluster 2, with the characteristics of a middle-class hotel, consists of 69 hotels. And for cluster 3 with the characteristics of upscale hotels consisting of 14 hotels*

Keywords: Jakarta, Hotel, *K-Medoids*, Cluster

1. Pendahuluan

Daerah Khusus Ibukota Jakarta atau dapat disingkat menjadi DKI Jakarta adalah ibu kota negara Indonesia. Jakarta merupakan satu-satunya kota di Indonesia yang kedudukannya setingkat provinsi sebagaimana diatur dalam Undang-Undang No. 18 Tahun 1965 tentang Pokok-pokok Pemerintahan Daerah. Jakarta merupakan kota dengan mobilitas yang tinggi dan juga yang menyebabkan kondisi perekonomian di Jakarta terus meningkat. Selain menjadi pusat bisnis dan politik, Jakarta juga memiliki komoditas di sektor pariwisata yang cukup memadai diantaranya Kepulauan Seribu, Dufan, Taman Mini Indonesia Indah, dan masih banyak lagi. Disamping tempat wisata yang memadai, Jakarta juga menyediakan akomodasi berupa hotel, selain untuk tempat beristirahat bagi wisatawan, hotel juga digunakan sebagai tempat untuk para pebisnis yang hanya singgah beberapa hari di Jakarta atau untuk warga lokal yang hendak menghabiskan waktu liburannya menginap di hotel.

Sebagai calon konsumen, tentu mempertimbangkan dari segi kenyamanan, harga, maupun fasilitasnya. Hotel pun memiliki kelasnya tersendiri mulai dari bintang 1 hingga hotel dengan bintang 5, pemberian bintang ini dilakukan oleh Dinas Pariwisata Daerah atau instansi resmi lainnya. Dalam pemberian bintang pada hotel ini, terdapat kriteria tertentu untuk pemberiannya. Selain bintang pada hotel, biasanya para calon penghuni hotel mencari hasil *review* pada *website* atau aplikasi terkait sektor pariwisata seperti *tiket.com*, *Traveloka*, *Agoda* dan masih banyak lagi.

Penulis meninjau terhadap beberapa penelitian pada tahun-tahun sebelumnya yang berkaitan diantaranya yaitu membahas tentang pengelompokan dalam menentukan pemesanan hotel dengan menggunakan *clustering* metode *K-Medoids*. Pada penelitian ini menghasilkan 3 *cluster*, untuk *cluster* 1 terdiri dari 2.343 transaksi, *cluster* 2 terdiri dari 1.833 transaksi, dan untuk *cluster* 3 terdiri dari 4.530 transaksi penelitian ini disusun oleh Kurniawan dkk [2]. Terdapat pula tinjauan terkait dengan penerapan *clustering* metode *K-Medoids* yang membahas tentang pengelompokan data potensi kebakaran hutan atau lahan berdasarkan persebaran titik panas yang disusun oleh Pramesti dkk. Pada penelitian ini didapatkan 2 *cluster*. Pada *cluster* pertama dengan karakteristik potensi tinggi dengan hasil rata-rata *brightness* sebesar $344,47^{\circ}K$ dan untuk *cluster* kedua dengan karakteristik potensi sedang dengan hasil rata-rata *brightness* sebesar $318,8^{\circ}K$ [3].

Berdasarkan hasil tinjauan tersebut, penulis akan melakukan pengelompokan untuk melihat karakteristik terhadap data hotel yang ada pada *website tiket.com* di wilayah Jakarta dengan menggunakan salah satu metode *clustering* non hierarki yaitu *K-Medoids Cluster Analysis* dikarenakan metode ini memiliki kelebihan tersendiri, yaitu mampu mengatasi nilai pencilon pada data[1].

2. Metodologi Penelitian

2.1.Data dan Sumber Data

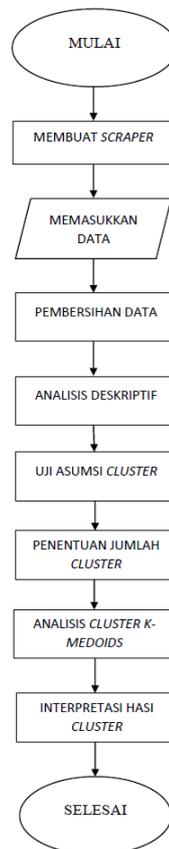
Data yang digunakan merupakan data primer yang diambil melalui *website tiket.com* pada produk hotel dengan menggunakan *scraper* yang telah penulis buat menggunakan bahasa pemrograman *Python* pada bulan Maret 2022. Variabel merupakan objek penelitian atau karakteristik yang menjadi perhatian sebuah penelitian. Maka dari itu, variabel yang digunakan penulis disajikan pada Tabel 1 sebagai berikut.

Tabel 1 Tabel Definisi Variabel

Nama Variabel	Definisi Variabel	Satuan
Bintang	Penilaian terhadap fasilitas dan pelayanan (bintang 1-5)	Tingkat atau Kelas
Review	Ulasan singkat berupa nilai 1-5.	Skala
Harga	Nilai yang harus dibayarkan konsumen	Rupiah

2.2. Metode Penelitian

Salah satu metode yang digunakan pada penelitian ini adalah statistika deskriptif untuk mengetahui gambaran umum pada data hotel dan untuk mengelompokkannya sendiri menggunakan analisis *cluster* dengan metode *K-Medoids*, penulis menggunakan *software Visual Code* untuk melakukan *scraping*, *Microsoft Excel* untuk pembersihan data dan juga *software R* untuk melakukan analisisnya. Untuk tahapan penelitian dijelaskan pada sebuah diagram alur seperti berikut:



Gambar 1. Diagram Alur Penelitian

3. Hasil dan Pembahasan

3.1. Analisis Deskriptif

Hasil dari analisis deskriptif ini digunakan sebagai gambaran umum pada data yang diperoleh dari hasil *scraping* pada *website tiket.com* [4]. Gambaran umum dari data yang digunakan Penulis tersaji dalam Tabel 2 sebagai berikut.

Tabel 2 Tabel Definisi Variabel

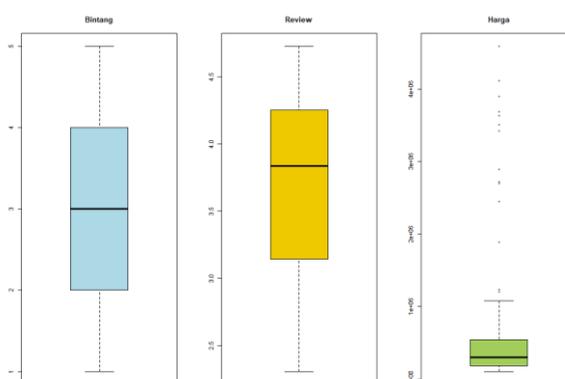
Nilai	Variabel	
	Review	Harga
Minimum	2,305	110.000
Kuartil 1	3,162	188.220
Median	3,836	306.391
Kuartil 3	4,252	540.595
Maksimum	4,727	4.585.900
Mean	3,705	664.005

Berdasarkan Tabel 2 diperoleh hasil statistik deskriptif dari data hotel, untuk variabel Ulasan nilai minimumnya adalah 2,305, kuartil 1 bernilai 3,162, median bernilai 3,836, kuartil 3 bernilai 4,252, nilai maksimumnya 4,727 dan rata-rata dari variabel Ulasan adalah 3,705. Dan untuk variabel Harga nilai minimumnya adalah 110.000, kuartil 1 bernilai 188.220, mediannya bernilai 306.391, kuartil 3 bernilai 540.595, nilai maksimumnya adalah 4.585.900 dan untuk rata-rata dari variabel Harga adalah 664.005.



Gambar 2. Diagram Batang Variabel Bintang

Berdasarkan diagram batang di atas dapat dilihat bahwa hotel dengan berbintang 1 jumlahnya adalah 11, untuk hotel berbintang 2 jumlahnya adalah 46, hotel berbintang 3 jumlahnya adalah 33, hotel berbintang 4 jumlahnya adalah 23 dan untuk hotel berbintang 5 jumlahnya adalah 17. Sehingga total sampel yang digunakan penulis adalah 130.



Gambar 3. Boxplot Variabel

Berdasarkan *boxplot* yang didapatkan pada Gambar 3 untuk variabel Bintang dan Review tidak terdapat nilai *outlier*, tetapi pada variabel harga terdapat beberapa data yang memiliki nilai *outlier* atau pencilan. Sehingga metode yang cocok digunakan pada analisis ini yaitu metode *K-Medoids Clustering* karena yang menjadi pusat *cluster* adalah objek representatifnya yang kemudian menempatkan setiap objek disekitarnya dengan medoids, sehingga cocok untuk data yang mengandung *outlier* [5].

3.2. Uji Asumsi Kluster

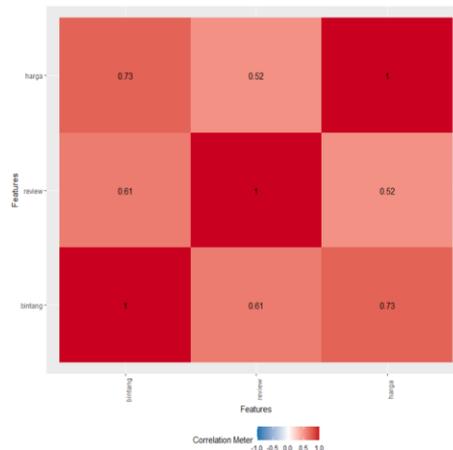
Dalam melakukan uji *cluster*, terdapat dua uji asumsi yang harus terpenuhi yang pertama yaitu uji kelayakan sampel yang harus representatif terhadap populasinya, dan kedua yaitu uji multikolinearitas untuk mengetahui ada atau tidaknya korelasi antar variabel independennya.

3.2.1. Uji Kelayakan Sampel

Untuk menguji kelayakan sampel representatif terhadap populasinya dilakukan uji kelayakan dengan menggunakan nilai *Kaiser Meyer Olkin* (KMO) [6]. *Output* nilai KMO yang diperoleh untuk data hotel adalah sebagai berikut. Diperoleh nilai *Kaiser Meyer Olkin* (KMO) yang diperoleh sebesar 0,6829297 yang artinya nilai KMO > 0,5, maka dari itu dapat disimpulkan bahwa sampel yang diambil merepresentasikan populasi yang ada dan sampel layak digunakan untuk analisis kluster.

3.2.2. Uji Multikolinearitas

Uji selanjutnya yang harus terpenuhi adalah uji multikolinearitas antar variabel, uji ini digunakan untuk mengetahui apakah terdapat korelasi atau hubungan yang kuat antar variabel bebasnya, berikut adalah hasil koefisien korelasi antar variabel yang digunakan.



Gambar 4. Plot Korelasi

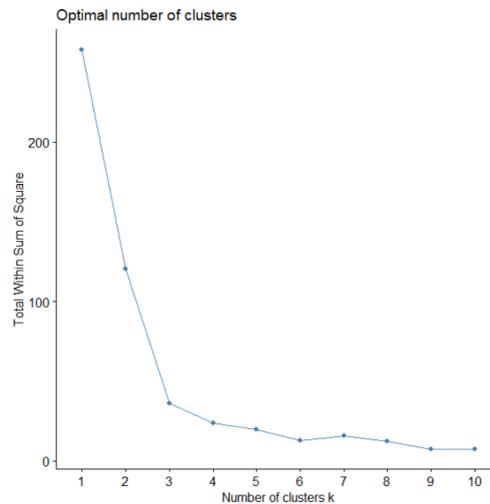
Kesamaan antar variabel dapat dilihat dari koefisien korelasi antar pasangan objek, berdasarkan hasil korelasi pada Gambar 4 menurut [7] terjadi multikolinearitas apabila nilai koefisien korelasinya lebih dari 0,85 sedangkan variabel yang digunakan Penulis tidak terdapat nilai koefisien korelasi yang lebih dari 0,85. Sehingga dapat disimpulkan bahwa variabel bebas yang digunakan layak digunakan karena tidak terjadi multikolinearitas.

3.3. Analisis Kluster Metode K-Medoids

Setelah memenuhi uji kelayakan sampel dan uji multikolinearitas, maka Penulis melanjutkan analisis *Cluster* menggunakan metode *K-Medoids*.

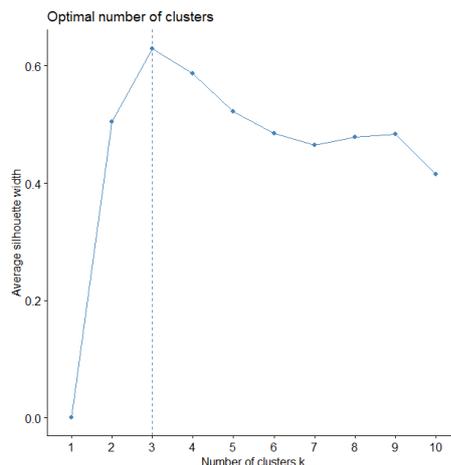
3.3.1. Penentuan Jumlah Kluster Optimal

Untuk mengetahui bagaimana pengelompokan atau hasil *clustering* terhadap harga dan *review* hotel di DKI Jakarta, langkah awalnya yaitu menentukan jumlah *cluster* optimalnya, pada penelitian kali ini penulis menggunakan metode *Sum of Square Error* atau biasa disebut dengan metode *Elbow* dan metode *Silhouette*.



Gambar 5. Grafik *Cluster Optimal Metode Elbow*

Berdasarkan plot pada Gambar 5. dapat dilihat bahwa garis mulai melandai pada *cluster* ke 3 atau nilai $k=3$ menurut [8] dalam penentuan jumlah *cluster* terbaiknya dapat dilihat dengan membandingkan jumlah *cluster* yang membentuk siku di suatu titik. Berbeda dengan nilai sebelumnya ($k=2$, $k=1$) yang menunjukkan kecuraman yang cukup tinggi karena semakin landai titik *cluster* semakin optimum jumlah *cluster*-nya. Selanjutnya penulis menggunakan metode *Silhouette* didapatkan hasil sebagai berikut.

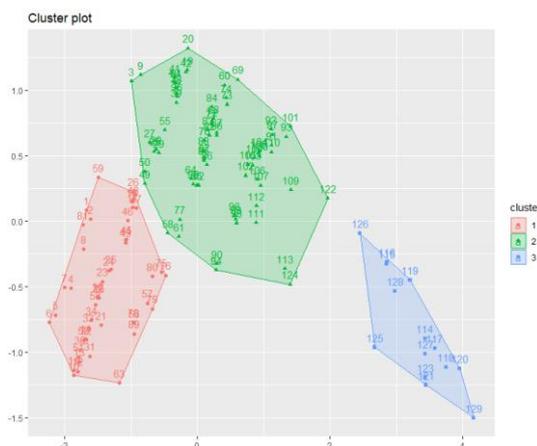


Gambar 6. Grafik *Cluster Optimal Metode Silhouette*

Berdasarkan grafik pada Gambar 6. dengan menggunakan metode *Silhouette* nilai *cluster* optimumnya juga menunjukkan di angka 3 atau $k=3$. Maka dari itu, setelah melakukan pendekatan dengan metode *Elbow* dan *Silhouette* dimana keduanya menunjukkan jumlah *cluster* 3 kelompok. Oleh karena itu, penulis menggunakan jumlah *cluster* optimalnya adalah 3 kelompok.

3.3.2 Penerapan Kluster Metode *K-Medoids*

Proses *cluster* dilakukan dengan membagi jumlah *cluster* sebanyak 3 kelompok sesuai dengan hasil penentuan jumlah *cluster*. Data yang telah memenuhi uji asumsi kemudian diolah menggunakan *softwawre Rstudio*. Berikut merupakan hasil dari pengelompokkan hotel berdasarkan variabel bintang, harga dan *review*.



Gambar 7. Plot *Cluster K-Medoids*

Dari visualisasi pada Gambar 7 hasil *cluster* dengan jumlah kelompoknya adalah 3 ($k=3$) dimana anggota *cluster 1* berwarna merah, anggota *cluster 2* berwarna hijau, dan anggota *cluster 3* berwarna biru. Berikut merupakan tabel anggota kelompok untuk masing-masing *cluster*.

Tabel 3 Tabel Anggota *Cluster K-Medoids*

<i>Cluster</i>	Jumlah Anggota
1	47
2	69
3	14

Berdasarkan tabel 4 didapatkan bahwa *cluster 1* memiliki 47 anggota, *cluster 2* memiliki 69 anggota, dan *cluster 3* memiliki 14 anggota.

Untuk mengetahui karakteristik dari tiap *cluster* yang telah didapatkan, penulis melakukan perhitungan rata-rata terhadap setiap variabel, maka didapatkan hasil dalam Tabel 5.

Tabel 4 Hasil Rata-Rata *Cluster K-Medoids*

<i>Cluster</i>	Review	Harga	Keterangan
1	2,89	193.372	Bawah
2	4,09	458.775	Menengah
3	4,56	3.255.478	Atas

Berdasarkan nilai rata-rata pada tiap *cluster* yang terdapat pada Tabel 5 dapat diketahui ciri-ciri dari setiap *cluster*, sehingga dapat diinterpretasikan sebagai berikut:

1. *Cluster 1* terdiri dari 47 hotel yang merupakan hotel kelas bawah. Hal itu terjadi karena anggotanya rata-rata terdiri dari hotel dengan 2 bintang, nilai *review* yang rendah dan harga dari hotel pada *cluster 1* lebih murah dibandingkan *cluster 2* dan *cluster 3*.
2. *Cluster 2* terdiri dari 69 hotel yang merupakan hotel kelas menengah. Hal itu dapat terjadi dikarenakan anggotanya rata-rata terdiri dari hotel dengan 3 bintang, nilai *review* yang relatif lebih tinggi dibandingkan *cluster 1*, namun harga lebih mahal dibandingkan *cluster 1* tetapi tidak semahal *cluster 3*.
3. *Cluster 3* terdiri dari 14 hotel yang merupakan hotel kelas atas. Hal itu dapat terjadi dikarenakan anggotanya rata-rata terdiri dari hotel dengan 5 bintang, nilai *review* yang relatif lebih tinggi dibandingkan *cluster 1* dan *cluster 2*. Dan memiliki harga yang paling tinggi dibandingkan *cluster 1* dan *cluster 2*.

4. Kesimpulan

Berdasarkan pembahasan sebelumnya penulis mendapatkan kesimpulan bahwa dari pengelompokan data hotel yang diambil dari tiket.com berdasarkan variabel bintang, *review*, dan harga dengan menggunakan analisis *cluster* metode *K-Medoids* diperoleh jumlah *cluster* sebanyak 3 kelompok dengan jumlah anggota *cluster* pertama sebanyak 47 hotel, *cluster* kedua sebanyak 69 hotel, dan *cluster* ketiga sebanyak 14 hotel.

Dari penelitian yang telah dilakukan, diperoleh karakteristik atau ciri khusus dari masing-masing *cluster* adalah sebagai berikut *Cluster* 1 merupakan pengelompokan hotel kelas bawah, dengan rata-rata anggotanya hotel dengan bintang 2 dengan nilai *review* yang rendah serta harga hotel yang paling murah dibandingkan *cluster* 2 dan *cluster* 3, *Cluster* 2 merupakan pengelompokan hotel kelas menengah, dimana rata-rata anggotanya adalah hotel dengan bintang 3 dengan nilai *review* yang lebih tinggi dibandingkan *cluster* 1 serta harga hotel lebih mahal dibandingkan *cluster* 1 tetapi tidak lebih mahal dibandingkan *cluster* 3, *Cluster* 3 merupakan pengelompokan hotel kelas atas, dengan seluruh anggotanya adalah hotel dengan bintang 5 dengan nilai *review* paling tinggi serta harga hotel yang paling tinggi dibandingkan *cluster* 1 dan *cluster* 2.

5. Daftar Pustaka

- [1] Eiyanto, M. N. Mara and N. Satyahadewi, "Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis," 2013.
- [2] W. Kurniawan, A. Rifai, W. Gata and D. Gunawan, "Analisis Algoritma K-Medoids Clustering Dalam Menentukan Pemesanan Hotel," *Swabum E-Journal*, 2020.
- [3] D. F. Pramesti, M. T. Furqon and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," 2017.
- [4] P. Agus, in *Panduan Laboratorium Statistik Inferensial*, Jakarta, Gramedia Widiasana Indonesia, 2007.
- [5] K. G. Soni and A. Patel, "Comparative Analysis of K-Means and K-Medoids Algorithm on IRIS Data," *International Journal of Computational Intelligence Research*, pp. 899-906, 2017.
- [6] R. A. Johnson and D. W. Wichern, *Applied Multivariate Statistical Analysis Sixth Edition*, New Jersey, 2007.
- [7] A. Widarjono, *Ekonometrika: Pengantar dan Aplikasinya*, Jakarta, 2013.
- [8] S. Madhulatha, "An Overview on Clustering Methods," *IOSR Journal of Engineering*, 2012.