

## ANALISIS PENGELOMPOKAN DAYA TARIK OBYEK WISATA BERDASARKAN JENISNYA MENGGUNAKAN METODE K-MEANS PADA DATA PEMPROV JABAR

Anggun Helia<sup>1</sup>, Mulyawan<sup>2</sup>, Cep Lukman Rohmat<sup>3</sup>, Fathurrohman<sup>4</sup>

<sup>1</sup>Teknik Informatika, STMIK IKMI Cirebon

<sup>2</sup>Sistem Informasi, STMIK IKMI Cirebon

<sup>3,4</sup>Rekayasa Perangkat Lunak, STMIK IKMI Cirebon

Jl. Perjuangan No.10B, Karyamulya, Kec. Kesambi, Kota Cirebon, Indonesia

anggunhelia49@gmail.com

### ABSTRAK

Provinsi Jawa Barat merupakan salah satu provinsi di Indonesia yang memiliki daya tarik pariwisata tinggi baik dari kalangan wisatawan lokal maupun mancanegara. Hal ini dapat dilihat dari jumlah kunjungan yang kian tahun semakin naik dari tahun-tahun sebelumnya. Peningkatan kunjungan wisatawan akan memberikan pengaruh baik bagi masyarakat sekitar. Di samping itu, walaupun sudah ada usaha-usaha untuk meningkatkan kunjungan wisatawan, masih terdapat masalah ketidakseimbangan jumlah wisatawan di beberapa daerah. Oleh karena itu, diperlukan usaha untuk meratakan distribusi kunjungan wisatawan. Salah satu cara yang dapat dilakukan adalah dengan mengelompokkan objek wisata yang memiliki potensi untuk dikembangkan dengan menggunakan penerapan clustering metode algoritma k-means. Atribut yang digunakan dalam penelitian ini meliputi Nama Provinsi, Nama, Kabupaten/Kota, Alam, Buatan, Budaya, Satuan, Tahun dengan perolehan hasil *Clustering* rendah, sedang, dan tinggi dengan nilai DBI yang diperoleh dari  $K=4$  adalah 0.146. *Cluster 0* merupakan Tingkat Daya Tarik Obyek Wisata Rendah dengan anggota cluster berjumlah 16 kabupaten, *Cluster 1* merupakan Tingkat Daya Tarik Obyek Wisata sedang dengan anggota cluster berjumlah 9, *Cluster 2* merupakan Tingkat Daya Tarik Obyek Wisata tinggi dengan anggota cluster berjumlah 1 kabupaten, dan *Cluster 3* merupakan Tingkat Daya Tarik Obyek Wisata tinggi dengan anggota cluster berjumlah 1 kota.

**Kata kunci :** *Clustering, Data Mining, K-Means, Kunjungan Wisata*

### 1. PENDAHULUAN

Pariwisata merupakan salah satu sektor yang dapat meningkatkan perekonomian untuk kepentingan masyarakat lokal di sekitar destinasi wisata. Dari sektor tersebut, pariwisata ditetapkan sebagai sektor penggerak (*leading sector*) karena dalam jangka pendek, menengah, maupun jangka panjang diprediksi akan mempunyai pertumbuhan yang positif serta dianggap sebagai sektor yang strategis serta mampu menjadi media integrasi program antar sektor. Khusus di Jawa Barat, banyak sekali tempat wisata yang setiap tahunnya menarik ribuan wisatawan baik dari dalam dan luar negeri untuk menikmati indahnya dan serunya berwisata ke Jawa Barat.

Peran pariwisata dalam meningkatkan ekonomi masyarakat sangat signifikan. Saat ini, Indonesia memiliki beragam destinasi wisata yang memiliki potensi, baik yang bersifat buatan maupun berbasis alam. Di Jawa Barat, upaya pengembangan potensi pariwisata menjadi fokus untuk meningkatkan pendapatan masyarakat. Keberadaan destinasi wisata memberikan peluang ekonomi bagi masyarakat setempat, karena destinasi tersebut dianggap sebagai potensi ekonomi.

Penelitian tentang daya tarik obyek wisata juga telah dilakukan oleh beberapa penelitian sebelumnya, misalnya penelitian yang telah dilakukan oleh Maulidia L dalam sebuah artikel yang berjudul Penerapan Data Mining dalam Mengelompokkan Kunjungan Wisatawan ke Objek Wisata Unggulan di Prov. DKI Jakarta dengan K-Means yang

dipublikasikan pada Jurnal Informatika Sunan Kalijaga (2018). Jurnal ini menganalisis penggunaan metode K-Means dalam mengelompokkan jumlah wisatawan internasional yang mengunjungi Provinsi DKI Jakarta dilakukan melalui aplikasi teknologi data mining. Data penelitian diperoleh dari Badan Pusat Statistik (BPS) Provinsi DKI Jakarta dan mencakup statistik jumlah pengunjung wisata pada periode 2007-2013.. Data tersebut dikelompokkan menjadi 3 *cluster* yaitu C1 = jumlah kunjungan wisatawan tinggi, C2 = jumlah kunjungan wisatawan sedang dan C3 = jumlah kunjungan wisatawan rendah. Nilai centroid akhir yang digunakan pada C1 = 15.438.488, C2 = 4.464.577 dan C3 = 342.332. Sehingga hasil pengelompokan C1 = Taman Impian Jaya Ancol, C2 = Taman Mini Indonesia Indah Dan Kebon Binatang Ragunan dan C3 = Monumen Nasional, Museum Nasional, Museum Satria Mandala, Museum Sejarah Jakarta dan Pelabuhan Sunda Kelapa. Hasil pengelompokan C3 menjadi rekor bagi pemerintah Prov. DKI. Jakarta [1].

Penelitian lain terkait daya tarik obyek wisata juga telah dilakukan oleh Al-Fahmi B dkk dengan judul Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran pada jurnal JSC-Tech *Journal of Computer Science and Technology* (2023) menjelaskan bahwa penggunaan teknik clustering dapat memfasilitasi penentuan label kelas pada data yang belum memiliki klasifikasi. Metode K-Means terkenal karena kemudahannya dan kemampuannya

untuk mengelompokkan data yang besar dan outlier dengan kecepatan yang tinggi. Hasil dari pengolahan data menggunakan algoritma K-Means telah melibatkan lima iterasi, di mana pengelompokan dilakukan secara acak untuk tiga kluster, yaitu *cluster* 1, *cluster* 2, dan *cluster* 3. Pada iterasi tersebut, teridentifikasi bahwa cluster 1 memuat sebanyak 24 data dengan persentase (50%), *cluster* 2 berjumlah 11 data dengan persentase (23%), dan *cluster* 3 berjumlah 13 data dengan persentase (27%) [2].

Berdasarkan beberapa penelitian yang telah dilakukan sebelumnya, dapat disimpulkan bahwa teknik pengelompokan data objek wisata dapat diaplikasikan pada skenario lain dengan atribut yang berbeda. Oleh karena itu, penelitian ini akan menerapkan teknik pengelompokan data daya tarik objek wisata berdasarkan kabupaten/kota, dengan atribut yang akan digunakan yaitu nama provinsi, nama kabupaten/kota, alam, buatan, budaya satuan, dan tahun dengan metode yang digunakan yaitu teknik data mining clustering. Teknik pengelompokan data dalam data mining, yang dikenal sebagai clustering, adalah suatu metode yang mampu mengelompokkan berbagai jenis data berdasarkan karakteristik uniknya, sehingga kelompok yang terbentuk mencerminkan data dengan sifat yang serupa [3]. Dalam penelitian ini, metode yang diterapkan menggunakan algoritma K-Means, di mana pendekatan ini mampu mengelompokkan data berdasarkan pusat cluster atau centroid yang terdekat dengan setiap data [4].

Dari tahun ke tahun, kunjungan wisata di Provinsi Jawa Barat mengalami kenaikan. Dengan adanya kenaikan tersebut fasilitas penunjang kebutuhan wisata pun perlu ditingkatkan. Hal ini merupakan tantangan, khususnya bagi pemerintah dalam mengembangkan objek wisata di Provinsi Jawa Barat agar dapat menarik lebih banyak wisatawan. Salah satu permasalahan yang dihadapi adalah disparitas dalam jumlah kunjungan wisatawan antara objek wisata yang terletak di kota dan kabupaten. Infrastruktur pendukung pariwisata di kota-kota telah berkembang pesat, sementara di daerah kabupaten, kondisinya jauh dari memadai bahkan beberapa di antaranya dianggap tidak layak. Oleh karena itu, diperlukan upaya untuk meratakan distribusi kunjungan wisatawan dengan mengelompokkan objek wisata yang memiliki potensi untuk dikembangkan melalui teknik pengelompokan (clustering). Hal ini bertujuan untuk mengidentifikasi objek wisata yang memiliki potensi dan menetapkan strategi pengembangan yang tepat.

Penelitian ini bertujuan untuk menerapkan teknik data mining, khususnya algoritma K-Means Clustering, dalam menganalisis daya tarik obyek wisata di Provinsi Jawa Barat. Dalam hal ini, penerapan datamining mampu menjadi solusi dalam menganalisa data. *Clustering* termasuk ke dalam *descriptive methods*, dan juga termasuk *unsupervised learning* dimana tidak ada pendefinisian kelas objek sebelumnya. Sehingga clustering dapat digunakan

untuk menentukan label kelas bagi data-data yang belum diketahui kelasnya. Metode K-Means termasuk dalam *partitioning clustering* yang memisahkan data ke daerah bagian yang terpisah.

Berdasarkan uraian diatas, maka peneliti tertarik untuk melakukan clustering daya tarik obyek wisata di Provinsi Jawa Barat menggunakan metode K-Means clustering. Sehingga dalam penelitian ini dapat diketahui tingkat pengaruh daya tarik obyek wisata di masing-masing kabupaten/kota di provinsi Jawa Barat. Variabel pada penelitian ini yaitu nama provinsi, nama kabupaten/kota, alam, buatan, budaya, satuan, dan tahun yang terdiri dari 18 nama kabupaten dan 9 nama kota di Provinsi Jawa Barat.

## 2. LANDASAN TEORI

### 2.1. Data Mining

Data mining, atau yang sering disebut *Knowledge Discovery in Database* (KDD), adalah suatu proses yang mencakup pengumpulan dan pemanfaatan sejarah data untuk menemukan pola, keterkaitan, dan informasi yang tersembunyi dalam set data yang berskala besar. Kegiatan ini menghasilkan keluaran yang dapat digunakan untuk membuat keputusan di masa mendatang. Salah satu metode yang sering diterapkan dalam data mining adalah teknik pengelompokan atau clustering.

### 2.2. Algoritma K-Means

Pendekatan yang sering digunakan dalam teknik pengelompokan adalah Algoritma K-Means, dikarenakan kemudahan pengoperasiannya. Algoritma K-Means dianggap lebih cepat dibandingkan dengan pendekatan hierarkis, sehingga teknik pengelompokan ini menjadi pilihan yang umum digunakan untuk pengolahan informasi dalam skala besar [5]. K-Means merupakan upaya yang dilakukan untuk mengelompokkan data objek yang ada ke dalam satu jenis atau beberapa kelompok atau *cluster* berdasarkan keunikannya, di mana penentuan titik terdekatnya dilakukan melalui proses yang berulang-ulang [6]. Adapun langkah-langkah yang dilakukan dalam pemrosesan algoritma K-Means sebagai berikut :

- Tentukan *cluster* secara sembarang sebagai *centroid* (titik sentral).
- Hitung menggunakan rumus *Euclidian Distance* untuk menentukan jarak data yang paling mendekati centroid. Rumus *Euclidian Distance* yaitu :

$$d(i, j) = \sqrt{\sum_j^m (C_{ij} - C_{kj})^2}$$

keterangan :

xi = data kriteria

µj = Centroid Pada Cluster ke-j

- Kumpulkan dari masing-masing data pada *cluster* dengan jarak terpendek.

$$\text{Min} \sum_{k=1}^k \text{dik} = \sqrt{\sum_j^m (C_{1j} - C_{kj})^2}$$

- d. Update nilai centroid data dengan cara menghitung nilai rata-rata cluster.

$$C_{kj} = \frac{\sum_{i=1}^p x_{ij}}{p}$$

- e. Ulangi langkah 1, 2, 3 hingga seluruh anggota cluster hingga tidak mengalami perubahan.
- f. Tentukan nilai sentral cluster terakhir yang mana hal ini dapat memastikan kesesuaian data dengan syarat pada langkah 1, 2, dan 3 telah dikerjakan dan memperoleh nilai yang sesuai.

**2.3. Penelitian Terdahulu**

Paper (Lion Ferdinand Marin et al, 2023) tentang pengelompokan tempat wisata di Kabupaten Teluk Wondama. Riset dilakukan karena jarak yang cukup jauh antar tempat wisata di Kabupaten Teluk Wondama, mencapai 1-2 jam perjalanan antar destinasi. Tujuan pengelompokan ini adalah untuk mempermudah pemerintah kabupaten dalam meningkatkan kualitas tempat wisata. Hasil analisis klaster menunjukkan bahwa terdapat dua klaster utama. Klaster pertama, disusun oleh distrik Kuri Wamesa, Raisey, Roswar, Wasior, dan Wondiboy, memiliki jumlah kunjungan yang lebih sedikit. Sementara klaster kedua, terdiri dari distrik Teluk Duairi, Roon, dan Windesi, menunjukkan jumlah kunjungan yang lebih tinggi [7].

Paper (Edy Satria et al., 2019) membahas Saat ini, sektor pariwisata Indonesia berkontribusi sekitar 4% dari total perekonomian. Pada tahun 2019, Pemerintah Indonesia memiliki target untuk meningkatkan kontribusi sektor pariwisata menjadi 8% dari Produk Domestik Bruto (PDB). Target ini mengimplikasikan bahwa dalam waktu empat tahun, jumlah pengunjung perlu ditingkatkan dua kali lipat menjadi sekitar 20 juta wisatawan. Penelitian ini membahas implementasi teknik clustering, khususnya menggunakan metode K-Means, dalam mengelompokkan jumlah kunjungan wisatawan mancanegara berdasarkan kebangsaan dan bulan kedatangan. Sumber data penelitian berasal dari Badan Pusat Statistik Nasional yang mencatat jumlah kunjungan wisatawan mancanegara. Hasil penelitian menunjukkan bahwa pada tahun 2017 hingga 2018, terdapat dua kluster utama. Cluster 1 (klaster tinggi) mencakup empat negara, yaitu China, Malaysia, Singapura, dan Timor Leste. Sementara itu, Cluster 2 (klaster rendah) melibatkan 87 negara lainnya [8].

Paper (Naning Savitri et al., 2021) membahas mengenai klasifikasi kunjungan wisatawan pada berbagai destinasi wisata di Kulon Progo, bertujuan untuk memajukan objek wisata yang memiliki tingkat kunjungan beragam. Proses pengelompokan dilakukan dengan menggunakan algoritma K-Means, dan hasilnya menghasilkan tiga kelompok. Kelompok pertama terdiri dari 20 objek wisata dengan jumlah pengunjung yang rendah, kelompok kedua melibatkan

15 objek wisata dengan tingkat kunjungan yang sedang, sementara kelompok ketiga mencakup dua objek wisata dengan tingkat kunjungan yang tinggi [9].

Paper (Bambang Setio Purnomo et al, 2021) membahas mengenai penerapan data mining untuk mengelompokkan kunjungan wisatawan di Kota Yogyakarta, penelitian ini mengevaluasi jumlah wisatawan yang datang berdasarkan lokasi wisata di Yogyakarta. Penelitian ini menggunakan metode algoritma K-Means dengan desain teknik yang mengacu pada Knowledge Discovery in Databases (KDD), dan hasilnya adalah terbentuk tiga kelompok [10].

Paper (Suastika Yulia Riska et al, 2023) membahas mengenai pengelompokan negara-negara berdasarkan jumlah wisatawan mancanegara yang datang ke Indonesia, dimulai dari yang paling banyak hingga yang paling sedikit. Hasil pengelompokan ini dapat menjadi panduan bagi pemerintah untuk merancang strategi yang efektif dalam meningkatkan ketertarikan wisatawan untuk mengunjungi Indonesia. Algoritma yang diterapkan dalam proses Clustering adalah K-Means dan K-Medoids, dengan menerapkan nilai k=2, k=3, dan k=5. Proses evaluasi digunakan metode davies bouldin index. Dengan menggunakan Algoritma K-Medoids, hasil cluster terbaik adalah dengan menggunakan k=2 dengan nilai davies bouldin index -0,567. Hasil cluster terbaik dalam kasus ini adalah dengan menggunakan Algoritma K-Means dengan k=5, dan dengan hasil nilai davies bouldin index - 0.302 [11].

**3. METODE PENELITIAN**

Metode penelitian yang digunakan dalam penelitian ini adalah metode penelitian kuantitatif. Metode kuantitatif adalah metode penelitian yang menggunakan proses data-data yang berupa angka sebagai alat menganalisis dan melakukan kajian penelitian, terutama mengenai apa yang sudah diteliti . Penelitian ini akan dilakukan dalam beberapa tahapan dan langkah-langkah penelitian seperti gambar di bawah ini :



Gambar 1. Tahapan Metode Penelitian

Tahapan penerapan algoritma K-Means pada data Daya Tarik Obyek Wisata Berdasarkan Jenisnya di Jawa Barat terdiri dari beberapa langkah. Berikut adalah tahapan-tahapan tersebut :

- a. Pengumpulan Data  
Tahap awal adalah mengumpulkan data Daya Tarik Obyek Wisata di Jawa Barat. Data ini dapat mencakup berbagai informasi seperti nama\_provinsi, nama\_kabupaten, alam, buatan, budaya, satuan dan tahun. Data ini harus dalam format yang sesuai dan siap untuk diproses.
- b. Pemrosesan Data  
Data yang telah dikumpulkan mungkin memerlukan pemrosesan tambahan, seperti membersihkan data dari nilai-nilai kosong atau data yang tidak valid, serta melakukan normalisasi data jika variabel-variabel memiliki skala yang berbeda.
- c. Pemilihan Jumlah Kluster (K)  
Langkah selanjutnya adalah memilih jumlah kluster yang akan dibentuk oleh algoritma K-Means. Jumlah kluster ini harus dipilih dengan hati-hati, karena dapat mempengaruhi hasil analisis.
- d. Inisialisasi Pusat Kluster dan Iterasi K-Means  
Pada awalnya, titik-titik pusat kluster dipilih secara acak dari data. Posisi pusat kluster ini akan digunakan sebagai titik awal dalam proses iteratif algoritma K-Means.
- e. Evaluasi Kluster  
Setelah proses iterasi selesai, langkah selanjutnya adalah mengevaluasi hasil kluster yang ditemukan. Analisis visual dan metode evaluasi lainnya dapat digunakan untuk memahami dan menginterpretasi setiap kluster yang dihasilkan.

**4. HASIL DAN PEMBAHASAN**

**4.1. Hasil Penelitian**

Penelitian ini menghasilkan beberapa temuan yang penting dan memberikan wawasan mengenai klusterisasi data dalam konteks penelitian ini. Penelitian ini menggunakan metode analisis dengan tahapan *Data Selection, Data Preprocessing, Data Transformation, Data Mining, Evaluation dan Knowledge*. Hasil clustering ini memberikan gambaran yang lebih jelas tentang pola dan relasi diantara data sebelumnya kompleks.

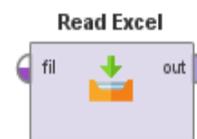
**4.1.1. Data**

Data yang digunakan penelitian ini diperoleh melalui sumber Open Data Jabar dengan format tampilan excel yang terstruktur dengan judul Data Daya Tarik Obyek Wisata Tahun 2014-2022. Data ini terdiri dari 726 dataset dan memiliki 7 atribut yaitu Nama Provinsi, Nama, Kabupaten/Kota, Alam, Buatan, Budaya, Satuan, dan Tahun beserta penjelasannya dapat dilihat pada Tabel 1 berikut.

Tabel 1. Daya Tarik Obyek Wisata

No	Tipe	Atribut	Keterangan
1	Nama Provinsi	Polynomial	Untuk atribut dengan lebih dari satu kategori
2	Nama Kabupaten/Kota	Polynomial	Untuk atribut dengan lebih dari satu kategori
3	Alam	Integer	Untuk file dengan nilai integer (angka tanpa koma)
4	Buatan	Integer	Untuk file dengan nilai integer (angka tanpa koma)
5	Budaya	Integer	Untuk file dengan nilai integer (angka tanpa koma)
6	Satuan	Integer	Untuk file dengan nilai integer (angka tanpa koma)
7	Tahun	Polinomial	Untuk atribut dengan lebih dari satu kategori

Dari Tabel 1, pada aplikasi RapidMiner untuk membaca data tersebut menggunakan operator Read Excel. Dapat dilihat pada Gambar 2.



Gambar 2. Operator Read Excel Pada RapidMiner

Operator Read Excel dalam RapidMiner digunakan untuk mengimpor data dari file Excel kedalam lingkungan RapidMiner. Dengan menggunakan operator ini, dapat memuat file Excel yang berisi data yang dipilih.

**4.1.2. Data Selection**

Tahapan selection digunakan untuk menyeleksi atau memilih data yang akan diolah. Data yang diolah yaitu data Daya Tarik Obyek Wisata. Tahapan ini dilakukan untuk memilih data yang akan diproses di RapidMiner, tahapan ini dilakukan di Microsoft Excel dan di RapidMiner. Pada Microsoft Excel menghapus data yang bernilai null, dan memilih tahun dari 2022. Pada RapidMiner menggunakan operator *Set Role* untuk menentukan id pada dataset, dapat dilihat pada Gambar 3.



Gambar 3. Operator Set Role Pada RapidMiner

Parameter pada operator *Set Role* yang digunakan dapat dilihat pada Gambar 3 dengan menggunakan atribut *name* yaitu Nama Kabupaten/Kota menjadi target role yaitu *id* dapat dilihat pada Tabel 2 berikut.

Tabel 2. Parameter Operator Set Role

No.	Parameter	Isi
1.	<i>attribute name</i>	<i>Nama Kabupaten/Kota</i>
2.	<i>target role</i>	<i>Id</i>

Dari hasil pembacaan operator *Set Role* didapat informasi pada Tabel 3 berikut.

Tabel 3. Hasil Pembacaan Operator Set Role

No	Uraian	Keterangan
1	<i>Examples</i>	27
2	<i>Special Attributes</i>	1
3	<i>Regular Attributes</i>	6
	Nama Provinsi	<i>Polynomial</i>
	Alam	<i>Integer</i>
	Buatan	<i>Integer</i>
	Budaya	<i>Integer</i>
	Satuan	<i>Polynomial</i>
	Tahun	<i>Polynomial</i>

#### 4.1.3. Data Preprocessing

Sebelum melaksanakan Data Mining, langkah awal yang penting adalah menjalani serangkaian proses, termasuk membersihkan duplikasi data, memeriksa inkonsistensi dalam data, dan memperbaiki kesalahan yang mungkin muncul, seperti kesalahan cetak (*tipografi*). Selain itu, proses *enrichment* data dilakukan untuk memperkaya data yang telah ada dengan menambahkan informasi tambahan yang relevan dan diperlukan dalam Kontak Penemuan Pengetahuan (KDD), termasuk data atau *informers eksternal* yang dapat meningkatkan kualitas dan keberagaman data. Dalam tahap ini diperoleh data sebanyak 27 *record* pada tahun 2022 dengan 7 atribut yang akan digunakan yaitu Nama Provinsi, Nama, Kabupaten/Kota, Alam, Buatan, Budaya, Satuan, Tahun. Berikut adalah proses *processing* pada *RapidMiner* menggunakan operator *Select Attributes* dapat dilihat pada Gambar 4 berikut.



Gambar 4. Operator *Select Attributes* pada RapidMiner

Operator *Select Attributes* dalam *RapidMiner* digunakan untuk memilih, menghapus, atau mengatur atribut dalam dataset. Ini memungkinkan mengubah tipe data, dan melakukan transformasi lainnya pada atribut sebelum proses analisis data. Parameter yang digunakan pada operator *Select Attributes* dapat dilihat pada Tabel 4 berikut.

Tabel 4. Parameter pada Operator *Select Attributes*

No	Parameter	Keterangan
1	<i>Type</i>	<i>include attributes</i>
2	<i>attribute filter type</i>	<i>a subset</i>
3	<i>attribute</i>	Nama Kabupaten/Kota
		Alam
		Buatan
		Budaya

Hasil dari penggunaan operator *Select Attributes* diperoleh informasi pada Tabel 5 berikut.

Tabel 5. Hasil Operator *Select Attributes*

No	Nama Atribut	Jenis Data	Missing
1	Nama Kabupaten/Kota	<i>Polynomial</i>	0
2	Alam	<i>Integer</i>	0
3	Buatan	<i>Integer</i>	0
4	Budaya	<i>Integer</i>	0

Dari hasil result dari statistik dataset 4 atribut seperti tampak pada Tabel 5, diketahui bahwa tidak ada atribut yang memiliki nilai *missing*. Untuk memeriksa konsisten atau tidak konsistennya dataset yang digunakan diperiksa per-*record* secara langsung dan menunjukkan bahwa dataset memiliki data yang konsisten terhadap nilainya.

#### 4.1.4. Data Transformation

Pada penelitian ini data yang digunakan adalah data yang relevan dan konsisten. Setelah dilakukan pengecekan pada hasil statistik data, terlihat bahwa data dalam penelitian ini sudah bernilai angka. Maka dari itu, langkah *Data Transformation* tidak dilakukan.

NAMA KABUPATEN/KOTA	Jenis	Min	Max	Jumlah
NAMA KABUPATEN/KOTA	Polynomial	0	1	27
ALAM	Integer	0	1	182
BUATAN	Integer	0	2	71
BUDAYA	Integer	0	1	175

Gambar 5. Review Statistic Data

#### 4.1.5. Data Mining

Langkah ini berfungsi untuk menentukan metode atau teknik yang paling sesuai dalam mengidentifikasi pola atau informasi menarik dalam data yang telah dipilih. Dalam penelitian ini, proses data mining diimplementasikan dengan menggunakan metode Clustering, dengan memanfaatkan Algoritma K-Means. Pada tahap ini juga peneliti melakukan proses clustering, menguji dan mengevaluasi hasil dari proses tersebut dalam platform *RapidMiner*. Operator yang digunakan adalah operator Clustering (K-Means) dapat dilihat pada Gambar 6 berikut.



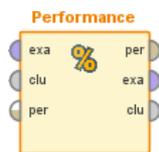
Gambar 6. Operator Clustering pada RapidMiner

Parameter yang digunakan pada k-Means atau Clustering dapat dilihat pada Tabel 6 berikut.

Tabel 6. Parameter pada Operator Clustering

No.	Parameter	Isi
1	<i>add cluster attribute</i>	Digunakan
2	<i>k</i>	4
3	<i>max runs</i>	10
4	<i>determine good star values</i>	Digunakan
5	<i>measure types</i>	<i>Bregman Divergences</i>
6	<i>divergence</i>	<i>Squared Euclidean Distance</i>
7	<i>max optimization steps</i>	100

Pada saat menggunakan operator clustering kita perlu mengetahui berapa kluster yang terbaik dalam dataset ini, lalu dalam penelitian ini menggunakan performance Davies Bouldin Index (DBI) dapat dilihat pada Gambar 7 berikut.



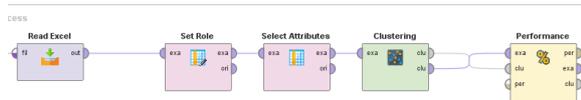
Gambar 7. Operator Performance untuk Davies Bouldin Index pada RapidMiner

Parameter yang digunakan pada operator Performance dapat dilihat pada Tabel 7 berikut.

Tabel 7. Parameter pada Operator Performance

No.	Parameter	Isi
1	<i>Main criterion</i>	<i>Davies Bouldin</i>
2	<i>Normalize</i>	Digunakan
3	<i>Maximize</i>	Digunakan

Pada operator *Performance*, parameter *normalize* dan *maximize* digunakan karena jika hanya mencentang *normalize* saja akan menjadi minus, dan menggunakan atau mencentang *maximize* hanya menghilangkan minusnya saja tidak merubah *cluster* atau lainnya. Tahapan pada clustering dapat dilihat pada Gambar 8 berikut.



Gambar 8. Tahapan Clustering pada RapidMiner

4.1.6. Evaluation

Teori evaluasi yang digunakan dalam penelitian ini adalah *Davies Bouldin Index* (DBI).

Dalam Tabel 8, terdapat jumlah cluster yang telah dimodelkan dan nilai DBI yang terkait.

Tabel 8. Nilai DBI Setiap Jumlah Cluster

Cluster	Nilai DBI
K2	0.252
K3	0.165
K4	0.146
K5	0.228
K6	0.171
K7	0.186
K8	0.178
K9	0.235
K10	0.173

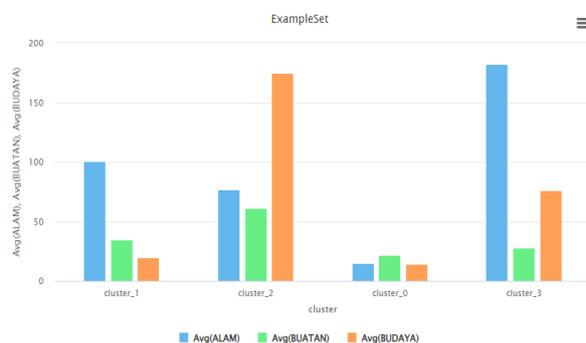
Mengacu pada prinsip DBI, nilai yang dianggap baik dalam data mining adalah semakin kecil atau mendekati nol. Dalam Tabel 8 nilai DBI yang paling baik adalah *cluster 4*, dengan nilai DBI yang mendekati nol yaitu 0.146. Oleh karena itu, pengelompokan dilakukan dengan menggunakan 4 *cluster*.

4.2. Pembahasan

Penelitian ini dilakukan dengan tujuan utama untuk mengelompokkan dataset Daya Tarik Obyek Wisata menggunakan algoritma K-Means. Tujuan spesifik mengidentifikasi pola atau kesamaan berdasarkan atribut-atribut terpilih serta mengevaluasi struktur dan jarak antar cluster yang dihasilkan. Hasil pengelompokan data Daya Tarik Obyek Wisata menggunakan algoritma K-Means telah diinterpretasikan dengan mempertimbangkan temuan yang sesuai dengan konsep, teori, dan penelitian terdahulu yang relevan dengan lingkup penelitian ini. Interpretasi ini memberikan wawasan yang mendalam tentang struktur kluster dan signifikansi temuan yang dihasilkan.

4.2.1. Pengelompokkan Menggunakan Teknik Clustering

Penelitian ini berhasil mengimplementasikan teknik *clustering*, dari proses tersebut menghasilkan visualisasi data dalam bentuk histogram pada Gambar 4.8 berikut.



Gambar 9. Cluster Visualisasi Histogram

Bentuk histogram yang ditunjukkan pada Gambar 4.8 ini merupakan visualiasasi dari hasil pemodelan cluster pada dataset daya tarik obyek wisata dengan

deskripsi histogram berwarna biru adalah alam, histogram berwarna hijau adalah buatan dan histogram berwarna jingga adalah budaya. Dari pengujian diatas dapat disimpulkan bahwa cluster 0 memiliki nilai rata-rata Alam sebesar 14,938, Buatan sebesar 21,438 dan Budaya sebesar 14,188. Cluster 1 memiliki nilai rata-rata Alam sebesar 100,889, Buatan sebesar 34,889 dan Budaya sebesar 19,556. Cluster 2 memiliki nilai rata-rata Alam sebesar 77, Buatan sebesar 61 dan Budaya sebesar 175. Cluster 3 memiliki nilai rata-rata Alam sebesar 182, Buatan sebesar 28 dan Budaya sebesar 76.

**4.2.2. Klaster yang dihasilkan oleh Algoritma K-Means**

Algoritma K-Means Clustering dapat digunakan untuk mengelompokkan data Daya Tarik Obyek Wisata dengan atribut Nama Provinsi, Nama, Kabupaten/Kota, Alam, Buatan, Budaya, Satuan, Tahun dengan hasil Cluster Model dan jumlah anggota setiap cluster dapat dilihat pada Gambar 10. berikut.

**Cluster Model**

```
Cluster 0: 16 items
Cluster 1: 9 items
Cluster 2: 1 items
Cluster 3: 1 items
Total number of items: 27
```

Gambar 10. Hasil Jumlah Item dari 4 Cluster

Dari Gambar 10, menunjukkan bahwa data telah berhasil dikelompokkan menjadi 4 klaster yaitu Cluster 0 dengan jumlah 16 items, Cluster 1 dengan jumlah 9 items, Cluster 3 dengan jumlah 1 items, dengan total jumlah item adalah 27. Jumlah item yang berbeda menunjukkan adanya variasi dalam sifat atau karakteristik tersebut. Cluster dengan jumlah yang lebih tinggi menunjukkan adanya lebih banyak daya tarik obyek wisata. Selanjutnya analisis nilai rata-rata centroid pada setiap cluster dari atribut yang telah ditentukan. Hasil dari nilai rata-rata atribut dapat dilihat dari Gambar 11 berikut.

Atribut	cluster_0	cluster_1	cluster_2	cluster_3
ALAM	14.938	100.889	77	182
BUATAN	21.438	34.889	61	28
BUDAYA	14.188	19.556	175	76

Gambar 11. Hasil Rata-rata Centroid dari Setiap Cluster

Dari Gambar 4.9 dan 4.10 dapat disimpulkan bahwa setiap cluster berdasarkan Daya Tarik Obyek Wisata Pada Pemprov Jabar dapat dianalisis sebagai berikut :

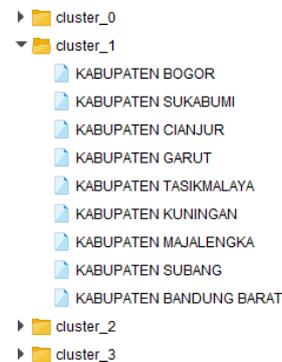
- a. *Cluster 0* merupakan Tingkat Daya Tarik Obyek Wisata Rendah dengan anggota cluster berjumlah 16 kabupaten yang meliputi Kabupaten Ciamis, Kabupaten Cirebon, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kota

Sukabumi, Kota Bandung, Kota Cirebon, kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar, dan Kabupaten Pangandaran.



Gambar 12. Anggota Cluster 0

- b. *Cluster 1* merupakan Tingkat Daya Tarik Obyek Wisata sedang dengan anggota cluster berjumlah 9 kabupaten yang meliputi Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Kuningan, Kabupaten Majalengka, Kabupaten Subang, dan Kabupaten Bandung Barat.



Gambar 13. Anggota Cluster 1

- c. *Cluster 2* merupakan Tingkat Daya Tarik Obyek Wisata tinggi dengan anggota cluster berjumlah 1 kabupaten yaitu Kabupaten Bandung.



Gambar 14. Anggota Cluster 2

- d. *Cluster 3* merupakan Tingkat Daya Tarik Obyek Wisata tinggi dengan anggota cluster berjumlah 1 kota yaitu Kota Bogor.



Gambar 15. Anggota Cluster 3

**4.2.3. Jarak Antar Klaster Terhadap Titik Pusat Klaster**

Selama penelitian, jarak antar klaster diukur dari titik pusat klaster terdekat dan terjauh untuk memahami distribusi atau perbedaan antar klaster berdasarkan atribut-atribut yang dianalisis dalam dataset Daya Tarik Obyek Wisata. Jumlah rata-rata dalam jarak centroid setiap cluster pada tabel 4.10.

Tabel 9. Jumlah rata-rata dalam Jarak Centroid setiap Cluster

Cluster	Rata-rata Jarak Centroid
K Pusat	246.288
K0	207.152
K1	370.593
K2	0.000
K3	0.000

**5. KESIMPULAN DAN SARAN**

Dari analisis data yang telah dilakukan, diperoleh hasil dan kesimpulan bahwa penggunaan algoritma K-Means dan metode Knowledge Discovery in Database (KDD) dalam pemodelan dan evaluasi menggunakan nilai Davies Bouldin Index (DBI) dapat mengoptimalkan jumlah klaster didapat klaster terbaik yaitu pada cluster 8 dengan nilai DBI 0.094, dengan deskripsi sebagai berikut : 1. Clustering atau perolehan hasil cluster ini menghasilkan pengelompokan daya tarik obyek wisata yang berdasarkan Nama Provinsi, Nama, Kabupaten/Kota, Alam, Buatan, Budaya, Satuan, Tahun dengan kategori rendah, sedang dan tinggi. Penjelasan dari 3 kategori tersebut dapat dijelaskan sebagai berikut : Cluster 0 merupakan Tingkat Daya Tarik Obyek Wisata Rendah dengan anggota cluster berjumlah 16 kabupaten yang meliputi Kabupaten Ciamis, Kabupaten Cirebon, Kabupaten Sumedang, Kabupaten Indramayu, Kabupaten Purwakarta, Kabupaten Karawang, Kabupaten Bekasi, Kota Sukabumi, Kota Bandung, Kota Cirebon, kota Bekasi, Kota Depok, Kota Cimahi, Kota Tasikmalaya, Kota Banjar, dan Kabupaten Pangandaran. Cluster 1 merupakan Tingkat Daya Tarik Obyek Wisata sedang dengan anggota cluster berjumlah 9 kabupaten yang meliputi Kabupaten Bogor, Kabupaten Sukabumi, Kabupaten Cianjur, Kabupaten Garut, Kabupaten Tasikmalaya, Kabupaten Kuningan, Kabupaten Majalengka, Kabupaten Subang, dan Kabupaten Bandung Barat. Cluster 2 merupakan Tingkat Daya Tarik Obyek Wisata tinggi dengan anggota cluster berjumlah 1 kabupaten yaitu Kabupaten Bandung. Cluster 3 merupakan Tingkat Daya Tarik Obyek Wisata tinggi dengan anggota cluster berjumlah 1 kota yaitu Kota Bogor. 2. Perfoma Cluster yang diperoleh

dari Analisis Pengelompokan Daya Tarik Obyek Wisata Berdasarkan Jenisnya Menggunakan Metode K-Means Pada Data Pemprov Jabar memiliki nilai rata-rata atribut sebesar Avg. within centroid distance: 246.288, Cluster 0=207.152, Cluster 1=370.593, Cluster 2=0.000, dan Cluster 3=0.000.

Hasil penelitian ini, peneliti mengharapkan dapat digunakan pada penelitian-penelitian selanjutnya, dan dengan pembagian data menggunakan clustering ini diharapkan dapat membantu pengguna dalam penelitian Obyek Wisata di Provinsi Jawa Barat. Dikarenakan keterbatasan waktu dan energi, peneliti menyadari bahwa hasil penelitian ini masih jauh dari kesempurnaan. Oleh karena itu, dalam penelitian selanjutnya, dapat dilakukan perbandingan dengan metode clustering yang berbeda.

**DAFTAR PUSTAKA**

[1] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means," *JISKA (Jurnal Informatika Sunan Kalijaga)*, vol. 2, no. 3, p. 167, 2018, doi: 10.14421/jiska.2018.23-06.

[2] B. M. Al-Fahmi, E. Rahmawati, and T. Sagirani, "Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 2, pp. 141–149, 2023, doi: 10.25077/teknosi.v9i2.2023.141-149.

[3] D. Triyansyah and D. Fitrihanah, "Analisis Data Mining Menggunakan Algoritma K-Means Clustering Untuk Menentukan Strategi Marketing," *Jurnal Telekomunikasi dan Komputer*, vol. 8, 2018, doi: 10.22441/incomtech.v8i2.4174.

[4] N. Azwanti, "Segmentasi Tingkat Pemakaian Material dengan Data Mining Clustering," *Jurnal Komputer Terapan*, vol. 4, no. 2, 2018, [Online]. Available: <http://jurnal.pcr.ac.id>

[5] A. P. Disi and B. Sumargo, "PENGELOMPOKAN PENGGUNA INTERNET DENGAN METODE K-MEANS CLUSTERING," *Jurnal Statistika dan Aplikasinya (JSA)*, vol. 3, no. 1, 2019.

[6] E. D. S. Mulyani *et al.*, "Clustering Wilayah Dan Pelanggaran Berkendaraan Menggunakan Algoritma K-Means Pada Data Satlantas Polres Tasikmalaya Kota," vol. 8, no. 1, 2019.

[7] L. F. Marini and C. D. Suhendra, "Penggunaan Algoritma K-Means Pada Aplikasi Pemetaan Klaster Daerah Pariwisata," *Jurnal Media Informatika Budidarma*, vol. 7, no. 2, pp. 707–713, 2023, doi: 10.30865/mib.v7i2.5558.

[8] E. Satria *et al.*, "Prosiding Seminar Nasional Riset Information Science (SENARIS) Penerapan Clustering dalam Mengelompokkan Jumlah Kunjungan Wisatawan Mancanegara Dengan Metode K-Means," no. September, p.

- 462, 2019, [Online]. Available: <https://www.bps.go.id/dynamic/table/2018/07/30/1548/jumlah-kunjungan->
- [9] S. Naning, R. Pranata, A. N. M. Clara, and O. S. Rahajeng, "Pengelompokan Kunjungan Wisata Kabupaten Kulon Progo Tahun 2019 Menggunakan K-Means Clustering," *JUST IT: Jurnal Sistem ...*, vol. 12, no. 1, pp. 38–45, 2022, [Online]. Available: <https://jurnal.umj.ac.id/index.php/just-it/article/view/8249%0Ahttps://jurnal.umj.ac.id/index.php/just-it/article/download/8249/7242>
- [10] B. Setio and P. Prasetyaningrum, "Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means," *Journal of Computer Science and Technology (JCS-TECH)*, vol. 1, no. 1, pp. 27–32, 2021, doi: 10.54840/jcstech.v1i1.9.
- [11] S. Y. Riska and L. Farokhah, "Perbandingan Hasil Evaluasi Algoritma K-Means dan K-Medoid Berdasarkan Kunjungan Wisatawan Mancanegara ke Indonesia," *Journal of Information Technology*, vol. 8, no. 1, p. hal. 1-8, 2023.