JURNAL ILMIAH
MULTIDISIPLIN ILMU
Halaman Jurnal: https://journal.smartpublisher.id/index.php/jimi
Halaman UTAMA Jurnal : https://journal.smartpublisher.id/

DOI: https://doi.org/10.69714/3hhfj353

IMPLEMENTASI K-MEANS *CLUSTER*ING DALAM PENGELOMPOKAN DATA KUNJUNGAN WISATAWAN ASING DI INDONESIA

Miftahul Arif Aldi [a*], Zaehol Fatah [b]
[a] Teknologi Informasi, dyroth299@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur
[b] Sistem Informasi, zaeholfatah@gmail.com, Universitas Ibrahimy, Situbondo Jawa Timur
[*] korespondensi

ABSTRACT

Clustering is a data mining technique used for grouping data based on specific similarities. This study implements K-Means Clustering to analyze foreign tourist visit data in Indonesia in 2024. Using the Knowledge Discovery in Database (KDD) methodology, the research involves five stages: Data Selection, preprocessing, Transformation, data mining, and Evaluation. Data Clustering was conducted using RapidMiner software, experimenting with different cluster counts (k=2 to k=7) to determine the optimal number of clusters. Results indicate that three clusters (k=3) with the smallest Davies-Bouldin Index (DBI) value were optimal. This Clustering approach categorizes tourists into low, medium, and high visit groups, assisting policymakers in strategic tourism development. The findings support capacity planning and seasonal marketing strategies to optimize Indonesia's tourism sector.

Keywords: *Clustering, K-Means, travelers, RapidMiner, Clustering.*

Abstrak

Klasterisasi adalah teknik penambangan data yang digunakan untuk mengelompokkan data berdasarkan kesamaan tertentu. Penelitian ini menerapkan Klasterisasi *K-Means* untuk menganalisis data kunjungan wisatawan asing di Indonesia pada tahun 2024. Menggunakan metodologi *Knowledge Discovery in Database* (KDD), penelitian ini melibatkan lima tahap: Pemilihan Data, praproses, Transformasi, penambangan data, dan Evaluasi. Klasterisasi data dilakukan menggunakan perangkat lunak *RapidMiner*, dengan eksperimen jumlah klaster yang berbeda (k=2 hingga k=7) untuk menentukan jumlah klaster yang optimal. Hasil penelitian menunjukkan bahwa tiga klaster (k=3) dengan nilai *Davies-Bouldin Index* (DBI) terkecil adalah yang optimal. Pendekatan klasterisasi ini mengkategorikan wisatawan menjadi kelompok kunjungan rendah, sedang, dan tinggi, yang dapat membantu pembuat kebijakan dalam pengembangan pariwisata yang strategis. Temuan ini mendukung perencanaan kapasitas dan strategi pemasaran musiman untuk mengoptimalkan sektor pariwisata Indonesia.

Kata Kunci: *Clustering*, *K-Means*, wisatawan, *RapidMiner*, pengelompokan.

1. PENDAHULUAN

Clustering atau pengelompokan data adalah salah satu teknik dalam data mining yang digunakan untuk mengelompokkan data berdasarkan kesamaan tertentu [1] Dalam sektor pariwisata, pengelompokan data menjadi hal yang penting untuk membantu pihak terkait dalam menganalisis pola kunjungan wisatawan serta merancang strategi pengembangan destinasi wisata yang lebih terarah dan efektif. Salah satu metode pengelompokan yang sering digunakan adalah *K-Means Clustering*, yang memungkinkan pengelompokan data ke dalam beberapa kelompok berdasarkan kesamaan atribut yang dimiliki oleh data tersebut.

Indonesia, sebagai tujuan wisata populer, menarik kedatangan jutaan wisatawan asing setiap tahunnya. Data mengenai jumlah, asal negara, serta preferensi destinasi wisatawan sangat beragam. Oleh karena itu, penerapan metode *K-Means Clustering* untuk mengelompokkan data kunjungan wisatawan asing di

Received Desember 1, 2024; Revised Desember 3, 2024; Accepted Januari 22, 2025; Published Februari 1, 2025

Indonesia sangat direkomendasikan. Dengan mengimplementasikan *K-Means*, data kunjungan dapat dibagi ke dalam beberapa *cluster* yang memiliki karakteristik serupa, sehingga memberikan pemahaman yang lebih mendalam tentang pola kunjungan wisatawan dan potensi pasar yang dapat dikembangkan.

Algoritma *K-Means* untuk *Clustering* telah digunakan dalam sejumlah penelitian sebelumnya untuk menerapkan topik analisis *Clustering*. Penelitian "*Data Mining Clustering Menggunakan Algoritma K-Means Pada Data Kunjungan Wisatawan Di Kabupaten Karawang"*[2].Teknik *K-Means* digunakan dalam penelitian tersebut untuk menerapkan pengelompokan data mining. Aplikasi *RapidMiner* digunakan untuk menerapkan pengelompokan data mining. Tujuan analisis pengelompokan adalah untuk menentukan nilai k yang ideal untuk mengklasifikasi objek wisata. Berdasarkan hasil penelitian tersebut, kelompok 1 mewakili kelompok wisatawan rendah, kelompok 2 mewakili kelompok wisatawan sedang, dan kelompok 3 mewakili kelompok wisatawan tinggi. Jumlah k yang ideal adalah 3.. Kemudian pada penelitian "Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode *K-Means*"[3]. Tiga klaster, yakni klaster 1, 2, dan 3, diidentifikasi dalam penelitian tersebut. Dengan rata-rata 15.611 hingga 46.783 pengunjung, Klaster 1 memiliki kategori kunjungan wisatawan sedang. Ini mencakup destinasi wisata Kraton, Taman Pintar, dan Kebun Binatang Gembira Loka. Dari 46.784 hingga 91.566 pengunjung di Klaster 2, 23% masuk ke objek wisata di Taman Pintar dan Kebun Binatang Gembira Loka. Dari 15.610 pengunjung di Klaster 3, 27% masuk ke Taman Pintar dan museum, sehingga masuk kategori rendah. Menurut data, pengunjung datang ke tempat-tempat tersebut pada berbagai waktu dalam setahun. Lalu pada penelitian "*Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Mancanegara Di Prov. Sulawesi Selatan Dengan K-Means Dan SVM*"[4]. Penelitian tersebut bertujuan untuk menganalisis penggunaan data mining dalam mengklasifikasi jumlah pengunjung mancanegara ke prefektur Sulawesi Selatan menggunakan *K-Means*. Data yang digunakan berasal dari BPS Prov. Sulawesi Selatan dan dikelompokkan menjadi dua *cluster*: pengunjung terbanyak adalah C1, dengan hasil dari Malaysia, dan jumlah pengunjung paling sedikit adalah C0, dengan hasil dari Singapura, Jepang, Korea Selatan, Taiwan, China, India, Filipina, Hong Kong, Thailand, Australia, USA, UK, Belanda, Jerman, Prancis, Rusia, Arab Saudi, Mesir, Uni Emirat Arab, Pearl of the Persian Gulf, dan Swiss. Kemudian data tersebut diproses kembali dengan SVM untuk menentukan nilai precision dan recall, dan memperoleh akurasi 100,00% dalam aplikasi *RapidMiner*. Kemudian Penelitian "*Data Mining Klasterisasi Customer Segmantation Netflix Menggunakan Metode Kmeans Dengan RapidMiner*"[5]. Meskipun pada variable yang digunakan pada penelitian tersebut berbeda, akan tetapi penelitian tersebut menunjukan bahwasanya metode *K-Means* dapat digunakan dalam pengklasteran sebuah data dan dapat menghasilkan data yang dibutuhkan untuk menjadi sebuah informasi nantinya.

Secara keseluruhan, penelitian yg dilakukan ini tidak hanya berkontribusi pada pemahaman tentang perilaku wisatawan asing tetapi juga memberikan dasar bagi pengambilan keputusan strategis dalam pengembangan pariwisata di Indonesia. Dengan memanfaatkan teknik *K-Means Clustering*, diharapkan para pemangku kepentingan dapat merumuskan kebijakan yang lebih tepat sasaran dalam menarik dan mempertahankan wisatawan asing.

2. TINJAUAN PUSTAKA

2.1. Data Mining

Data mining adalah metode yang memungkinkan para penggunanya untuk mengakses data yang besar dalam waktu yang relatif cepat. Atau dengan kata lain data mining merupakan suatu alat dan aplikasi menggunakan analisis statistik pada data melalui suatu proses ekstraksi atau penggalian data dan informasi yang belum diketahui sebelumnya. Secara sederhana data mining merupakan proses penggalian suatu data yang berujung pada penemuan informasi terbaru dengan cara mencari pola atau aturan tertentu dari sejumlah data yang sangat besar, sehingga cara kerja dari data mining sebenarnya adalah untuk memeriksa *database* yang berukuran besar guna menemukan pola atau bentuk yang baru sehingga berguna dalam proses pengambilan keputusan. [6]

2.1.1. *Cluster*ing

Clustering adalah untuk mengelompokkan data dengan karakteristik yang sama ke satu wilayah yang sama dan data dengan karakteristik yang berbeda ke wilayah yang lain. Manfaat *Clustering* sebagai segmentasi data yang berguna untuk memprediksi dan menganalisis masalah bisnis, serta mengidentifikasi obyek dalam berbagai bidang[7].

2.1.2. K-Means

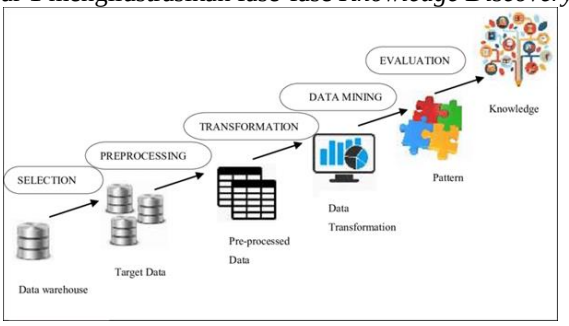
K-Means merupakan salah satu algoritma yang bersifat *unsupervised learning*. *K-Means* memiliki fungsi untuk mengelompokkan data ke dalam data *cluster*. Algoritma ini dapat menerima data tanpa ada label kategori. *K-Means Clustering* Algoritma juga merupakan metode *non-hierarchy*. Metode *Clustering* Algoritma adalah mengelompokkan beberapa data ke dalam kelompok yang menjelaskan data dalam satu kelompok memiliki karakteristik yang sama dan memiliki karakteristik yang berbeda dengan data yang ada di kelompok lain. *Cluster* Sampling adalah teknik pengambilan sampel di mana unit-unit populasi dipilih secara acak dari kelompok yang sudah ada yang disebut '*cluster*, nah *Clustering* atau klasterisasi adalah salah satu masalah yang menggunakan teknik *unsupervised learning*.[8]

2.1.3. Rapidminer

Rapid Miner merupakan perangkat lunak yang dibuat oleh Dr. Markus Hofmann dari *Institute Of Technology Blanchardstown* dan Rafl Klinkenberg dari rapid-i.com dengan tampilan GUI sehingga memudahkan pengguna dalam menggunakan perangkat lunak ini. Perangkat lunak ini bersifat open source dan dibuat dengan menggunakan bahasa Jawa di bawah lisensi GNU *Public License* dan *RapidMiner* dapat dijalankan di sistem operasi manapun. Dengan menggunakan Rapid Miner, tidak dibutuhkan kemampuan koding khusus, karena semua fasilitas sudah disediakan. *RapidMiner* dikhususkan untuk penggunaan Data Mining. [9]

3. METODOLOGI PENELITIAN

Pendekatan *Knowledge Discovery in Database* (KDD) merupakan metodologi penelitian yang digunakan. Pemilihan data, praproses, transformasi, penggalian data, dan evaluasi/interpretasi merupakan lima langkah dalam proses ini[10]. Gambar 1 mengilustrasikan fase-fase *Knowledge Discovery in Database* (KDD).



Gambar 1 Tahapan *Knowledge Discovery in Database* (KDD)

3.1.1. Data *Selection*

Teknik *Knowledge Discovery in Database* (KDD) menyatakan bahwa pemilihan data merupakan langkah awal dalam proses data mining. Sebelum beralih ke tahap *preprocessing* , Data *Selection* merupakan proses memilih data mana dari kumpulan data yang akan digunakan. Dalam penelitian ini, seleksi data dilakukan dengan memfokuskan pada data Jumlah Kunjungan Wisatawan Mancanegara per bulan Menurut Kebangsaan (Kunjungan) Tahun 2024.

3.1.2. *Preprocessing*

Tahap *preprocessing* adalah proses pembersihan data, yang meliputi penghilangan data yang bernilai null atau hilang, penghapusan atribut yang kurang relevan dengan penelitian, serta pembersihan data yang mengandung anomali. Selain itu, pada tahap ini juga diberikan identitas atau id pada data yang akan dianalisis. Penentuan id dilakukan menggunakan atribut yang bersifat unik, yaitu nama Kebangsaan. Atribut yang digunakan dalam tahap ini meliputi Januari, Februari, Maret, April, Mei, Juni, Juli, Agustus, September, dan Oktober.

3.1.3. *Transformation*

Pada tahap transformasi, data diubah agar sesuai dengan tipe data yang dibutuhkan untuk proses selanjutnya, sehingga siap untuk diolah. Pada tahap ini, data non-numerik diubah menjadi data numerik sesuai kebutuhan analisis .

3.1.4. Data Mining

Tahap data mining merupakan inti dari proses pencarian informasi dari dataset yang digunakan. Dalam tahap ini, informasi baru yang dapat memberikan wawasan antara data diekstraksi untuk membantu pengambilan

keputusan bagi pihak atau organisasi yang terlibat . Proses data mining dilakukan menggunakan perangkat lunak *RapidMiner* v.10.2.

3.1.5. Evaluation

Pada tahap evaluasi, hasil analisis ditampilkan dalam format yang mudah dipahami, baik dalam bentuk tabel maupun grafik . Evaluasi akhir dilakukan dengan menampilkan hasil dari proses *Clustering* data mengenai Jumlah Kunjungan Wisatawan Mancanegara per bulan Menurut Kebangsaan (Kunjungan) Tahun 2024.

4. HASIL DAN PEMBAHASAN

Berikut adalah hasil serta pembahasan dari penelitian implementasi *K-Means Clustering* dalam pengelompokan data kunjungan wisatawan asing di Indonesia.

4.1 *Data Selection*

Data yang digunakan merupakan data Jumlah Kunjungan Wisatawan Mancanegara per bulan Menurut Kebangsaan (Kunjungan) pada tahun 2024 dengan total jumlah data sebanyak 235 data yang terdiri dari 11 attribut.

4.2 *Preprocessing*

Langkah selanjutnya adalah praproses, yaitu memilih atribut yang akan digunakan dan menambahkan *id* atau identitas pada data yang akan diproses. Pada tahap ini, operator *Select Attributes* digunakan untuk menghapus atribut yang tidak akan digunakan..

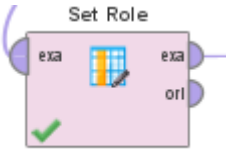
Gambar 2 Operator *Select Attributes*

Untuk memilih karakteristik mana yang akan digunakan dan mana yang tidak, opsi *subset* dipilih dalam argumen tipe filter atribut. Hasil dari pemilihan kualitas yang digunakan adalah sebagai berikut.

Tabel 1 Atribut yang dipilih

Attributes	*Selected Attributes*
November	Kebangsaan
Desember	Januari
Tahunan	Februari
	Maret
	April
	Mei
	Juni
	Juli
	Agustus
	September
	Oktober

Langkah selanjutnya adalah memilih *id* data setelah memilih atribut. Data yang dipilih harus unik atau bebas dari duplikat. *Set Role* adalah operator yang digunakan.



Gambar 3 Operator *Set Role*

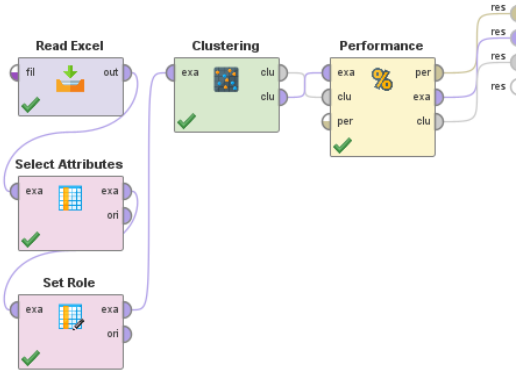
Atribut Kebangsaan dipakai sebagai *id* atau identitas dalam data.

Tabel 2 Attribut ID

Attributes	***Selected* Atributes**
Kebangsaan	id

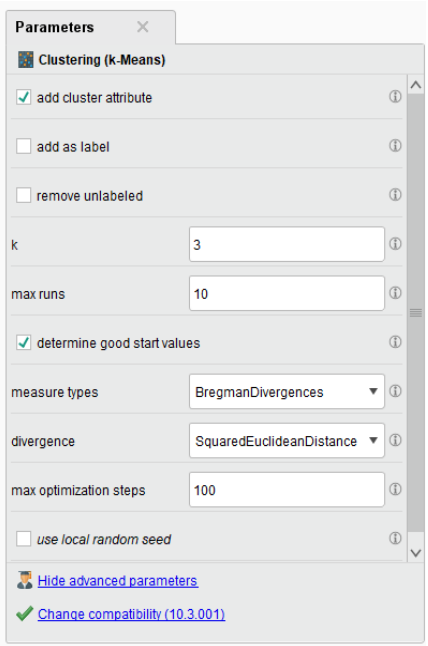
4.3 Data Mining

Pada tahap ini, pengelompokan dilakukan menggunakan operator Pengelompokan *K-Means*, dan nilai *Davies Bouldin* digunakan untuk mengukur kinerja pengelompokan menggunakan operator Kinerja Jarak Kluster.



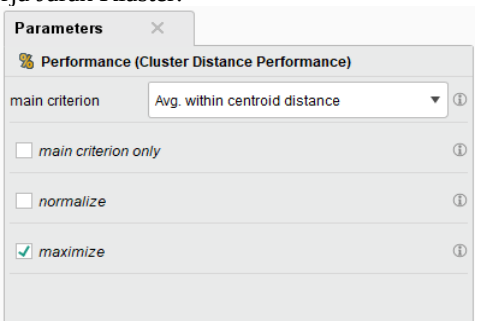
Gambar 4 Model proses data mining *K-Means*

Untuk menentukan jumlah k yang ideal, dilakukan enam kali percobaan dalam penelitian ini, dimulai dengan k=2, k=3, k=4, k=5, k=6, dan k=7.



Gambar 5 Parameter operator *K-Means*

Untuk mendapatkan hasil *Davies Bouldin* terbaik, operator *K-Means* menggunakan berbagai jenis pengukuran. Untuk memastikan bahwa temuan *Davies Bouldin* tidak negatif, klik tanda centang di bagian maksimum untuk operator Kinerja Jarak Klaster.



Gambar 6 Parameter *Cluster* Distance Performance

4.4 *Evaluation*

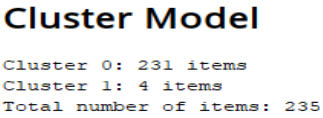
Tabel 3 adalah temuan DBI setelah pengelompokan menggunakan *K-Means* mengikuti pengujian sebelumnya.

Tabel 3 Hasil DBI

Jumlah (k)	DBI *Bregman Divergences*
2	0.282
3	0.550
4	0.362
5	0.467
6	0.420
7	0.451

Menggunakan parameter *Bregman Divergences*, yang memiliki nilai 0,349, tabel di atas menunjukkan bahwa DBI minimum terjadi pada k = 3.

Sebagai konsekuensi dari k=3 dan nilai DBI sebesar 0,422, terdapat 31 titik data pada kelompok 1, 61 titik data pada kelompok 2, dan 3 titik data pada kelompok 3. Gambar 7 dan Tabel 4 menunjukkan hasil dari penjelasan yang dijelaskan sebelumnya.



Gambar 7 Hasil *cluster* k=2

Tabel 4 Hasil *cluster* 0

Kebangsaan	Jumlah Kunjungan Wisatawan Mancanegara per bulan Menurut Kebangsaan (Kunjungan) Tahun 2024	
	Januari	…
Brunei Darussalam	747	…
Philippines	16.937	…
Thailand	8449	…
Vietnam	6.772	…
Laos	169	…
Kamboja	737	…
Myanmar/Burma	2818	…
Indonesia	35180	…
Hong Kong	1475	…

... . …

Berdasarkan tabel diatas, klaster 0 termasuk dalam kategori wisata rendah, yang memiliki jangkauan puluhan hingga ratusan mil.

Tabel 5 Hasil *cluster* 1

Kebangsaan	Jumlah Kunjungan Wisatawan Mancanegara per bulan Menurut Kebangsaan (Kunjungan) Tahun 2024	
	Januari	…
Malaysia	155213	…
Singapore	87248	…
China	81691	…
Australia	127097	…

Dari tabel di atas terlihat bahwa klaster 2 memiliki tiga set data dengan kisaran ratusan ribu. Dengan demikian, dapat dikatakan bahwa klaster 1 termasuk kelompok dengan jumlah wisatawan terbanyak.

5. KESIMPULAN DAN SARAN

Dengan klaster 0 yang mewakili kelompok wisatawan rendah dan klaster 0 yang mewakili kelompok wisatawan tinggi, dapat disimpulkan dari hasil pengelompokan di atas bahwa k = 2 merupakan jumlah wisatawan ideal. Oleh karena itu, diharapkan hasil pengelompokan ini akan mendukung taktik pemasaran musiman untuk memaksimalkan pendapatan dan membantu pemerintah dalam merencanakan kapasitas objek wisata, termasuk penginapan, transportasi, dan infrastruktur pendukung.

DAFTAR PUSTAKA

[1] S. Syam *et al.*, *Data Mining : Teori dan Penerapannya dalam Berbagai Bidang*. PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: https://books.google.co.id/books?id=hTAxEQAAQBAJ

[2] K. Gustipartsani, N. Rahaningsih, R. D. Dana, and I. Y. Mustafa, "DATA MINING *CLUSTERING* MENGGUNAKAN ALGORITMA *K-MEANS* PADA DATA KUNJUNGAN WISATAWAN DI KABUPATEN KARAWANG," 2023.

[3] B. S. Purnomo and P. T. Prasetyaningrum, "PENERAPAN DATA MINING DALAM MENGELOMPOKKAN KUNJUNGAN WISATAWAN DI KOTA YOGYAKARTA MENGGUNAKAN METODE *K-MEANS*," 2021.

[4] K. Wisatawan Mancanegara Di Prov Sulawesi Selatan Dengan *K-Means* Dan SVM Nero Caesar Gosari, "Penerapan Data Mining Dalam Mengelompokkan," vol. 8, no. 3, 2023, [Online]. Available: https://sulsel.bps.go.id/searchengine/result.html.

[5] N. Maymuna and Z. Fatah, "DATA MINING KLASTERISASI CUSTOMER SEGMANTATION NETFLIX MENGGUNAKAN METODE KMEANS DENGAN *RAPIDMINER*," vol. 6, pp. 30–41, Nov. 2024.

[6] Y. Ardilla *et al.*, *DATA MINING DAN APLIKASINYA*. Penerbit Widina, 2021. [Online]. Available: https://books.google.co.id/books?id=53FXEAAAQBAJ

[7] P. W. Rahayu *et al.*, *Buku Ajar Data Mining*. PT. Sonpedia Publishing Indonesia, 2024. [Online]. Available: https://books.google.co.id/books?id=vCruEAAAQBAJ

[8] C. Prianto and S. Bunyamin, *Pembuatan aplikasi Clustering gangguan jaringan menggunakan metode K-Means Clustering*. in Knowledge. Kreatif, 2020. [Online]. Available: https://books.google.co.id/books?id=y8TgDwAAQBAJ

[9] Z. Setiawan *et al.*, *BUKU AJAR DATA MINING*. PT. Sonpedia Publishing Indonesia, 2023. [Online]. Available: https://books.google.co.id/books?id=1nLVEAAAQBAJ

[10] R. M. Sari, A. Rizka, N. A. Putri, and A. Efriana, *Perhitungan Metode Clustering*. Serasi Media Teknologi, 2024. [Online]. Available: https://books.google.co.id/books?id=RIU0EQAAQBAJ