

## IMPLEMENTASI K-MEANS CLUSTERING MENGGUNAKAN RAPIDMINER DALAM PENGELOMPOKAN DATA KUNJUNGAN WISATAWAN ASING DI PROVINSI JAWA TIMUR

Edwin Wira Liyanto<sup>1</sup>, Ahmad Homaidi<sup>2</sup>, Ahmad Lutfi<sup>3</sup>

Jurusan Teknologi Informasi, Fakultas Sains dan Teknologi, Universitas Ibrahimy  
Jl. KHR. Syamsul Arifin No.1-2, Sukorejo, Situbondo 68374, Jawa Timur, Indonesia  
e-mail : <sup>1</sup>[agistasinta15@gmail.com](mailto:agistasinta15@gmail.com), <sup>2</sup>[ahmadhomaidi@ibrahimiy.ac.id](mailto:ahmadhomaidi@ibrahimiy.ac.id), <sup>3</sup>[ahmadlutfi.14@gmail.com](mailto:ahmadlutfi.14@gmail.com)

### ABSTRAK

Pariwisata adalah salah satu sektor yang mampu memberikan sumbangsi yang sangat signifikan kepada perekonomian di suatu negara, terlebih pengelolaan pariwisata yang baik dapat memberikan dampak positif bagi sosial dan ekonomi di suatu wilayah. Jawa Timur menjadi salah satu wilayah destinasi wisata alam yang cukup banyak menarik wisatawan mancanegara dengan daya tarik wisata, tradisi, budaya dan kearifan lokalnya yang menjadi target utama kunjungan wisatawan mancanegara. Pengelompokan wisatawan dilakukan dengan teknik Data Mining menerapkan algoritma k-means, analisis yang digunakan memakai metode Knowledge Discovery in Database (KDD). Dengan menerapkan algoritma K-Means clustering dalam rangka klasterisasi data Kunjungan Wisatawan Mancanegara (ASIA) ke Jawa Timur Melalui Bandara Juanda Tahun 2016 – 2023 yang diambil dari portal resmi BPS Jawa Timur. Dari data yang telah diolah ditemukan 2 cluster yang menghasilkan nilai dbi paling optimal yaitu 0.293 dengan cluster 0 berisi 2 data, cluster 1 48 data.

**Kata Kunci:** Pariwisata, K-Means, Data Mining.

### ABSTRACT

*Tourism is one sector that is able to make a very significant contribution to the economy of a country, especially good tourism management can have a positive impact on the social and economy of a region. East Java is one of the natural tourist destination areas that attracts quite a lot of foreign tourists with its tourist attractions, traditions, culture and local wisdom which are the main targets for foreign tourist visits. Tourist grouping is carried out using Data Mining techniques applying the k-means algorithm, the analysis used uses the Knowledge Discovery in Database (KDD) method. By applying the K-Means clustering algorithm in order to cluster data on Foreign Tourist Visits (ASIA) to East Java via Juanda Airport for 2016 - 2023 taken from the official BPS East Java portal. From the data that has been processed, 2 clusters were found which produced the most optimal dbi value, namely 0.293 with cluster 0 containing 2 data, cluster 1 48 data.*

**Keywords:** Tourism, K-Means, Data Mining.

## I. PENDAHULUAN

### 1.1. Latar Belakang Penelitian

Jawa Timur sebagai Provinsi paling timur dari kepulauan Jawa yang dipenuhi dengan keragaman dan kekayaan alam maupun budaya yang sangat menarik untuk pelajari secara eksploratif. Namun, masih banyak daerah di Jawa Timur yang tidak paham dan sadar sebesar apa potensi yang dimiliki.

Oleh karena itu menguraikan setiap potensi daerah dalam memanfaatkan kekayaan alamnya sehingga mampu membentuk satu ikon pesona daerah yang menjadi magnet untuk siapapun yang berkunjung ke tempat pariwisata tersebut. Semua potensi yang ada di Jawa Timur masih banyak untuk di eksplorasi lebih dalam hingga mampu menjadi sebuah daya tarik yang mampu mengungkit perekonomian yang di Jawa Timur secara signifikan.

Sektor pariwisata menjadi salah satu sektor yang memperoleh prioritas utama untuk memulihkan sektor ekonomi yang ada di suatu daerah.[1] Pariwisata sendiri juga merupakan ujung tombak dari peningkatan pendapatan daerah karena dari sektor pariwisata sendiri dapat menumbuhkan perekonomian masyarakat di segala bidang karena pariwisata merupakan sektor yang dapat terus berkembang.[2] Demi peningkatan kunjungan wisatawan terutama wisatawan mancanegara, dibutuhkanlah sebuah strategi dalam meningkatkan kontribusi di sektor wisata dalam menunjang keberlangsungan sektor wisata di suatu daerah. Tujuan utama wisatawan mancanegara datang yakni ingin mengenal lebih jauh tradisi, budaya serta kearifan lokal yang terdapat di suatu tempat. Pola hidup dan ilmu pengetahuan serta bermacam-macam cara hidup yang berwujud kegiatan yang dilakukan oleh masyarakat disekitar sebagai upaya menyelesaikan berbagai masalah kehidupan mereka adalah esensi dari kearifan lokal sendiri.[3] Banyak hal yang dapat dipelajari dari kearifan lokal yang ada di Jawa Timur, sehingga mampu memikat wisatawan asing untuk sekedar singgah dan menghabiskan waktunya di tempat wisata yang ada.

Pariwisata sendiri memiliki banyak efek pada perkembangan sebuah wilayah. Salah satu efek yang dapat dirasakan langsung oleh masyarakat yakni tersedianya banyak kesempatan pekerjaan bagi masyarakat yang bersentuhan langsung dengan pariwisata.[4] Mulai dari sektor transportasi, kuliner, hingga produk olahan yang berbasis perumahan maupun industri dapat merasakan langsung dampak dari pariwisata yang ada di suatu daerah. Berbagai produk pariwisata juga dapat terus berkembang seiring berjalannya waktu dan memiliki pengaruh jangka panjang jika dapat diolah dengan baik. Produktifitas dari pelayanan, objek wisata serta event yang mampu menarik lebih banyak wisatawan akan memberikan keuntungan terhadap suatu wilayah. Sedangkan wisatawan adalah sebuah komponen penting yang bersanding selaras dengan dunia pariwisata. Wisatawan memiliki beragam jenis, mancanegara dan lokal, tua dan muda, miskin dan kaya, semuanya memiliki keinginan dan harapan yang berbeda-beda. Semua kalangan wisatawan

memiliki tujuan yang sama yakni untuk mengunjungi suatu tempat wisata terlepas dari berbagai tujuannya namun tetap pada visi yang sama yakni bepergian dan melepas rutinitas setiap harinya.

Dengan terus berkembangnya zaman hingga kini trend pariwisata mengalami banyak perubahan yang cukup signifikan, dari pariwisata yang berbasis konvensional berubah menjadi pariwisata berbasis minat khusus.[5] Seperti wisata petualangan yang memberikan sensasi pengalaman yang luar biasa dan juga sensasional, di Jawa Timur sendiri begitu banyak wisata minat khusus seperti ini pada kegiatan wisata, *hiking*, *snorkeling*, dan lain sebagainya. Pada wisata minat khusus, wisata kuliner menjadi salah satu bagian wisata minat khusus. Jawa Timur dengan sejuta budaya dan letak geografisnya yang beragam memiliki banyak wisata kuliner yang dapat di jelajahi di setiap sudut daerahnya, mulai dari pantai hingga pegunungan selalu menawarkan kuliner dengan cita rasa yang beragam. Wisata edukasi yang tengah dikembangkan oleh banyak tempat wisata ternyata menjadi bagian dari wisata minat khusus, saat ini di Jawa Timur sudah banyak sekali menawarkan wisata edukasi seperti Jatim Park 2 dan 3 yang menawarkan wisata edukasi flora dan fauna bagi segala usia. Pilihan terakhir dari trend wisata edukasi adalah wisata olahraga yang cukup banyak menarik minat wisatawan, wisata olahraga biasanya menawarkan berbagai kegiatan yang menguji adrenalin seperti adalah surfing, rafting, snorkelin, tubing dan lain sebagainya. Dapat disimpulkan bahwa wisata minat khusus mungkin menjadi pilihan bagi para pelancong yang akan berlibur dengan cara berbeda. Wisata minat khusus pada umumnya mengajarkan kita untuk lebih menghargai alam dengan semua kekayaannya yang terkandung di dalamnya.

Semua tempat wisata yang ada di Jawa Timur memang mampu menarik minat wisatawan mancanegara untuk berkunjung. Namun tidak dipungkiri bahwa belum banyak wisata yang di ketahui oleh banyak masyarakat sendiri, terlebih wisatawan mancanegara yang lebih berfokus terhadap wisata-wisata yang sudah berkembang dan memiliki nilai jual yang pasti. Padahal tidak dapat dipungkiri

wisata yang ada di Jawa Timur tidak terbatas hanya itu saja, banyak sekali wisata yang mulai berkembang dan menjadi daya tarik wisatawan lokal. Hal tersebut tidak lepas dari kurangnya informasi wisatawan asing terhadap tempat-tempat wisata tersebut dan lebih banyak *tour guide* yang masih belum mengetahui nilai dari suatu tempat wisata. Belum adanya penelitian dan kajian terhadap tempat wisata baru juga menjadi keterbatasan yang harus segera ditemukan penyelesaiannya sehingga tidak hanya tempat wisata yang sudah memiliki nilai yang menjadi daya ungkit kunjungan wisatawan mancanegara melainkan wisata-wisata berkembang lainnya yang memiliki nilai budaya, tradisi dan kearifan lokal lainnya mampu mendorong hal tersebut.

Adanya pandemi Covid-19 menjadi krisis terbesar yang pernah di alami oleh seluruh dunia, tidak terkecuali Indonesia dan Jawa Timur pada khususnya. Hampir semua sektor mengalami penurunan yang sangat signifikan akibat pandemi Covid-19. Sektor kesehatan yang paling terganggu memberikan efek yang besar bagi ekonomi dan memunculkan beragam problematika yang saling berkaitan dan menciptakan efek domino yang tak terduga. Pariwisata pun tidak ketinggalan menjadi salah satu sektor yang mengalami kemunduran yang sangat signifikan akibat pandemi. Kebijakan isolasi mandiri serta tertutupnya pintu masuk negara melalui bandara nasional dan internasional mempengaruhi jumlah wisatawan mancanegara yang ingin berkunjung ke Indonesia, serta seluruh komoditas ekonomi yang menjadi satu kesatuan dalam kegiatan pariwisata, agen *travel*, hotel dan *guest house* menjadi beberapa komoditas yang terdampak langsung. Seiring berjalannya waktu, kondisi pariwisata di tahun 2023 semakin menunjukkan pemulihan. Pemerintah terus menggelontorkan berbagai program serta kebijakan dalam upaya mendukung kembalinya sektor pariwisata pasca pandemi Covid-19. Pengembangan pariwisata berupa program yang tepat sasaran sangat dibutuhkan untuk meningkatkan jumlah kedatangan wisatawan mancanegara. Strategi yang dikembangkan seperti meningkatkan infrastruktur, promosi pariwisata secara agresif dapat memberikan kemudahan berwisata ke destinasi pariwisata.

Sektor pariwisata Jawa Timur pada tahun 2023 mengalami perkembangan yang cukup baik jika dibandingkan dengan daerah lain di Indonesia, hal ini di tandai dengan naiknya nilai Produk Domestik Regional Bruto (PDRB ADHK 2010) Provinsi Jawa Timur sektor akomodasi serta produk makan minum mencapai 101,73 triliun rupiah. Nilai ini lebih tinggi dibandingkan tahun 2022 (PDRB ADHK 2010) yang mencapai 94,15 triliun rupiah (BPS Jawa Timur, 2024). Kenaikan ini menunjukkan bahwa sektor pariwisata di Jawa Timur mengalami kebangkitan pasca pandemi Covid-19. UNWTO sebagai organisasi pariwisata meramalkan bahwa Covid-19 akan berdampak pada (PDB) yang melonjak tajam sebesar 1,5 persen hingga 2,8 persen. Pada awal tahun sampai Mei tahun 2020, telah terjadi penyusutan yang begitu drastis pada banyaknya wisatawan mancanegara yang berkunjung dengan penurunan yakni mencapai 56%, dan turunnya penghasilan dari ekspor pariwisata mencapai 320 juta USD. Selain dampak tersebut, Covid-19 juga beresiko menghilangkan hingga 120 juta pekerjaan yang berhubungan dengan dunia wisata. Pariwisata sendiri merupakan sektor dengan sebagian besar kontribusinya berasal dari padat karya serta kegiatan usahanya mikro kecil menengah. Sempat terpuruk karena pandemi Covid-19, kini sektor pariwisata menunjukkan peningkatan yang cukup drastis. Pemulihan sektor pariwisata juga terlihat dari pencapaian Indonesia di forum internasional. *Travel and Tourism Competitiveness Index* (TTDI) Indonesia naik 12 peringkat menjadi posisi 32 di atas Malaysia, Thailand, Vietnam, dan Filipina. Bahkan Indonesia mencapai peringkat pertama dalam *Global Muslim Travel Index* (GMTI) menurut sumber dari BPJ Jawa Timur.

Diluar permasalahan kompleks yang terjadi di sektor pariwisata yang ada di Jawa Timur, pemerintah serta elemen masyarakat di dunia pariwisata harus terus berkolaborasi untuk terus meningkatkan nilai tambah sektor pariwisata khususnya kunjungan wisatawan mancanegara. Promosi yang tepat serta target audiens yang beragam dari berbagai negara yang berkunjung ke Jawa Timur menjadi point penting dalam mengambil keputusan. Data kunjungan wisatawan asing yang di peroleh dari data publik yang di publikasikan oleh BPS Jawa Timur

akan menjadi data utama dalam proses penelitian. Penelitian akan mengklasifikasikan mana saja negara yang layak untuk menjadi target promosi dengan menggunakan algoritma k-means *clustering*.

## 1.2. Landasan Teori

Keputusan yang tepat harus diambil dengan mengutamakan aspek data yang relevan sehingga mampu mewujudkan strategi yang baik dalam menentukan kebijakan. Penentuan kebijakan tersebut tidak hanya dapat dilakukan oleh pemerintah sebagai lading sektor pariwisata pada umumnya, masyarakat yang terlibat dalam usaha wisata sendiri dapat menggunakan penelitian ini sebagai landasan teori yang penting dalam mempromosikan tempat-tempat wisata di Jawa Timur. Data yang digunakan dalam penelitian ini adalah data kunjungan wisatawan mancanegara (ASIA) ke Jawa Timur melalui bandara Juanda Tahun 2016 – 2023 yang diambil melalui portal BPS Jawa Timur dengan sifat data publik. Data tersebut berasal dari buku Statistik Pariwisata Provinsi Jawa Timur yang diterbitkan oleh BPS Jawa Timur pada setiap tahun. Dari data tersebut akan dilakukan pemisahan secara kelompok negara mana yang memiliki jumlah wisatawan terbanyak hingga yang paling sedikit guna mencari sasaran yang tepat dari promosi wisata yang akan dilakukan nantinya. Data yang digunakan merupakan dataset kunjungan wisatawan mancanegara melalui bandara Juanda dari tahun 2016 s.d 2023 di Provinsi Jawa Timur. Data yang digunakan telah memiliki struktur data yang jelas, jenis yang tetap, serta format data yang tidak rubah, sehingga data ini disebut sebuah data terstruktur.[6]

Data mining merupakan suatu tahap penyimpanan informasi dan data yang menggunakan berbagai teknik seperti matematika, statistik, artificial intelligence (AI) serta machine learning untuk mencari dan menerjemahkan berbagai informasi atau pengetahuan yang berguna untuk disimpan di dalam database. Data mining merupakan sebagian proses dari *Knowledge Discovery in Databases* (KDD) yang tersusun dari beberapa langkah yakni pemilihan, pengolahan, transformasi, data mining hingga evaluasi hasil. Ada 5 fungsi pada data mining yang pertama yakni :

*classification*, ini merupakan fungsi untuk menyimpulkan berbagai macam deskripsi dari berbagai karakteristik tertentu dari sepotong kelompok. *Clustering*, merupakan fungsi yang mendeskripsikan suatu kelompok dari sebuah data sehingga mempunyai karakteristik khusus perbedaan khusus pada clustering adalah dimana pada *clustering* tidak ada definisi karakteristik awal seperti pada *classification*. *Association*, fungsi yang biasa digunakan untuk mengidentifikasi hubungan antara kejadian yang terjadi pada suatu waktu. *Sequencing*, memiliki fungsi yang hampir sama dengan *association*, *sequencing* mengidentifikasi hubungan yang berbeda pada suatu periode waktu. Terakhir yakni *forecasting* yaitu fungsi untuk memperkirakan nilai pada masa depan berdasarkan pola atau kebiasaan tertentu.[7] *Clustering* adalah proses mencari dan mengelompokkan data dalam suatu data ke dalam kelompok data yang memiliki kesamaan (*similarity*) antara data yang satu dengan yang lainnya.[8]

Tahapan pada algoritma K-means clustering adalah sebagai berikut:[9]

1. Menentukan banyaknya cluster
2. Menentukan titik centroid, untuk menentukan titik centroid awal pada iterasi yakni dilakukan secara manual menurut peneliti. Namun ketika tahap iterasi pertama hingga iterasi terakhir ketika nilai centroid tidak berubah lagi, dipastikan dengan mengaplikasikan rumus berikut :

$$V_{ij} = \frac{1}{N_i} \sum_{k=0}^{N_i} X_{kj}$$

Keterangan :

$V_{ij}$  = Centroid average pada klaster ke-i untuk variabel ke-j

$N_i$  = Banyak himpunan pada klaster ke-i

$i, k$  = index dari klaster

$j$  = index dari variabel

$X_{kj}$  = nilai dari data ke-k variabel ke-j untuk klaster tersebut.

3. Untuk memastikan jarak gunakan rumus *Euclidean Distance* guna menentukan jarak tiap titik centroid dengan titik pada tiap objek, sehingga ditemukan jarak garis lurus memisahkan antara dua titik dalam ruang Euclidean.

$$D_e = \sqrt{(x_i - s_i)^2 + (y_i - t_i)^2}$$

Keterangan :

De = Euclidean Distance (Jarak)

i = Banyak objek

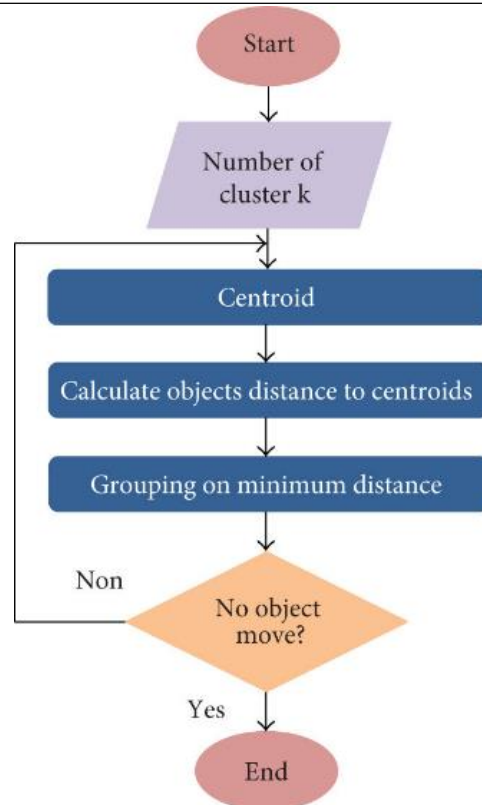
(x, y) = Titik Koordinat objek

(s, t) = Titik Koordinat centroid

Xkj = nilai dari data ke-k variabel ke-j untuk kluster tersebut.

4. Objek dikelompokkan sesuai anggota cluster dengan memperhitungkan jaraknya ke titik centroid paling dekat.
5. Lakukan pengulangan tahap 2 hingga tahapan 4 sampai nilai centroid yang keluaran menjadi tetap atau bernilai optimal, dan anggota kluster tidak ada yang berpindah ke kluster lainnya.

Bentuk algoritma k-means clustering adalah sebagai berikut :



Gambar 1. Flowchart K-means Clustering.

Untuk mengevaluasi dan menilai kluster dalam klasifikasi akan digunakan *Davies Bouldin Index* (DBI) berdasarkan nilai kohesi dan separasi (pemisahan). *Davies Bouldin Indeks* (DBI) juga disebut konsistensi dari serangkaian pengukuran klasifikasi indeks.[10] diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 *Davies Bouldin Index* (DBI) memiliki pengertian yakni perbandingan *average* jarak antar kluster untuk setiap kluster dengan kluster kelompok terdekatnya.[11] Rumus yang gunakan untuk menemukan kerangka kohesi dari sebuah kelompok ke-i yang disebut *Sum of Square Within Cluster* (SSW) dan diformulasikan sebagai berikut :

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_i, c_i)$$

Keterangan:

m = Banyak data pada kluster ke-i

c = Centroid kluster ke-i

d = Jarak (distance)

Sebaliknya fungsi yang digunakan untuk menemukan nilai pemisahan antar kelompok disebut *Sum of Square Between Cluster* (SSB) yang diformulasikan sebagai berikut:

$$SSB_{ij} = d(c_i, c_j)$$

Keterangan:

d = Jarak (distance)

c = Centroid klaster ke-i

Setelah menemukan nilai kohesi dan nilai separasi, dilanjutkan dengan pengukuran rasio ( $R_{ij}$ ) dengan tujuan untuk mencari nilai komparasi antara kelompok ke-i dan kelompok ke-j. Dan nantinya kelompok terbaik adalah klaster yang memiliki nilai afinitas terkecil dan nilai separasi terbesar. Rasio ( $R_{ij}$ ) dapat dihitung dengan menggunakan rumus berikut:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}}$$

Setelah nilai rasio ditemukan lalu gunakan untuk menghitung nilai Davies Bouldin Index (DBI) dengan cara berikut:

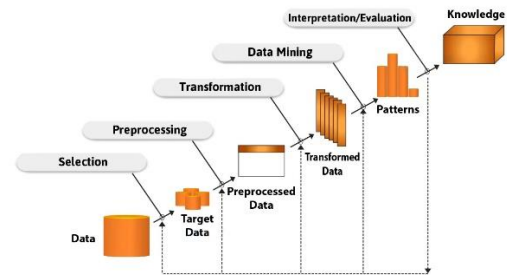
$$DBI = \sum_{i=1}^k \max_{i \neq j}$$

Setelah diketahui nilai k atau jumlah kluster dari persamaan Davies Bouldin Index (DBI) diambil kesimpulan jika semakin kecil nilai DBI yang diperoleh (non negatif  $\geq 0$ ) maka klaster tersebut semakin baik.[12]

## II. METODE PENELITIAN

### 2.1 Tahapan Penelitian

Dalam penelitian ini digunakan *Knowledge Discovery in Databases* (KDD) sebagai metode dalam mengevaluasi hasil pengolahan data. Pengelompokan data jumlah kunjungan wisatawan asing ke Jawa Timur akan dilakukan dengan memanfaatkan tool Rapidminer. Pengelompokan memanfaatkan operator atau fungsi yang ada pada Rapidminer. Operator yang akan diaplikasikan pada penelitian memiliki tujuan dan fungsi yang berbeda tergantung kebutuhan penelitian. Metodologis yang digunakan diuraikan pada gambar 1.



**Gambar 1.** Metode Knowledge Discovery in Databases (KDD)

Penjabaran gambar 1

1. Data dikumpulkan melalui laman resmi Badan Pusat Statistik Provinsi Jawa Timur mulai tahun 2016 hingga 2023 pada statistik pariwisata Provinsi Jawa Timur.
2. Pengolahan data mencakup proses pemilihan data mentah yang mengandung informasi yang tidak dibutuhkan dalam penelitian sehingga perlu di proses Kembali menjadi data set yang akan menjadi objek penelitian.
3. Transformasi data berupa perubahan format data hingga penyusunan data untuk kemudiann di proses secara manual menggunakan excel atau menggunakan tool aplikasi Rapidminer menggunakan algoritma k-means clustering.
4. Data yang telah diproses kemudian di evaluasi menggunakan dan disimpulkan menggunakan metode *Knowledge Discovery in Databases* (KDD) hingga menghasilkan informasi dan pengetahuan baru.

## III. HASIL DAN PEMBAHASAN

### 3.1 Perhitungan Manual Menggunakan Excel dengan K-Means Clustering

Pengolahan data dimulai dari menentukan jumlah klaster yang akan dibentuk. Ada 2 klaster yang akan di bentuk dengan kondisi jumlah klaster harus lebih sedikit dari jumlah data ( $k < n$ ).

Negara	2016	2017	2018	.....	2023
Brunei	2893	3568	1543	.....	1096
Darussalam	2844	3206	2711	.....	1858
Philippina	23382	24898	25692	.....	23832
Thailand	7274	9507	9175	.....	3969

Vietnam	1059	1326	1698	.....	1550
Hongkong	5824	6208	4245	.....	4056
India	4837	6107	6081	.....	3805
Jepang	5866	6381	5884	.....	3403
Korea Selatan	3856	5243	4707	.....	118
Pakistan	211	248	216	.....	194
Bangladesh	377	439	565	.....	143
Srilanka	204	213	269	.....	4110
Taiwan	13550	14168	9096	.....	41579
Tiongkok	16648	17511	18194	.....	4056
.....	.....	.....	.....	.....	.....
Afrika Lainnya	1640	935	558	.....	189

**Tabel 1.** Data Kunjungan Wisatawan berdasarkan negara dan tahun

Dari data awal yang ditampilkan akan tercipta 2 kluster ( $k=2$ ) dan dalam penentuan centroid awal ditentukan secara *random* oleh peneliti ataupun secara manual, pemilihan data awal ini akan sangat menentukan perhitungan centroid selanjutnya. Menggunakan rumus *Euclidean Distance* seperti berikut :

Data 1

Brunai Darussalam 2016, Centroid 1 2016 =

Centroid	2016	2017	.....	2023
Thailand (C1)	7274	9507	.....	3969
Norwegia (C2)	178	227	.....	96

**Tabel 2.** Centroid Awal

Pengenalan centroid pada Tabel 2 adalah C1 (7274,9507,9175,7058,712,4,1178,3969) dan C2 (178,227,109,108,51,0,38,96). Untuk iterasi selanjutnya bila terdapat data yang berpindah kluster, maka tentukan kembali nilai centroid baru yang dihitung dengan menggunakan nilai rata-rata pada setiap kluster. Iterasi atau perhitungan akan dihentikan apabila data yang muncul pada perhitungan baru dan sebelumnya tidak berpindah kluster.

Setelah ditetapkan nilai centroid awal selanjutnya lakukan perhitungan jarak menggunakan rumus *Euclidean Distance* untuk menentukan jarak antara setiap titik centroid dengan titik pada tiap objek, yaitu jarak garis lurus yang biasa antara dua titik dalam ruang Euclidean.

Negara	C1	C2
Singapura	√	
Thailand	√	
India	√	
Jepang	√	
Korea Selatan	√	
Taiwan	√	
Tiongkok	√	
Amerika Serikat	√	
Philippina		√
Prancis		√
Jerman		√
Inggris		√
Australia		√
Hongkong		√
Belanda		√
.....	.....	√
Afrika Lainnya		√

**Tabel 4.** Hasil Kluster

Dari 50 negara diperoleh hasil anggota kluster 1 sebanyak 8 negara dan kluster 2 sebanyak 42 negara. Kemudian dilakukan pencarian centroid yang baru kembali dengan menggunakan rumus  $\sum \frac{a_i}{n}$  yakni total nilai data pada setiap kluster dibagi dengan banyak anggota kluster.

Centroid	2016	2017	...	2023
Thailand (C1)	10190,5	11277,875	...	11320,125
Norwegia (C2)	925,785714285714	988,6429	...	578,857142857143

**Tabel 5.** Centroid Ke 2

Hitung kembali jarak antar centroid menggunakan rumus *Euclidean Distance* pada tiap objek hingga tiap anggota kluster tidak mengalami perubahan kluster.

Negara	C1	C2
Singapura	√	
Thailand	√	
Taiwan	√	
Tiongkok	√	
Amerika Serikat	√	
India		√
Jerman		√
Inggris		√
Philippina		√
Prancis		√

Jepang	√
Belanda	√
Korea Selatan	√
Australia	√
Brunei	√
Darussalam	√
.....	√
Afrika Lainnya	√

**Tabel 6.** Hasil Klaster Dengan Centroid Baru

Karena jumlah klaster pada iterasi ke 2 mengalami perubahan maka hitung kembali nilai centroid menggunakan rumus *Euclidean Distance* kembali hingga memunculkan nilai centroid ke 3 seperti pada table berikut.

Centroid	2016	2017	....	2023
(C1)	13393	14498,4	....	15859,4
(C2)	1187,6	1316,75555	....	790,5777

**Tabel 7.** Centroid Ke 3

Lakukan iterasi ke 3 pada data yang sama hingga memunculkan klaster terbaru dengan menggunakan nilai centroid ke 3 hingga keluarlah klaster baru .

Negara	C1	C2
Singapura	√	
Taiwan	√	
Tiongkok	√	
Thailand		√
Amerika Serikat		√
Brunei		√
Darussalam		√
Philippina		√
Vietnam		√
Hongkong		√
Korea Selatan		√
Jepang		√
India		√
Srilanka		√
Bangladesh		√
Pakistan		√
.....	.....	√
Afrika Lainnya		√

**Tabel 8.** Hasil Klaster Dengan Centroid Ke 3

Karena jumlah klaster pada iterasi ke 3 mengalami perubahan maka hitung kembali nilai centroid menggunakan rumus *Euclidean Distance*

kembali hingga memunculkan nilai centroid ke 4 seperti pada table berikut.

Centroid	2016	2017	....	2023
(C1)	17860	14498,4	....	15859,4
(C2)	1421,851	1316,755	....	790,5777

**Tabel 9.** Centroid Ke 4

Setelah dilakukan iterasi sebanyak 4 kali dengan nilai centroid ke 4 dan centroid ke 5 tidak mengalami perubahan diperoleh hasil data anggota klaster 1 sebanyak 3 negara yakni Singapura, Taiwan dan Tiongkok dan klaster 2 sebanyak 47 negara yang di dapat dilihat dari tabel 10.

Negara	C1	C2
Singapura	√	
Taiwan	√	
Tiongkok	√	
Brunei		√
Darussalam		√
Philippina		√
Thailand		√
Vietnam		√
Hongkong		√
India		√
Jepang		√
Korea Selatan		√
Pakistan		√
Bangladesh		√
Srilanka		√
Asia Lainnya		√
.....	.....	√
Afrika Lainnya		√

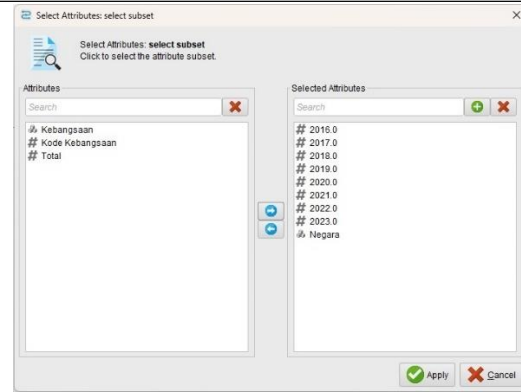
**Tabel 10.** Hasil Klaster Dengan Centroid 4

Setelah dilakukan penghitungan manual dengan menggunakan Excel kemudian akan dilakukan percobaan k-means *clustering* menggunakan tool Rapidminer pada penelitian ini, algoritma k-means *clustering* akan diimplementasikan menggunakan tools RapidMiner (Altair AI Studio) versi 2024.0.1 Dataset digunakan terdiri dari 50 negara dengan kunjungan sejak tahun 2016 hingga tahun 2023. Langkah-langkah dalam proses ini meliputi:

1. Pemilihan Data

Pemilihan data akan menggunakan seleksi data pada Rapidminer. Tujuannya adalah untuk

menentukan atribut dari data set yang akan digunakan pada proses selanjutnya sehingga data yang tidak diperlukan pada penelitian dapat disingkirkan dan tidak melalui proses selanjutnya. Operator Select Attributes merupakan fungsi yang akan digunakan pada proses seleksi data. Operator Select Attributes sendiri memiliki fungsi untuk melakukan proses pemilihan atribut dari data set yang digunakan. Untuk memilih beberapa atribut, gunakan subset pada parameter dari operator Select Atribut. Dataset kunjungan wisatawan mancanegara ke jawa timur berjumlah 12 atribut. Awal proses seleksi data akan milih atribut pada data yang nantinya akan muncul 9 atribut data yang digunakan dalam penelitian ini.



**Gambar 2.** Operator Select Attributes

Setelah dataset diberikan fungsi *select attributes* maka data set yang akan kita gunakan juga berubah dan menjadi data set yang baru dengan tabel dan data yang telah di pilih pada operator *select attribute* seperti pada Gambar 3 berikut :

No	Nama Atribut	Keterangan
1	Kode Kebangsaan	Tidak Digunakan
2	Kebangsaan	Tidak Digunakan
3	Negara	Digunakan
4	2016	Digunakan
5	2017	Digunakan
6	2018	Digunakan
7	2019	Digunakan
8	2020	Digunakan
9	2021	Digunakan
10	2022	Digunakan
11	2023	Digunakan
12	Total	Tidak Digunakan

**Tabel 7.** Atribut Data Set

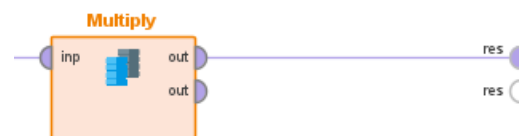
Dari table 7 diketahui bahwa peneliti hanya menggunakan 9 atribut data dari total 12 atribut pada dataset yaitu atribut Negara, 2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023. Melalui data dari dari Tabel 7, akan diterapkan pada pemilihan data pada operator Select Attributes dengan cara memindahkan data yang berada pada kolom kiri ke kolom sebelah kanan seperti Gambar 2 berikut:

Row No.	Negara	2016.0	2017.0	2018.0	2019.0	2020.0	2021.0	2022.0	2023.0
1	Brunei Darus	2893	3568	1543	1234	188	0	412	1096
2	Philippna	2844	3205	2711	2679	348	19	843	1858
3	Singapura	23382	24898	29892	27751	3958	28	9288	23832
4	Thailand	7274	9507	9175	7958	712	4	1178	3889
5	Vietnam	1059	1325	1898	2005	253	1	838	1555
6	Hongkong	5824	6208	4245	1315	1	0	0	1
7	India	4837	6107	6081	5769	792	4	1653	4056
8	Jejang	5866	6381	5884	5721	782	97	892	3805
9	Korea Selatan	3856	5243	4707	4881	833	84	760	3403
10	Palندان	211	248	216	173	35	0	39	118
11	Bangladesh	377	459	585	637	108	0	95	194
12	Sri Lanka	204	213	289	245	80	0	39	143
13	Taiwan	13550	14388	9096	9247	1256	12	394	4110

**Gambar 3.** Data Set Penggunaan Operator Select Attributes

## 2. Transformasi Data

Langka transformasi data dilakukan di Rapidminer dengan memanfaatkan operator *multiply* untuk menciptakan tiruan objek pada aplikasi RapidMiner. Operator ini memngut objek melalui *port input* dan membawa tiruannya ke *port output*. Setiap *port* yang terkoneksi akan membuat salinan yang *independen*. Apabilah mengubah satu tiruan tidak akan berpengaruh pada tiruan yang lainnya.

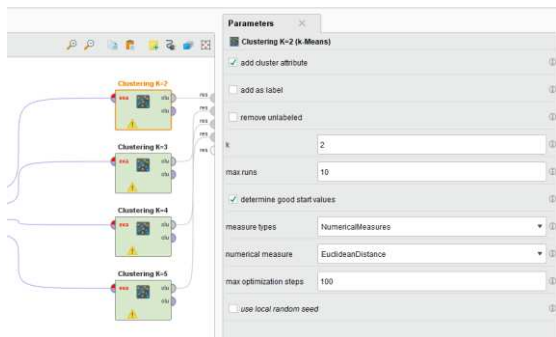


**Gambar 4.** Data Set Penggunaan Operator Filter Examples

## 3. Pengelolaan Data atau Data Mining

Penggunaan Algoritma k-means clustering digunakan pada Rapidminer dengan memunculan operator k-means yang dapat ditemukan pada folder

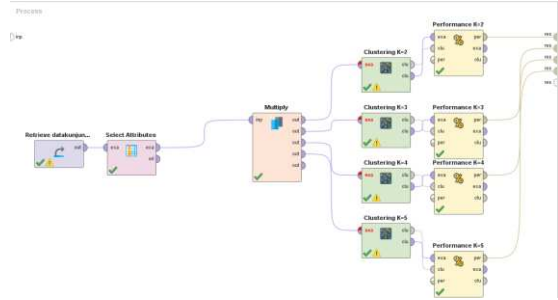
operator *Modeling* lalu pilih *Segmentation* dan tarik operator k-means pada jendela *Process*. Algoritma k-means *clustering* dipakai untuk membuat pengelompokan pada dataset wisatawan asing yang berkunjung ke Jawa Timur dengan jumlah kluster terpilih menggunakan nilai DBI (*Davies Bouldin Index*) teroptimal. Gunakan nilai parameter jumlah kelompok sebanyak 2 ( $k = 2$ ) dengan nilai iterasi sebanyak 10 kali, dibuat pula perbandingan dengan menggunakan kluster 3 ( $k=3$ ) hingga kluster 5 ( $k=5$ ) sebagai perbandingan kluster terbaik dari ke 4 kluster tersebut. Gunakan parameter *measure types* dan terapkan *Numerical Measure* dengan menggunakan pilihan *Euclidean Distance*.



**Gambar 5.** Operator K-Means Clustering di Rapidminer

#### 4. Interpretasi atau Evaluasi (*Interpretation/Evaluation*)

Pada proses interpretasi penelitian ini yakni menggunakan operator *Performance* atau (*Cluster Distance Performance*) karena k-means *clustering* termasuk pengelompokan berdasarkan *distance* atau jarak maka kita gunakan *cluster distance performance* untuk mengukur kualitas kluster kita berbasis jarak. Parameter Operator ini memiliki 2 jenis kriteria, yaitu *average within centroid distance* dan juga *Davies Bouldin Index* (DBI). Di tahap evaluasi untuk model penelitian ini digunakan kriteria utama *Davies Bouldin Index* (DBI) dengan parameter *maximize* untuk menghasilkan nilai DBI non-negatif ( $> = 0$ ).



**Gambar 6.** Proses Davies Bouldin Index (DBI) di Rapidminer

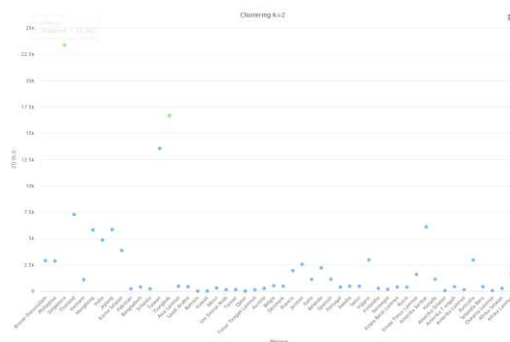
Setelah melakukan 10 percobaan menggunakan *tools* Rapidminer untuk mencari angka DBI paling optimal, dengan permulaan penelitian menerapkan kelompok  $K=2$  hingga  $K=5$  diperoleh kesimpulan bahwa kluster  $K=2$  mempunyai nilai DBI terkecil yakni 0.293 seperti pada tabel berikut.

No	Jumlah Kluster	Nilai DBI
1	K=2	0.293
2	K=3	0.469
3	K=4	0.310
4	K=5	0.310

**Tabel 8.** Nilai DBI Pada Setiap Kluster

#### 5. Hasil Evaluasi

Dari penelitian ini dapat disimpulkan bahwa klusterisasi wisatawan mancanegara di Provinsi Jawa Timur dapat dikelompokkan menjadi 2 kluster menggunakan k-means *clustering*. Klusterisasi data wisatawan mancanegara di Provinsi Jawa Timur di proyeksikan dengan kluster 0 sebanyak 2 anggota dan kluster 1 sebanyak 48 anggota seperti pada gambar berikut.



**Gambar 7.** Hasil Klusterisasi di Rapidminer

### IV. KESIMPULAN DAN SARAN

#### 4.1 KESIMPULAN

Pada penelitian ini ditemukan 2 klaster wisatawan mancanegara yang berkunjung ke Jawa Timur klaster 1 diisi oleh 3 negara dan klaster 2 diisi oleh 47 negara dengan perhitungan menggunakan Excel. Dari klasterisasi Dataset wisatawan asing yang berkunjung ke Jawa Timur melalui bandara Juanda disimpulkan ke dalam 2 klaster dengan klaster 1 berisi negara-negara dari Asia seperti Tiongkok, Taiwan, Singapura sedangkan untuk klaster ke 2 diisi negara dari Eropa, Asia, Timur Tengah dan Amerika. Pada klaster ke 2 dapat di jadikan acuan data sebagai target promosi pariwisata Jawa Timur sehingga lebih banyak negara-negara dari klaster 2 yang mengunjungi Jawa Timur.

#### 4.2 SARAN

Dari hasil penelitian telah disimpulkan jumlah klaster dan anggota klaster pada masing-masing klaster. Data anggota klaster tersebut dapat menjadi data yang mampu membantu pariwisata Jawa Timur untuk menentukan promosi seperti apa yang harus dilakukan serta target promosi seperti apa yang akan di pilih dalam rangka meningkatkan kunjungan wisatawan mancanegara ke Jawa Timur. Selain itu hasil penelitian dapat menjadi rujukan bagia penelitian lainnya dengan metode yang berbeda sehingga mampu menghasilkan data yang lebih bermanfaat untuk kedepannya.

#### REFERENSI

- [1] M. Tobing, “Pengaruh Jumlah Obyek Wisata, Tingkat Penghunian Kamar, Dan Jumlah Kunjungan Wisatawan Terhadap Pendapatan Asli Daerah Kabupaten Simalungun,” *Jurnal Ekuilnomi*, vol. 3, no. 2, pp. 127–139, 2021, doi: 10.36985/ekuilnomi.v3i2.263.
- [2] D. Yanti, I. Ramadhan, D. Yunita, and M. R. Lubis, “Peran Media Sosial Dalam Mempromosikan Pariwisata Di Desa Perkebunan Bukit Lawang,” *Jurnal Pariwisata*, vol. 11, no. 1, pp. 1–13, 2024, doi: 10.31294/par.v11i1.21228.
- [3] R. Njatrijani, “Kearifan Lokal Dalam Perspektif Budaya Kota Semarang,” *Gema*

*Keadilan*, vol. 5, no. 1, pp. 16–31, 2018, doi: 10.14710/gk.2018.3580.

- [4] R. R. Aji, R. W. D. Pramono, and D. H. Rahmi, “Kontribusi Sektor Pariwisata Terhadap Ekonomi Wilayah Di Provinsi Jawa Timur,” *Jurnal Planoearth*, vol. 3, no. 2, p. 57, 2018, doi: 10.31764/jpe.v3i2.600.
- [5] P. Priyanto, “Pengembangan Potensi Desa Wisata Berbasis Budaya Tinjauan Terhadap Desa Wisata Di Jawa Tengah,” *Jurnal Vokasi Indonesia*, vol. 4, no. 1, 2016, doi: 10.7454/jvi.v4i1.53.
- [6] R. C. Prihandari, “Data Mining: Konsep Dan Aplikasi Menggunakan Rapidminer (Series: Supervised Learning Dan Unsupervised Learning),” p. 8, 2022.
- [7] I. Ahmad, S. Samsugi, and Y. Irawan, “Implementasi Data Mining Sebagai Pengolahan Data,” *Jurnal Teknoinfo*, vol. 16, no. 1, p. 46, 2022.
- [8] M. Iqbal, “Klasterisasi Data Jamaah Umroh Pada Auliya Tour & Travel Menggunakan Metode K-Means Clustering,” *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 5, no. 2, pp. 97–104, 2019, doi: 10.33330/jurteks.v5i2.352.
- [9] N. Wakhidah, “Clustering Menggunakan K-Means Algorithm ( K-Means Algorithm Clustering ),” *Fakultas Teknologi informasi*, vol. 21, no. 1, pp. 70–80, 2014.
- [10] X. Ling, Q. Tu, M. Jin, W. Wang, Y. Cui, and J. Zhu, “Research on Distributed Resource Aggregation Technology Based on Hierarchical Agglomerative Clustering Analysis,” *Journal of Physics: Conference Series*, vol. 2399, no. 1, 2022, doi: 10.1088/1742-6596/2399/1/012034.
- [11] A. R. Hedar, A. M. M. Ibrahim, A. E. Abdel-Hakim, and A. A. Sewisy, “K-means cloning: Adaptive spherical K-means clustering,” *Algorithms*, vol. 11, no. 10, pp. 1–21, 2018, doi: 10.3390/a11100151.

- [12] S. Butsianto and N. Saepudin, “Penerapan Data Mining Terhadap Minat Siswa Dalam Mata Pelajaran Matematika Dengan Metode K-Means,” *Jurnal Nasional Komputasi dan Teknologi Informasi (JNKTI)*, vol. 3, no. 1, pp. 51–59, 2020, doi: 10.32672/jnkti.v3i1.2008.