

KLASTERISASI DATA MINING JENIS BOOKING PADA HOTEL THE HIVE

**Muhammad Raka Ayunda Wahab¹, Trisakti Tumbelaka², Al Fajri Mulya³
Universitas Bina Sarana Informatika**

E-mail: rakaayunda716@gmail.com¹, saktitumbelaka@gmail.com²,
alfajri.mulya0903@gmail.com³

Abstract

This study aims to support the management of The Hive Hotel in monitoring visitor data based on the types of bookings used during the period from January to December 2023. Using RapidMiner software and the K-Means clustering method, this research segments hotel guest data into several groups based on similarities in booking characteristics. The methods employed include observation, primary data collection from the hotel, and data processing using the K-Means clustering technique. The initial data consists of 12 variables with a total of 3,587 rows, which were then selected and cleaned to ensure optimal clustering results. The study's findings reveal that hotel guests can be divided into three main groups: low, medium, and high, with adequate accuracy based on metrics such as Euclidean Distance and the Davies Bouldin Index. Key findings from this research include the identification of guest booking patterns that help management enhance services, personalization, pricing, and capacity management more efficiently.

Keywords — Data Clustering; Hotel Data Analysis; The Hive Hotel; RapidMiner; Data Mining.

Abstrak

Penelitian ini bertujuan untuk mendukung manajemen hotel The Hive dalam memantau data pengunjung berdasarkan jenis pemesanan yang digunakan selama periode Januari hingga Desember 2023. Menggunakan perangkat lunak RapidMiner dan metode K-Means clustering, penelitian ini mengelompokkan data tamu hotel ke dalam beberapa segmen berdasarkan kesamaan karakteristik pemesanan. Metode yang digunakan mencakup observasi, pengumpulan data primer dari pihak hotel, serta pengolahan data menggunakan teknik K-Means clustering. Data awal terdiri dari 12 variabel dengan total 3587 baris data, yang kemudian dipilih dan dibersihkan untuk memastikan hasil clustering yang optimal. Hasil penelitian mengungkapkan bahwa tamu hotel dapat dibagi ke dalam tiga kelompok utama: rendah, sedang, dan tinggi, dengan tingkat akurasi yang memadai berdasarkan metrik seperti Euclidean Distance dan Davies Bouldin Index. Temuan utama dari penelitian ini mencakup identifikasi pola pemesanan tamu hotel yang membantu manajemen dalam meningkatkan layanan, personalisasi, penetapan harga, dan pengelolaan kapasitas secara lebih efisien.

Kata Kunci — Klasterisasi Data; Analisis Data Hotel; Hotel The Hive; RapidMiner; Data Mining.

1. PENDAHULUAN

Perkembangan teknologi dalam dunia bisnis kini telah menjadi elemen penting dalam mempermudah pengelolaan data, termasuk di sektor perhotelan. Industri perhotelan sendiri mengalami pertumbuhan yang sangat pesat seiring dengan munculnya banyak hotel baru, yang meningkatkan tingkat persaingan dalam sektor ini. Berdasarkan PP No.56/2011, hotel merupakan bangunan yang menyediakan layanan penginapan berbayar bagi pengunjung, dan untuk menghadapi kompetisi yang semakin ketat, pengelolaan data

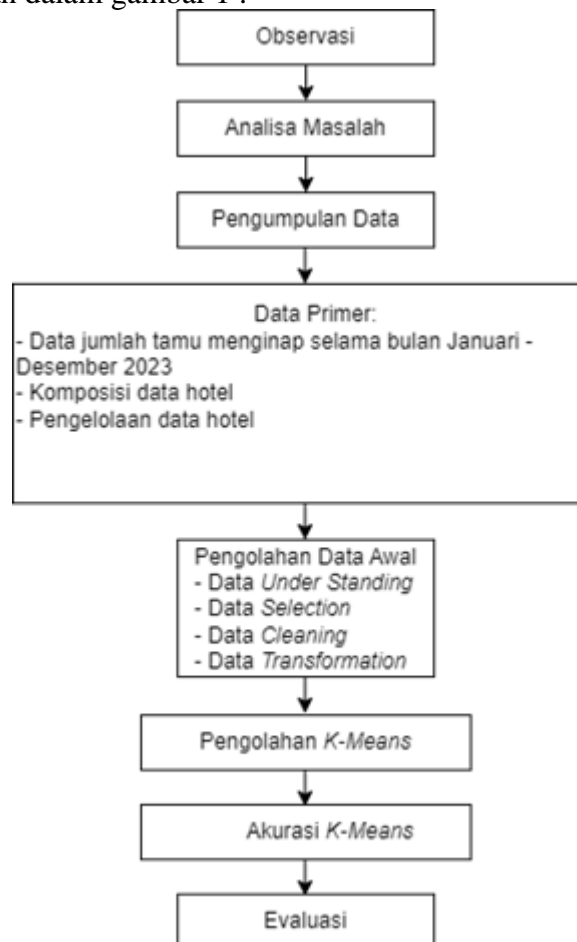
pengunjung secara efektif sangat diperlukan. Di Hotel The Hive, penerapan sistem pengelolaan data yang baik diharapkan mampu meningkatkan kualitas layanan yang diberikan kepada tamu hotel.

Salah satu pendekatan dalam pengelolaan data pengunjung yang dapat diterapkan adalah dengan melakukan pengelompokan data berdasarkan jenis booking yang dilakukan selama periode tertentu. Dengan demikian, pola-pola preferensi tamu dapat diidentifikasi, yang selanjutnya dapat dimanfaatkan untuk memaksimalkan layanan yang diberikan. Berbagai penelitian terdahulu telah membahas pengelompokan data menggunakan teknologi digital di berbagai industri, termasuk perhotelan, sebagai langkah untuk meningkatkan daya saing. Metode K-Means menjadi salah satu teknik yang umum digunakan untuk pengelompokan data yang kompleks secara efektif dan efisien.

Penelitian ini bertujuan untuk mengaplikasikan metode K-Means yang diimplementasikan melalui software RapidMiner guna mengelompokkan data pengunjung Hotel The Hive berdasarkan jenis booking yang dilakukan selama periode Januari hingga Desember 2023. Manajemen hotel dapat merancang strategi pelayanan yang lebih terfokus dan efisien dengan memahami preferensi tamu melalui pola booking yang telah dianalisis. Hal ini tidak hanya akan membantu dalam meningkatkan kualitas layanan, tetapi juga dapat mendorong efisiensi operasional yang sesuai dengan kebutuhan dan harapan para pengunjung.

2. METODE PENELITIAN

Proses penelitian ini diawali dengan studi literatur sampai penyusunan laporan seperti yang dijabarkan dalam gambar 1 :



Gambar 1. Tahapan Penelitian

a. Observasi

Observasi dilakukan dengan memperhatikan dan memahami situasi serta mencatat detail penting serta mengumpulkan data mengenai tamu yang menginap, dengan tujuan untuk memperoleh informasi yang diperlukan.

b. Analisis masalah

Analisis masalah dapat memberikan pemahaman yang lebih mendalam tentang potensi, batasan, dan strategi untuk meningkatkan kepuasan tamu menggunakan metode k-means clustering dalam konteks studi.

c. Penumpukan data

Data yang dihimpun merupakan data primer berupa informasi dari pihak hotel The Hive. Observasi lokasi dan pengumpulan data dilakukan dengan cara mencari jurnal tentang perhitungan klasifikasi atau klusterisasi data menggunakan K-Means, lalu mempelajari referensi menurut para ahli dan peraturan pemerintah atau kementerian terkait.

d. Pengolahan data awal

Pengolahan data dalam K-Means clustering mencakup beberapa tahap penting yang perlu diperhatikan untuk memastikan kualitas hasil clustering yang optimal. Berikut adalah penjelasan singkat untuk setiap tahap dalam pengolahan data awal (preprocessing) pada K-Means:

1) Data under standing

Data awal dari pihak hotel The Hive dengan bentuk file excel dengan jumlah variable sebanyak 12 kolom berupa nomor, nama tamu, pax, type, room, arrival (tanggal dan bulan), departemen, ngt, rate code, discon, net rate, dan company (jenis booking) sebanyak 3587 baris (pada bulan januari).

2) Data Selection

Pemilihan data dengan cara menghilangkan variable yang tidak diperlukan yaitu nomor, nama tamu, pax, type, room, departemen, ngt, rate code, discon, dan net rate, dan menyisakan variable arrival (tanggal dan bulan) dan company (jenis booking).

3) Data Cleaning

Menghapus data publikasi dan inkonsisten data, mengelompokan data pemesanan berdasarkan bulan dan berdasarkan jenis booking, maka hasil dari proses tersebut untuk tersebut untuk pengelompokan data dan cleaning data.

4) Data Transformation

Proses penggabungan data yang sudah diseleksi dimasukkan kedalam software rapidminer lalu di proses sinkronisasi dengan data read excel untuk diproses import data, lalu dipilih format data untuk Metode Booking sebagai id dan bulan Januari - Desember sebagai integer.

3. HASIL DAN PEMBAHASAN

a. Data understanding

File data Excel dari Hotel The Hive berisi 3587 baris dan 12 kolom informasi, termasuk rincian tamu, tipe kamar, tanggal kedatangan, lama menginap, tarif, diskon, serta jenis pemesanan, mencakup seluruh transaksi bulan Januari dalam Tabel 1.

Tabel 1. Data awal riwayat tamu menginap

From: 01-Jan-2023		To: 31-Dec-2023		Residents (Predicted)															
Department	Room	Rate	Room Type	Arrival	Departure	Rate	Rate Code	Discount	Net Rate	Company	Residence	Residence	Residence	Residence	Residence	Residence	Residence	Residence	Residence
Hotel The Hive	01-Jan-2023	31-Dec-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023
Hotel The Hive	01-Jan-2023	31-Dec-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023
Hotel The Hive	01-Jan-2023	31-Dec-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023
Hotel The Hive	01-Jan-2023	31-Dec-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023	01-Jan-2023

Sumber: Data Hotel The Hive 2023

b. Data selection

Data dipilih dengan menghilangkan variabel yang tidak diperlukan, menyisakan arrival (tanggal, bulan, tahun) dan *ComProfile* (jenis booking), yang ditampilkan dalam Tabel 2.

Tabel 2. Pemilihan data yang digunakan

From	To
01/Jan/2023	01/Feb/2023
Best Western Premier The Hive	
Bedrooms	
Arrival	ComProfile
01/Jan/2023	C. WALK IN
01/Jan/2023	A. BOOKING.COM. NOCOMMINBESTCHEQUE
01/Jan/2023	A. TIKET
01/Jan/2023	A. TRAVELOKA.COM T&T
01/Jan/2023	A. TRAVELOKA.COM T&T
01/Jan/2023	A. TRAVELOKA.COM T&T

Sumber : Dokumen Pribadi

c. Data cleaning

Data telah dibersihkan dari duplikasi dan inkonsistensi, dikelompokkan berdasarkan bulan dan jenis *booking*, serta disusun ulang dengan kolom Metode *Booking* dan bulan Januari-Desember 2023, seperti ditampilkan dalam Tabel 3.

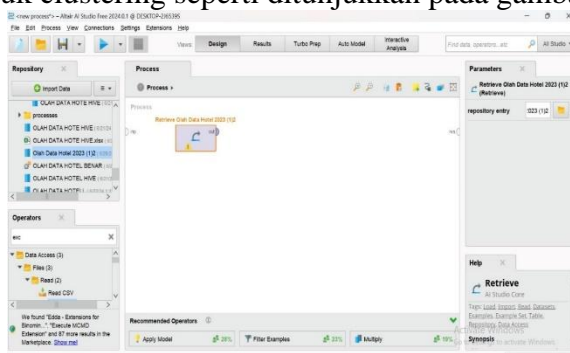
Tabel 3. Hasil pembersihan dan pemeriksaan data

Metode Booking	Tahun 2023											
	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	September	Oktober	November	Desember
Booking.com	271	271	570	423	444	401	511	271	51	3	4	3
Tiket.com	203	220	300	300	194	140	302	10	57	3	0	253
traveleka	1000	886	1000	1000	1000	842	776	25	3	0	0	513
Website Hotel	625	512	400	442	513	417	430	510	450	524	400	400
Pegada	0	57	70	145	110	100	100	23	10	0	0	0
tripAdvisor	2	2	6	12	25	22	10	90	173	100	227	100
Wika Pality	24	13	6	0	0	10	17	10	10	17	10	10
Wika Combed	72	54	57	61	70	72	100	200	110	60	62	175

Sumber : Dokumen Pribadi

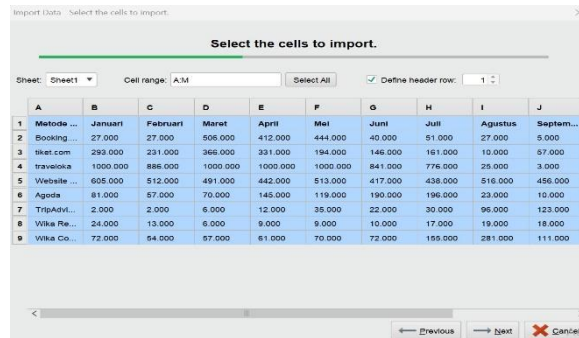
b. Data transformation

Data yang telah diseleksi diimpor ke *RapidMiner*, disinkronkan dengan data Excel, diformat dengan Metode *Booking* sebagai ID dan bulan Januari-Desember sebagai integer, kemudian diproses untuk clustering seperti ditunjukkan pada gambar 2 hingga 6.

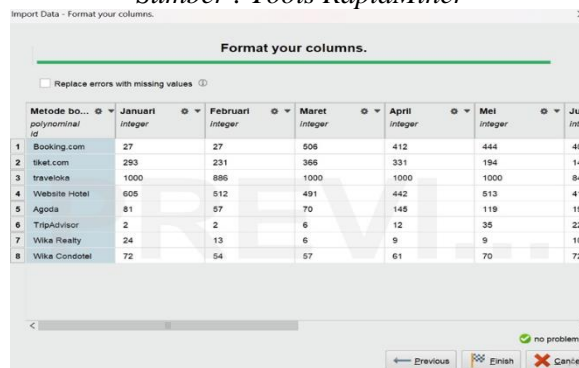


Gambar 2. Import data pada rapidminer

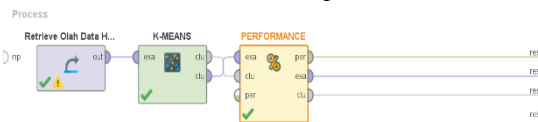
Sumber : Tools RapidMiner



Gambar 3. Kolom data untuk diimport
Sumber : Tools RapidMiner



Gambar 4. Pemilihan format kolom data untuk proses clustering
Sumber: Tools RapidMiner



Gambar 5. Proses clustering
Sumber : Tools RapidMiner

Pengolahan K-Means

Penelitian ini menggunakan data hasil *cleaning* yang dikelompokkan menjadi tiga kluster (0, 1, dan 2) menggunakan algoritma K-means, seperti ditunjukkan pada gambar 6.

	Tahun 2022											
Metode Booking	Januari	Februari	Marat	April	Mei	Juni	Juli	Agustus	September	Oktober	November	Desember
Booking.com	27	27	505	412	444	40	51	27	5	3	4	3
iket.com	293	231	366	331	194	194	161	10	51	2	0	25
traveloka	1000	886	1000	1000	1000	1000	841	776	29	0	0	513
Website Hotel	605	512	491	442	513	417	438	515	459	524	490	491
Agoda	81	57	70	145	110	100	196	23	18	0	0	0
TripAdvisor	2	2	6	12	35	12	31	98	123	389	237	100
Wika Realty	24	13	6	9	9	13	17	18	18	23	18	30
Wika Condotel	72	54	57	61	70	72	155	281	111	68	92	130

Gambar 6. Akurasi K-Means
Sumber: Dokumen Pribadi

Proses perhitungan untuk melakukan clustering pada kuantitas pengunjung yang menginap di Hotel The Hive menggunakan metode K-Means melibatkan beberapa langkah.

- 1) Jumlah cluster yang sesuai ditentukan dengan menghitung nilai k, dan dalam penelitian ini k = 3 dipilih untuk mewakili kelompok pengunjung dengan jumlah menginap rendah, sedang, dan tinggi dalam cluster 0.
- 2) Titik data dipilih secara acak sebagai pusat cluster, dengan data dari Agoda (nomor 5), Wika Realty (nomor 7), dan Wika Condotel (nomor 8) digunakan sebagai titik awal dalam penelitian ini.
- 3) Proses pengelompokan melibatkan penentuan jumlah booking terendah dari setiap data dan menghitung jarak dari pusat cluster, dengan hasil perhitungan dan pemilihan

centroid disajikan dalam Tabel 4.

Tabel 4. Data secara random sebagai centorid (pusat data).

Data	Centroid Awal	Januari	Februari	Maret	April	Mei	Jun	Juli	Agustus	September	Oktober	November	Desember
05C0		81	57	70	145	119	100	196	23	10	0	0	0
07C1		24	13	6	9	9	10	17	19	18	27	16	10
00C2		72	54	57	61	70	72	155	281	111	68	92	135

Sumber : Dokumen Pribadi

Tahap perhitungan per tiga metode booking yaitu Booking.com, Tiket.com, dan Traveloka jarak pusat cluster untuk C0, C1 dan C2 sebagai berikut:

a) Cluster 0.

$$\begin{aligned}
 & \sqrt{(293 - 81)^2} + \sqrt{(231 - 57)^2} + \\
 & \sqrt{(366 - 70)^2} + \sqrt{(331 - 145)^2} + \\
 & \sqrt{(194 - 119)^2} + \sqrt{(146 - 190)^2} + \\
 & \sqrt{(161 - 196)^2} + \sqrt{(10 - 23)^2} + \\
 & \sqrt{(57 - 10)^2} + \sqrt{(3 - 0)^2} + \sqrt{(0 - 0)^2} + \\
 & \sqrt{(252 - 0)^2} \\
 & = 1.5031
 \end{aligned}$$

$$\begin{aligned}
 & \sqrt{(293 - 24)^2} + \sqrt{(231 - 13)^2} + \\
 & \sqrt{(366 - 6)^2} + \sqrt{(331 - 9)^2} + \\
 & \sqrt{(194 - 9)^2} + \sqrt{(146 - 10)^2} + \\
 & \sqrt{(161 - 17)^2} + \sqrt{(10 - 19)^2} + \\
 & \sqrt{(57 - 18)^2} + \sqrt{(3 - 27)^2} + \sqrt{(0 - 16)^2} + \\
 & \sqrt{(252 - 10)^2} \\
 & = 1.4436
 \end{aligned}$$

$$\begin{aligned}
 & \sqrt{(293 - 72)^2} + \sqrt{(231 - 54)^2} + \\
 & \sqrt{(366 - 57)^2} + \sqrt{(331 - 61)^2} + \\
 & \sqrt{(194 - 70)^2} + \sqrt{(146 - 72)^2} + \\
 & \sqrt{(161 - 155)^2} + \sqrt{(10 - 281)^2} + \\
 & \sqrt{(57 - 111)^2} + \sqrt{(3 - 68)^2} + \\
 & \sqrt{(0 - 92)^2} + \sqrt{(252 - 135)^2} \\
 & = 1.5031
 \end{aligned}$$

b) Cluster 1

$$\begin{aligned}
& \sqrt{(27 - 81)^2} + \sqrt{(27 - 57)^2} + \\
& \sqrt{(506 - 70)^2} + \sqrt{(412 - 145)^2} + \\
& \sqrt{(444 - 119)^2} + \sqrt{(40 - 190)^2} + \\
& \sqrt{(51 - 196)^2} + \sqrt{(27 - 23)^2} + \\
& \sqrt{(5 - 10)^2} + \sqrt{(3 - 0)^2} + \sqrt{(4 - 0)^2} + \\
& \sqrt{(3 - 0)^2} \\
& = 2.095
\end{aligned}$$

$$\begin{aligned}
& \sqrt{(27 - 24)^2} + \sqrt{(27 - 13)^2} + \\
& \sqrt{(506 - 6)^2} + \sqrt{(412 - 9)^2} + \\
& \sqrt{(444 - 9)^2} + \sqrt{(40 - 10)^2} + \\
& \sqrt{(51 - 17)^2} + \sqrt{(27 - 19)^2} + \sqrt{(5 - 18)} \\
& + \sqrt{(3 - 27)^2} + \sqrt{(4 - 16)^2} + \\
& \sqrt{(3 - 10)^2} \\
& = 2.095
\end{aligned}$$

$$\begin{aligned}
& \sqrt{(27 - 24)^2} + \sqrt{(27 - 13)^2} + \\
& \sqrt{(506 - 6)^2} + \sqrt{(412 - 9)^2} + \\
& \sqrt{(444 - 9)^2} + \sqrt{(40 - 10)^2} + \\
& \sqrt{(51 - 17)^2} + \sqrt{(27 - 19)^2} + \sqrt{(5 - 18)} \\
& + \sqrt{(3 - 27)^2} + \sqrt{(4 - 16)^2} + \\
& \sqrt{(3 - 10)^2} \\
& = 2.095
\end{aligned}$$

c) Cluster 2

$$\begin{aligned}
& \sqrt{(1000 - 81)^2} + \sqrt{(886 - 57)^2} + \\
& \sqrt{(1000 - 70)^2} + \sqrt{(1000 - 145)^2} + \\
& \sqrt{(1000 - 119)^2} + \sqrt{(841 - 190)^2} + \\
& \sqrt{(776 - 196)^2} + \sqrt{(25 - 23)^2} + \\
& \sqrt{(3 - 10)^2} + \sqrt{(0 - 0)^2} + \sqrt{(1 - 0)^2} + \\
& \sqrt{(511 - 0)^2} \\
& = 4.286
\end{aligned}$$

$$\begin{aligned}
& \sqrt{(1000 - 24)^2} + \sqrt{(886 - 13)^2} + \\
& \sqrt{(1000 - 6)^2} + \sqrt{(1000 - 9)^2} + \\
& \sqrt{(1000 - 9)^2} + \sqrt{(841 - 10)^2} + \\
& \sqrt{(776 - 17)^2} + \sqrt{(25 - 19)^2} + \\
& \sqrt{(3 - 18)} + \sqrt{(0 - 27)^2} + \sqrt{(1 - 16)^2} + \\
& \sqrt{(511 - 10)^2} \\
& = 4.286
\end{aligned}$$

$$\begin{aligned}
& \sqrt{(1000 - 72)^2} + \sqrt{(886 - 54)^2} + \\
& \sqrt{(1000 - 57)^2} + \sqrt{(1000 - 61)^2} + \\
& \sqrt{(1000 - 70)^2} + \sqrt{(841 - 72)^2} + \\
& \sqrt{(776 - 155)^2} + \sqrt{(25 - 281)^2} + \\
& \sqrt{(3 - 111)^2} + \sqrt{(0 - 68)^2} + \\
& \sqrt{(1 - 92)^2} + \sqrt{(511 - 135)^2} \\
& = 4.286
\end{aligned}$$

Pengelompokan data jenis booking Wika Condotel dilakukan dengan memasukkan rata-rata jumlah ke dalam klaster yang sesuai, di mana nilai tertinggi dimasukkan ke klaster 2, terendah ke klaster 0, dan sedang ke klaster 1.

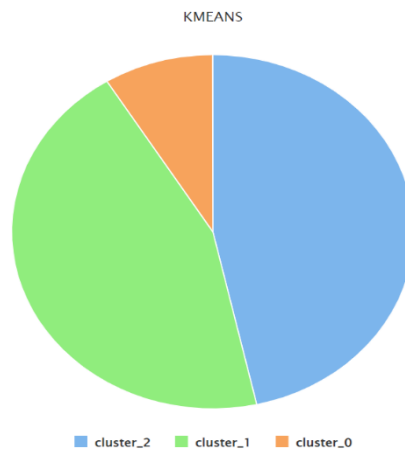
4) Berdasarkan pada tabel 5 menunjukkan hasil cluster model berdasarkan cluster 0, cluster 1 dan cluster 2.

Tabel 5. Cluster Model

Attribute	cluster_0	cluster_1	cluster_2
Januari	83.167	605	1000
Oktober	48.167	524	0
Agustus	78	516	25
Mai	145.167	513	1000
Februari	64	512	886
Maret	168.500	491	1000
Desember	96.667	491	511
November	58.167	450	1
September	54	456	3
April	161.667	442	1000
Juli	101.667	438	776
Juni	80	417	841

Sumber : Tools RapidMiner

5) Pengolahan data dengan RapidMiner menghasilkan diagram klasterisasi yang menunjukkan jumlah tamu tertinggi di klaster 2 (biru), sedang di klaster 1 (hijau), dan terendah di klaster 0 (oranye), seperti ditampilkan pada gambar 7.



Gambar 7. Diagram cluster.

Sumber : Tools RapidMiner

Akurasi K-Means

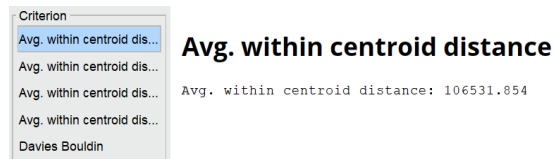
1. Akurasi diukur menggunakan rumus Euclidean Distance, dengan hasil 1.5031 untuk klaster 0, 2.095 untuk klaster 1, dan 4.286 untuk klaster 2, seperti terlihat pada Tabel 6.

Tabel 6. Hasil Perhitungan Manual

C0	C1	C2
1.5031	2.095	4.286

Sumber : Dokumen Pribadi

- urasi diukur menggunakan metrik Avg. within centroid distance dan Davies Bouldin, dengan hasil menunjukkan kinerja baik untuk model K-Means, yaitu nilai rata-rata Avg. within centroid distance sebesar 106531.854, Avg. within centroid distance_cluster_0 sebesar 142042.472, dan Davies Bouldin sebesar 0.228, seperti terlihat pada gambar 13 dan 15.



Gambar 8. Akurasi Avg. Within centroid distance
Sumber: Tools RapidMiner



Gambar 9. Akurasi Avg. Within centroid distance 0
Sumber: Tools RapidMiner



Gambar 10. Akurasi Davies Bouldin
Sumber: Tools RapidMiner

Penulis pada pengolahan data menggunakan software rapidminer dengan metode K-Means berhasil mengelompokkan tamu menginap berdasarkan jenis booking sejak bulan Januari - Desember 2023 dengan cukup baik. Cluster yang terbentuk memiliki karakteristik yang jelas dan dapat digunakan untuk pengambilan keputusan strategis oleh manajemen hotel dalam upaya meningkatkan layanan jenis booking dengan jumlah pemesanan terendah. Namun, terdapat beberapa tamu yang berada pada cluster sedang yang mungkin memerlukan metode pengelompokan tambahan guna hasil yang lebih akurat.

4. KESIMPULAN

- Berdasarkan perhitungan manual dan metode K-Means, delapan kolom data booking tamu hotel di The Hive dibagi menjadi tiga cluster. Cluster 0 (rendah) mencakup 6 jenis booking: Booking.com, Tiket.com, Agoda, TripAdvisor, Wika Realty, dan Wika Condotel. Cluster 1 (sedang) hanya mencakup Website Hotel, dan cluster 2 (tinggi) hanya mencakup Traveloka. Cluster 2 menunjukkan daya tarik pemesanan tertinggi dengan akurasi 4.286, sedangkan cluster 1 dan 0 memiliki jumlah pemesanan lebih rendah dengan akurasi masing-masing 2.095 dan 1.5031.
- Pengujian dengan sistem RapidMiner menggunakan metode K-Means menunjukkan

bahwa cluster 2 (Traveloka) memiliki jumlah tamu tertinggi dengan 7043 data. Cluster 1 (Website Hotel) memiliki jumlah tamu sedang dengan 5895 data, sedangkan cluster 0 (Booking.com, Tiket.com, Agoda, TripAdvisor, Wika Realty, dan Wika Condotel) memiliki jumlah tamu terendah dengan 6823 data.

5. DAFTAR PUSTAKA

- Annisa, K., Ginting, B. S., & Syar, M. A. (2022). Penerapan data mining pengelompokan data pengguna air bersih berdasarkan keluhannya menggunakan metode clustering pada PDAM Langkat. *Jurnal Sistem Informasi Kaputama (JSIK)*, 6(2), 165–179. <https://doi.org/10.59697/jsik.v6i2.167>
- Annur, H. (2022). Penerapan algoritma Naïve Bayes berbasis backward elimination untuk prediksi pemesanan kamar hotel. *Jurnal Ilmiah Ilmu Komputer Banthayo Lo Komputer*, 1(1), 1–5. <https://doi.org/10.37195/balok.v1i1.99>
- Atma, Y. A., & Montesori, S. (2022). Analisis data mining untuk menentukan profit perusahaan menggunakan metode K-Means. *Device: Journal of Information System, Computer Science and Information Technology*, 3(2), 29–36. <https://doi.org/10.46576/device.v3i2.2699>
- Febrian, A., Suarna, N., & Dwilestari, G. (2022). Penerapan algoritma K-Means untuk mengelompokkan data pengiriman paket di kantor pos Cirebon. *Jurnal Teknologi Technoscintia*, 15(1), 23–27. <https://doi.org/10.34151/technoscintia.v15i1.3858>
- Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan algoritma K-Means untuk clustering data obat-obatan. *Jurnal Nasional Teknologi dan Sistem Informasi*, 5(1), 17–24. <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>
- Mulyawan, Bahtiar, A., Dwilestari, G., Basysyar, F. M., & Suarna, N. (2021). Data mining techniques with machine learning algorithm to predict patients of heart disease. *IOP Conference Series: Materials Science and Engineering*, 1088(1), 012035. <https://doi.org/10.1088/1757-899X/1088/1/012035>
- Purba, W., Siawin, W., & H., H. (2019). Implementasi data mining untuk pengelompokan dan prediksi karyawan yang berpotensi PHK dengan algoritma K-Means clustering. *Jurnal Sistem Informasi dan Ilmu Komputer Prima (JUSIKOM PRIMA)*, 2(2), 85–90. <https://doi.org/10.34012/jusikom.v2i2.429>
- Sari, R. W., & Hartama, D. (2018). Data mining: Algoritma K-Means pada pengelompokan wisata asing ke Indonesia menurut provinsi. *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, 322–326.