

DATA MINING CLUSTERING MENGGUNAKAN ALGORITMA K-MEANS PADA DATA KUNJUNGAN WISATAWAN DI KABUPATEN KARAWANG

Kamaludin Gustipartsani¹, Nining Rahaningsih², Raditya Danar Dana³, Iva Yulia Mustafa⁴

¹ Program Studi Teknik Informatika, STMIK IKMI Cirebon

² Program Studi Komputerisasi Akuntansi, STMIK IKMI Cirebon

³ Program Studi Manajemen Informatika, STMIK IKMI Cirebon

⁴ Program Studi Perhotelan, Politeknik Pariwisata Prima Internasional

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

gustipartsani@gmail.com

ABSTRAK

Karawang merupakan sebuah Kabupaten di Provinsi Jawa Barat yang memiliki cukup banyak wisata di daerahnya. Dari berbagai tempat wisata yang ada di setiap Kecamatan di Kabupaten Karawang terdapat data jumlah kunjungan wisatawan setiap tahunnya. Pada data yang didapat dari situs web opendata.jabarprov.go.id masih terdapat banyak objek wisata yang sepi pengunjung. Hal tersebut dapat terjadi karena pengembangan wisata di Kabupaten Karawang masih kurang merata. Dalam hal ini, maka akan dilakukan pengelompokan wisata yang ditujukan untuk dasar pengambilan keputusan terkait peningkatan pengelolaan objek wisata. Dengan menggunakan *Knowledge Discovery in Database (KDD)*, penelitian ini akan menerapkan data mining *clustering* menggunakan algoritma *K-Means*. Implementasi data mining dilakukan dengan menggunakan tools *RapidMiner*. Analisa *clustering* bertujuan untuk mencari jumlah *k* yang optimal untuk pengelompokan objek wisata. Hasil dari penelitian ini diperoleh jumlah *k* yang optimal adalah 3 dengan kelompok 1 sebagai kelompok jumlah wisatawan rendah, kelompok 2 sebagai kelompok jumlah wisatawan sedang, dan kelompok 3 sebagai kelompok jumlah wisatawan tinggi.

Kata kunci: Data Mining, Clustering, K-Means, Pariwisata, Knowledge Discovery in Database

1. PENDAHULUAN

Kemajuan dalam bidang ilmu pengetahuan dan teknologi telah menghasilkan sejumlah perubahan dalam pola hidup masyarakat [1]. Dalam perkembangan teknologi informasi yang pesat, keterjangkauan akan informasi yang disediakan sangatlah penting [2]. Kemajuan pada bidang teknologi informasi berpengaruh penting pada beberapa sektor, salah satunya yaitu sektor pariwisata. Kabupaten Karawang yang dikenal dengan kota industri juga memiliki berbagai macam objek wisata budaya di daerahnya.[3]. Selain objek wisata budaya terdapat berbagai objek wisata alam di Kabupaten Karawang, maka dari itu untuk mendukung kemajuan sektor pariwisata di Kabupaten Karawang maka diperlukan publikasi terkait informasi objek wisata di Karawang [4].

Pada sektor wisata yang ada di Kabupaten Karawang masih terdapat beberapa tantangan yang harus dihadapi salah satunya adalah peningkatan pengelolaan objek wisata. Berdasarkan data yang diperoleh melalui website opendata.jabarprov.go.id terkait jumlah kunjungan wisatawan di Kabupaten Karawang, masih terdapat beberapa objek wisata yang memiliki jumlah wisatawan rendah tiap tahunnya. Berdasarkan data tersebut, ditemukan permasalahan seperti bagaimana cara meningkatkan pengelolaan pariwisata berdasarkan jumlah kunjungan wisatawan pada objek wisata. Untuk setiap strategi pembangunan dan peningkatan pariwisata tentunya berbeda berdasarkan keramaian pengunjung wisatawan. Maka dari itu pengelompokan pada jenis pariwisata

di Kabupaten Karawang berdasarkan jumlah kunjungan wisatawan akan berpengaruh untuk strategi peningkatan pengelolaan objek wisata. Pada konteks informatika, permasalahan yang dihadapi adalah bagaimana cara mengelola data jumlah kunjungan wisatawan pada objek wisata yang ada di Kabupaten Karawang untuk meningkatkan pengelolaan objek wisata. Data mining *clustering* dengan algoritma *K-Means* menjadi salah satu pilihan untuk menghadapi tantangan pada pengelolaan data kunjungan wisatawan pada objek wisata di Kabupaten Karawang.

Beberapa penelitian sebelumnya telah menerapkan terkait topik analisa *clustering* menggunakan algoritma *K-Means* untuk pengelompokan. Penelitian [5] telah menerapkan data mining *clustering* menggunakan metode *CRISP-DM* dengan algoritma *K-Means*. Penelitian tersebut menggunakan tools *Python* dan menghasilkan jumlah *cluster* sebanyak 3 *cluster* dengan pencarian *k* optimal menggunakan metode *Elbow*. Lalu penelitian berikutnya [6] menggunakan *clustering K-Means* untuk aplikasi pemetaan klaster, penelitian tersebut menggunakan metode *Elbow* dan metode *Silhouette Coefficient* dengan percobaan sebanyak 2 kali yaitu *k=2* dan *k=3* menghasilkan masing masing *k* optimal dengan metode *Elbow* memperoleh *k* optimal *k=3* dan metode *Silhouette Coefficient* memperoleh *k* optimal *k=2*. Penelitian selanjutnya [7] menerapkan data mining *clustering K-Means* dengan metode *KDD* menggunakan tools *WEKA 3.8.3 (waikato environment for knowledge analysis)* menghasilkan

cluster sebanyak 3 cluster. Dengan cluster 1 memiliki rata-rata persentase 50% kunjungan wisatawan, cluster 2 23% kunjungan wisatawan, dan cluster 3 27% kunjungan wisatawan.

Tujuan utama pada penelitian ini adalah mengelompokkan data jumlah kunjungan wisatawan pada objek wisata di Kabupaten Karawang menggunakan analisis *clustering* dengan algoritma *K-Means*. Tujuan dilakukan analisis *clustering* adalah untuk menemukan informasi baru terkait data kunjungan wisatawan di Kabupaten Karawang.

Pada penelitian ini metode yang digunakan adalah *Knowledge Discovery in Database (KDD)* menggunakan algoritma *K-Means Clustering*. Data mining yang sering juga disebut sebagai *Knowledge Discovery in Database (KDD)*, adalah proses yang mencakup pengumpulan dan pemanfaatan data historis untuk mengidentifikasi keteraturan serta pola hubungan dalam kumpulan data yang berskala besar [8]. Proses pada tahapan *Knowledge Discovery in Database (KDD)* terdiri dari beberapa tahapan yaitu pengumpulan data, pemakaian data historis yang memiliki tujuan menemukan pola atau korelasi pada data yang memiliki skala besar [9]. *Knowledge Discovery in Database (KDD)* adalah suatu langkah untuk dapat menemukan serta mengenali pola dalam data, dimana pola yang dapat ditemukan tersebut bersifat valid, memiliki keterbaruan, bermanfaat, dan dapat dimengerti [10].

Hasil penelitian ini diharapkan dapat memberikan wawasan terkait pengolahan data kepada para praktisi wisata. Pengelola objek wisata dapat secara langsung memanfaatkan hasil penelitian ini untuk merancang strategi pemasaran yang lebih terarah, meningkatkan pengalaman pengunjung, dan mengoptimalkan sumber daya. Selain manfaat praktis, penelitian ini membuka peluang untuk penelitian lebih lanjut di bidang Informatika terkait analisis data pariwisata, mendukung pengembangan model analisis yang lebih canggih. Dari segi teknologi, implementasi temuan ini dapat mendukung pengembangan sistem informasi pariwisata dan aplikasi berbasis data, meningkatkan efisiensi pengelolaan pariwisata, serta mempromosikan pertumbuhan sektor pariwisata secara keseluruhan. Kesuksesan penelitian ini bukan hanya memberikan manfaat praktis, tetapi juga melengkapi pemahaman tentang penerapan metode analisis data dalam pengelolaan destinasi pariwisata, berkontribusi pada perkembangan ilmu Informatika.

2. TINJAUAN PUSTAKA

2.1. Penelitian Terkait

Beberapa kajian sebelumnya terkait dengan analisa *clustering* menggunakan *K-Means* yaitu [11], yang membahas terkait strategi pembangunan infrastruktur dan menghasilkan *cluster* berjumlah 4 dengan cluster 1 berjumlah 2 kecamatan dengan kriteria rendah, cluster 2 berjumlah 4 kecamatan dengan kriteria rata-rata, cluster 3 berjumlah 3 kecamatan dengan kriteria tinggi, dan cluster 4 yang

berjumlah 2 kecamatan dengan kriteria sangat tinggi. Dengan kesimpulan pembangunan infrastruktur dapat dilakukan pada daerah dengan kriteria tinggi yakni C3 dan C4.

Penelitian lainnya yaitu [12], membahas terkait pengelompokkan wisatawan mancanegara berdasarkan asal negara dan bulan kunjungan yang menghasilkan *cluster* sebanyak 2 dengan cluster 1 yang terdiri dari 4 negara dan cluster 2 sebanyak 84 negara.

2.2. Data Mining

Data mining Data mining merupakan langkah-langkah yang dilakukan untuk mengambil informasi atau pengetahuan dari data yang sebelumnya tidak diketahui, tersirat, dan dianggap tidak berguna, yang berasal dari sejumlah besar data [13]. Langkah-langkah tersebut digunakan untuk mengidentifikasi dan memahami pola data sehingga menghasilkan informasi yang dibutuhkan [14]. Hasil dari data mining tersebut juga dapat digunakan untuk sebuah dasar pengambilan keputusan maupun kebijakan bagi sebuah organisasi pemerintah, maupun instansi terkait topik yang diteliti [15].

2.3. Clustering

Clustering merupakan suatu algoritma yang termasuk ke dalam kategori data unsupervised karena dapat mengklasifikasikan data tanpa label berdasarkan kemiripan datanya [16]. Clustering adalah di mana data dikelompokkan ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang tinggi, sedangkan tingkat kemiripan antar cluster rendah [17]. Jika data yang digunakan tidak memiliki kemiripan, maka data akan dikelompokkan dalam cluster yang berbeda. Tujuan dari clustering sendiri yaitu membagi data ke dalam beberapa cluster yang memiliki kemiripan sehingga menghasilkan sebuah informasi baru terkait data tersebut [9].

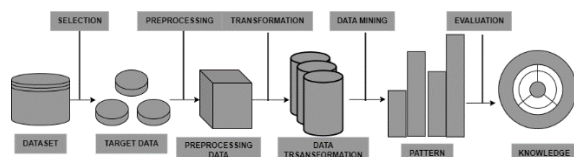
2.4. K-Means

K-Means merupakan algoritma pengelompokkan dengan kategori unsupervised yang cukup populer, karena memiliki beberapa keunggulan seperti efisien, cepat, dan ringkas dalam prosesnya [18]. Algoritma ini menghitung jarak antara objek data dengan pusat cluster, dan objek data yang dekat dengan pusat cluster dibagi ke dalam kategori yang sama [19]. K-Means juga telah digunakan pada berbagai aplikasi karena kesederhanaan dan konvergensinya yang cukup cepat sehingga membuat algoritma ini menjadi salah satu metode partisi yang paling efektif [20].

3. METODE PENELITIAN

Pada penelitian analisa cluster data jumlah kunjungan wisatawan di Kabupaten Karawang metode yang digunakan yaitu metode *Knowledge Discovery in Database (KDD)*. Metode KDD digunakan untuk mendapatkan pemahaman dari data jumlah kunjungan

wisatawan di Kabupaten Karawang, sehingga dapat menghasilkan suatu informasi guna mengambil keputusan untuk meningkatkan pengelolaan objek wisata (Gustientiedina et al., 2019). Metode ini memiliki 5 tahapan yaitu data selection, pre-processing, transformation, data mining, dan evaluation/interpretation (Ashari et al., 2022). Tahapan-tahapan pada metode KDD tersebut dapat dilihat pada gambar berikut.



Gambar 1. Tahapan Knowledge Discovery in Database (KDD)

3.1. Data Selection

Tahap awal pada proses data mining sesuai dengan metode *Knowledge Discovery in Database* (KDD) yaitu penyeleksian data. Seleksi data adalah tahapan pemilihan data dari set data yang digunakan sebelum mulai kepada tahapan penggalian informasi [9]. Dari dataset yang digunakan, penyeleksian dilakukan pada data wisatawan asing dan data tahun.

3.2. Preprocessing

Tahap ini merupakan tahap pembersihan data dengan membuang data yang bersifat null atau missing, membuang atribut yang kurang berpengaruh pada penelitian, membersihkan data yang memiliki anomali, dan memberikan identitas atau id pada data yang akan diteliti [10]. Penentuan id pada tahap ini dipilih dengan atribut yang bersifat unik yaitu nama objek wisata, sedangkan atribut yang digunakan hanyalah nama kabupaten, nama objek wisata dan jumlah wisatawan.

3.3. Transformation

Tahap transformasi data yaitu tahapan mengubah tipe data agar data dapat diolah sesuai kebutuhan sehingga data siap untuk diproses [21]. Pada tahap ini dilakukan transformasi dari tipe data non-numerik menjadi data numerik.

3.4. Data Mining

Tahapan ini merupakan proses pencarian informasi dari set data yang digunakan sehingga menghasilkan informasi baru antar data yang dapat dimanfaatkan untuk dasar pengambilan keputusan bagi pihak atau organisasi terkait yang berhubungan dengan set data yang digunakan [21]. Tahap ini dilakukan menggunakan *software Rapidminer v.10.2*.

3.5. Evaluation

Pada tahap ini hasil ditampilkan dengan cara yang mudah dipahami, dapat ditampilkan berupa tabel maupun gambar [19]. Tahap terakhir yaitu evaluasi

dengan menampilkan hasil dari *clustering* data jumlah kunjungan wisatawan di Kabupaten Karawang

4. HASIL DAN PEMBAHASAN

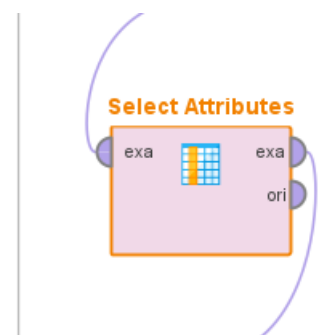
Berikut adalah hasil serta pembahasan dari penelitian analisa data analisa *cluster* data kunjungan wisatawan di Kabupaten Karawang menggunakan teknik clustering dengan algoritma *K-Means* dan metode *Knowledge Discovery in Database* (KDD).

4.1. Data Selection

Data yang digunakan pada penelitian ini merupakan data jumlah kunjungan wisatawan pada objek wisata di Kabupaten Karawang selama periode tahun 2020-2022 dengan total jumlah data sebanyak 346 dengan atribut terdiri dari 10 atribut.

4.2. Preprocessing

Langkah berikutnya adalah *preprocessing*, yaitu memilih atribut yang akan digunakan dan menambahkan id atau identitas pada data yang akan diolah. Pada tahap ini digunakan operator *Select Attributes* untuk membuang atribut yang tidak akan digunakan.



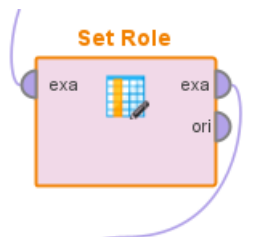
Gambar 2. Operator Select Attributes

Pada parameter *attribute filter type* dipilih opsi *a subset* untuk dapat memilih attribute yang akan digunakan dan tidak akan digunakan. Berikut adalah hasil dari pemilihan atribut yang digunakan.

Tabel 1. Atribut yang dipilih

Attributes	Selected Attributes
kategori_wisatawan	jumlah_wisatawan
kode_kabupaten	nama_kecamatan
kode_kecamatan	nama_objek_wisata
nama-kabupaten	
No	
satuan	

Setelah memilih atribut, selanjutnya adalah memilih id pada data. Data yang dipilih haruslah bersifat unik atau tidak memiliki duplikat. Operator yang digunakan adalah *Set Role*.



Gambar 3. Operator Set Role

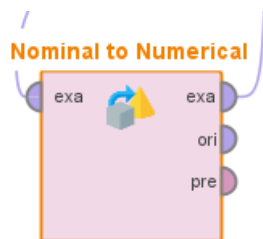
Pada data yang digunakan, atribut yang dijadikan ID atau identitas adalah atribut nama_objek_wisata.

Tabel 2. Atribut ID

Attribute Name	Target Role
nama_objek_wisata	id

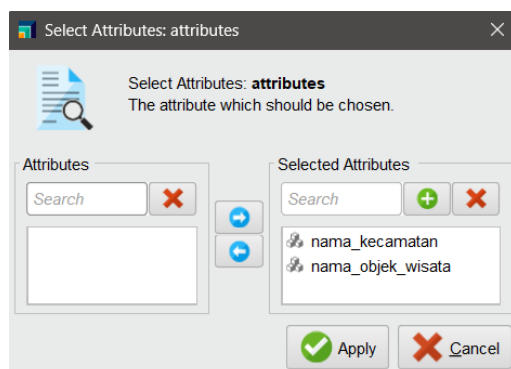
4.3. Transformation

Setelah pemilihan atribut dan penentuan id, maka tahapan berikutnya adalah transformasi data. Data yang telah dipilih harus diubah kedalam bentuk numerik untuk dapat melihat jumlah k optimal berdasarkan performa *Davies Bouldin*. Operator yang digunakan adalah *Nominal to Numerical*.



Gambar 4. Operator Nominal to Numerical

Pada parameter *attribute filter type*, pilih opsi *subset* untuk dapat memilih atribut yang akan di transformasi. Pada coding type nya gunakan *unique integers*. Atribut yang diubah adalah atribut yang telah dipilih sebelumnya menggunakan operator *Select Attributes*.



Gambar 5. Atribut Nominal to Numerical

Setelah selesai memilih atribut maka jalankan proses agar dapat melihat hasil transformasi data. Berikut adalah hasil dari transformasi data *Nominal to Numerical*.

Tabel 3. Hasil transformasi data

Nama objek wisata	Nama Kecamatan	Jumlah wisatawan
Maqom Mbah Rubiah	0	34
Eks Kantor Kec. Rengasdengklok	1	135
Situs Wadon Cibuaya	2	252
...	...	...
MAKAM PASAREAN	17	19132
Monumen Tugu Rengasdengklok	1	19206
WATERBOOM KAMPUNG TURIS	4	24191
Makam Syekh Quro - Syekh Bentong	7	28282
Pantai Sedari	2	123562

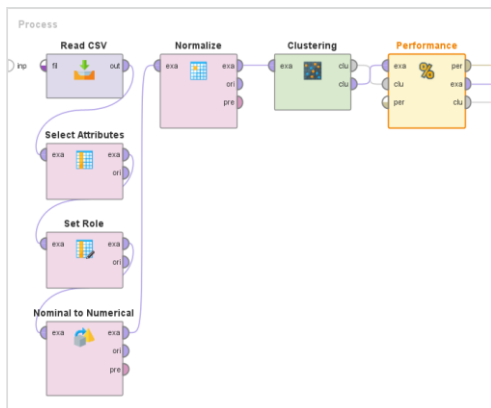
Berdasarkan tabel diatas, dapat dilihat bahwa jarak nilai antara data nama kecamatan dengan data jumlah wisatawan sangat berbeda jauh. Hal tersebut dapat mempengaruhi hasil *clustering* menjadi kurang optimal sehingga perlu dilakukan normalisasi data. Operator yang digunakan adalah operator *Normalize*. Pada parameter *attribute filter type* dipilih opsi *subset* untuk memilih atribut yang akan di normalisasikan. Pada parameter *method* digunakan tipe *range transformation* agar hasil normalisasi data berada pada range yang telah ditentukan, pada penelitian ini nilai minimal dan maksimal range nya berdasarkan nilai default yaitu dari 0,0 sampai 1,0. Atribut yang dinormalisasi adalah atribut nama kecamatan dan jumlah wisatawan. Berikut adalah hasil dari normalisasi data.

Tabel 4. Hasil normalisasi data

Nama objek wisata	Nama Kecamatan	Jumlah wisatawan
Maqom Mbah Rubiah	0	0
Eks Kantor Kec. Rengasdengklok	0,045	0,000
Situs Wadon Cibuaya	0,091	0,001
...	...	...
MAKAM PASAREAN	0,773	0,059
Monumen Tugu Rengasdengklok	0,045	0,059
WATERBOOM KAMPUNG TURIS	0,182	0,074
Makam Syekh Quro - Syekh Bentong	0,381	0,087
Pantai Sedari	0,091	0,379

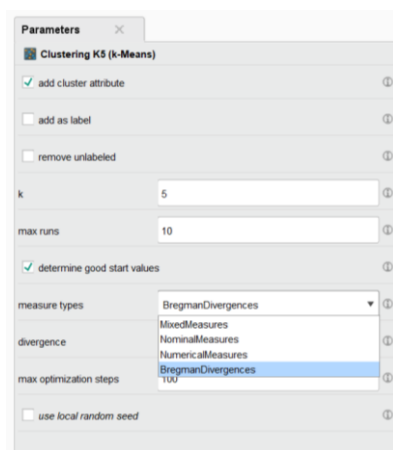
4.4. Data Mining

Tahap pemrosesan data dengan menggunakan algoritma *K-Means clustering*. Pada tahap ini digunakan operator *k-Means Clustering* untuk pengelompokkan dan operator *Cluster Distance Performance* untuk mengukur hasil performa *clustering* berdasarkan nilai *Davies Bouldin*.



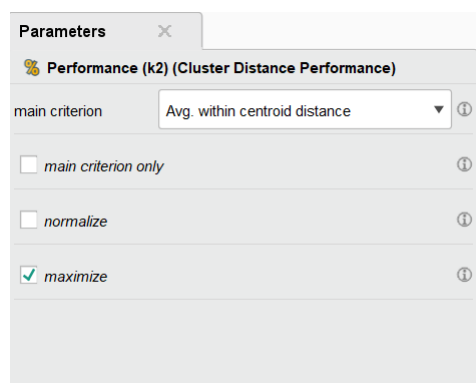
Gambar 6. Model proses data mining *k-means clustering*

Pada penelitian ini, dilakukan uji coba sebanyak 6 kali mulai dari $k=2, k=3, k=4, k=5, k=6, k=7$ untuk mencari jumlah k yang optimal.



Gambar 7. Parameter operator *k-means*

Penelitian ini juga mengujikan beberapa jenis *measure types* pada operator *K-Means* untuk mencari hasil *Davies Bouldin* terbaik. Untuk operator *Cluster Distance Performance*, klik centang pada bagian *maximize* agar hasil *Davies Bouldin* tidak bersifat negatif.



Gambar 8. Parameter *Cluster Distance Performance*

4.5. Evaluation

Setelah melakukan beberapa pengujian diatas, berikut adalah hasil *DBI* dari *clustering* menggunakan *K-Means*.

Tabel 5. Hasil *DBI*

Jumlah k	DBI Numerical Measures	DBI Bregman Divergences
2	0,483	0,448
3	0,521	0,422
4	0,622	0,609
5	0,521	0,557
6	0,619	0,609
7	0,560	0,568

Berdasarkan tabel diatas, dapat dilihat bahwa *DBI* terkecil berada pada jumlah $k=3$ menggunakan parameter *Bregman Divergences* dengan nilai 0,422. Untuk parameter *Nominal Measures* tidak dapat digunakan karena tidak terdapat data bersifat nominal, dan untuk parameter *Mixed Measures* akan sama hasilnya dengan *Numerical Measures* karena data yang diuji telah bersifat numerik semua.

Cluster Model

Cluster 0: 31 items
Cluster 1: 61 items
Cluster 2: 3 items
Total number of items: 95

Gambar 9. Hasil cluster $k3$

Hasil dari $k=3$ dengan nilai *DBI* 0,422 yaitu kelompok 1 memiliki total 31 data, kelompok 2 memiliki total 61 data, dan kelompok 3 memiliki total 3 data.

Tabel 6. Hasil *cluster 0*

Nama Kecamatan	Nama Objek Wisata	Jumlah Wisatawan
Pakisjaya	Makam Syekh Zakaria	844
Pedes	Lemah Duwur Lanang	909
Cibuaya	Candi Wadon	976
Rawamerta	Kolam Tirta Pitu Mulang	1590
Batujaya	Situs Damar (Seg Iii)	1897
...	...	...

Berdasarkan tabel diatas, dapat diketahui bahwa *cluster 0* adalah kelompok dengan jumlah wisatawan sedang dengan rentang jarak mulai dari ratusan hingga puluhan ribu jumlah wisatawan.

Tabel 7. Hasil *cluster* 1

Nama Kecamatan	Nama Objek Wisata	Jumlah Wisatawan
Karawang Barat	Maqom Mbah Rubiah	34
Rengasdengklok	Eks Kantor Kec. Rengasdengklok	135
Cibuaya	Situs Wadon Cibuaya	252
Pakisjaya	Situs Lingga	265
Tegalwaru	Curug Cikarapyak	344
...	...	...

Dilihat dari tabel diatas, *cluster* 1 termasuk kedalam kelompok wisatawan rendah dengan rentang jarak dari puluhan hingga ratusan.

Tabel 8. Hasil *cluster* 2

Nama Kecamatan	Nama Objek Wisata	Jumlah Wisatawan
Cibuaya	Pantai Sedari	123562
Pakisjaya	Pantai Tanjung Pakis	144733
Lemah Abang	Maqom Syech Bentong	242565

Berdasarkan tabel diatas, dapat dilihat bahwa *cluster* 2 terdiri dari 3 data dengan rentang jarak ratusan ribu. Dengan begitu maka dapat disimpulkan bahwa *cluster* 2 termasuk kedalam kelompok jumlah wisatawan tinggi.

5. KESIMPULAN DAN SARAN

Berdasarkan hasil dari *clustering* diatas dapat diambil kesimpulan bahwa jumlah k yang optimal adalah $k=3$, dengan *cluster* 0 sebagai kelompok jumlah wisatawan sedang, *cluster* 1 sebagai kelompok jumlah wisatawan rendah dan *cluster* 2 sebagai kelompok wisatawan tinggi. Dengan begitu, hasil dari *clustering* ini diharapkan dapat membantu pemerintah untuk dapat mengembangkan objek wisata berdasarkan kelompok tersebut.

Saran untuk penelitian kedepannya adalah dapat meneliti dengan metode *clustering* lain seperti dengan menggunakan algoritma *Fuzzy*, *X-Means*, *K-Medoids*, dll. Saran lainnya adalah untuk dapat mencari permasalahan baru yang lebih relevan untuk data kunjungan wisatawan di Kabupaten Karawang.

DAFTAR PUSTAKA

- [1] Y. F. Taopan, M. R. Oedjoe, and A. N. Sogen, "Dampak Perkembangan Teknologi Informasi dan Komunikasi Terhadap Perilaku Moral Remaja di SMA Negeri 3 Kota Kupang," *J. Kependidikan J. Has. Penelit. dan Kaji. Kepustakaan di Bid. Pendidikan, Pengajaran dan Pembelajaran*, vol. 5, no. 1, p. 61, 2019, doi: 10.33394/jk.v5i1.1395.
- [2] I. Hizbullah and M. Salmin, "Perencanaan Strategis Sistem Informasi/Teknologi Informasi Menggunakan Framework TOGAF Pada Dinas Pariwisata Kabupaten Pulau Morotai," *Teknika*, vol. 10, no. 2, pp. 122–127, 2021, doi: 10.34148/teknika.v10i2.355.
- [3] Muhamad Ridwan and I. Meitasari, "Potensi Pariwisata Situs Candi Jiwa Batujaya Kabupaten Karawang," *Satwika Kaji. Ilmu Budaya dan Perubahan Sos.*, vol. 7, no. 2, pp. 447–459, 2023, doi: 10.22219/satwika.v7i2.28126.
- [4] D. Widiyawati, D. Dedih, and W. Wahyudi, "Implementasi Metode Maut Dan Saw Dalam Pemilihan Tempat Wisata Di Kabupaten Karawang," *J. Interkom J. Publ. Ilm. Bid. Teknol. Inf. dan Komun.*, vol. 17, no. 2, pp. 71–80, 2022, doi: 10.35969/interkom.v17i2.231.
- [5] B. M. Al-Fahmi, E. Rahmawati, and T. Sagirani, "Penerapan K-Means Clustering Pada Pariwisata Kabupaten Bojonegoro Untuk Mendukung Keputusan Strategi Pemasaran," *J. Nas. Teknol. dan Sist. Inf.*, vol. 9, no. 2, pp. 141–149, 2023, doi: 10.25077/teknosi.v9i2.2023.141-149.
- [6] L. F. Marini and C. D. Suhendra, "Penggunaan Algoritma K-Means Pada Aplikasi Pemetaan Klaster Daerah Pariwisata," *J. Media Inform. Budidarma*, vol. 7, no. 2, pp. 707–713, 2023, doi: 10.30865/mib.v7i2.5558.
- [7] B. Setio and P. Prasetyaningrum, "Penerapan Data Mining Dalam Mengelompokkan Kunjungan Wisatawan Di Kota Yogyakarta Menggunakan Metode K-Means," *J. Comput. Sci. Technol.*, vol. 1, no. 1, pp. 27–32, 2021, doi: 10.54840/jcstech.v1i1.9.
- [8] S. Handoko, F. Fauziah, and E. T. E. Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020, doi: 10.35760/tr.2020.v25i1.2677.
- [9] A. Wibowo, Moh Makruf, Inge Virdyna, and Farah Chikita Venna, "Penentuan Klaster Koridor TransJakarta dengan Metode Majority Voting pada Algoritma Data Mining," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 5, no. 3, pp. 565–575, 2021, doi: 10.29207/resti.v5i3.3041.
- [10] T. Hartati and Y. Arie Wijaya, "Analisis Data Lalu Lintas Jaringan Di Kantor Cangehgar Cyber Operation Center Menggunakan Algoritma K-Means Network Traffic Data Analysis At Cangehgar Cyber Operation Center Office Using K-Means Algorithm," *J. Ilm. NERO*, vol. 7, no. 1, pp. 75–84, 2022.
- [11] M. Siahaan, "Data Mining Strategi Pembangunan Infrastruktur Menggunakan Algoritma K-Means," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 3, pp. 316–324, 2022, doi: 10.32736/sisfokom.v11i3.1453.
- [12] E. Satria et al., "Penerapan Clustering dalam Mengelompokkan Jumlah Kunjungan Wisatawan Mancanegara Dengan Metode K-Means," no. September, p. 462, 2019, [Online]. Available:

- <https://www.bps.go.id/dynamictable/2018/07/30/1548/jumlah-kunjungan->
- [13] Nurhachita and E. S. Negara, "A comparison between deep learning, naïve bayes and random forest for the application of data mining on the admission of new students," *IAES Int. J. Artif. Intell.*, vol. 10, no. 2, pp. 324–331, 2021, doi: 10.11591/ijai.v10.i2.pp324-331.
 - [14] B. G. Sudarsono and S. P. Lestari, "Clustering Penerima Beasiswa Yayasan Untuk Mahasiswa Menggunakan Metode K-Means," *J. Media Inform. Budidarma*, vol. 5, no. 1, p. 258, 2021, doi: 10.30865/mib.v5i1.2670.
 - [15] E. Luthfi and A. W. Wijayanto, "Analisis perbandingan metode hirerarchical, k-means, dan k-medoids clustering dalam pengelompokan indeks pembangunan manusia Indonesia," *Inovasi*, vol. 17, no. 4, pp. 761–773, 2021, doi: 10.30872/jinv.v17i4.10106.
 - [16] I. F. Ashari, E. Dwi Nugroho, R. Baraku, I. Novri Yanda, and R. Liwardana, "Analysis of Elbow, Silhouette, Davies-Bouldin, Calinski-Harabasz, and Rand-Index Evaluation on K-Means Algorithm for Classifying Flood-Affected Areas in Jakarta," *J. Appl. Informatics Comput.*, vol. 7, no. 1, pp. 89–97, 2023, doi: 10.30871/jaic.v7i1.4947.
 - [17] H. Gunawan and V. Purwayoga, "Data Mining Menggunakan Algoritma K-Means Clustering Untuk Mengetahui Potensi Penyebaran Virus Corona Di Kota Cirebon," *J. Sisfokom (Sistem Inf. dan Komputer)*, vol. 11, no. 1, pp. 1–8, 2022, doi: 10.32736/sisfokom.v11i1.1316.
 - [18] R. Sebastian and C. Julianne, "Application of Data Mining to Group the Spread of Covid-19 in West Java Province, Indonesia Using the K-Means Algorithm," *Indones. J. Artif. Intell. Data Min.*, vol. 5, no. 2, pp. 83–90, 2022, [Online]. Available: <https://ejournal.uin-suska.ac.id/index.php/IJAIDM/article/view/18721>
 - [19] D. Ramdhan, G. Dwilestari, R. D. Dana, A. Ajiz, and K. Kaslani, "Clustering Data Persediaan Barang Dengan Menggunakan Metode K-Means," *MEANS (Media Inf. Anal. dan Sist.*, vol. 7, no. 1, pp. 1–9, 2022, doi: 10.54367/means.v7i1.1826.
 - [20] Z. kai Feng, W. jing Niu, R. Zhang, S. Wang, and C. tian Cheng, "Operation rule derivation of hydropower reservoir by k-means clustering method and extreme learning machine based on particle swarm optimization," *J. Hydrol.*, vol. 576, no. June, pp. 229–238, 2019, doi: 10.1016/j.jhydrol.2019.06.045.
 - [21] Haris Kurniawan, Sarjon Defit, and Sumijan, "Data Mining Menggunakan Metode K-Means Clustering Untuk Menentukan Besaran Uang Kuliah Tunggal," *J. Appl. Comput. Sci. Technol.*, vol. 1, no. 2, pp. 80–89, 2020, doi: 10.52158/jacost.v1i2.102.