

KLASTERISASI KARAKTERISTIK WISATAWAN MANCANEGERA MENGGUNAKAN METODE K-MEANS CLUSTERING

Annisa Agustin Mahardika¹, Eka N Kencana^{2§}, I Komang Gde Sukarsa³,
Ketut Jayanegara⁴, IGN Lanang Wijayakusuma⁵, I Wayan Sumarjaya⁶

^{1,3,4,5,6}Program Studi Matematika, Fakultas MIPA – Universitas Udayana
Email: ¹mannisaagustin@gmail.com, ³gedesukarsa@unud.ac.id, ⁴ktjayanegara@unud.ac.id,
⁵lanang_wijaya@unud.ac.id, ⁶sumarjaya@unud.ac.id

²Kelompok Studi Sosiometrika – Universitas Udayana
Email: i.putu.enk@unud.ac.id

[§]Corresponding Author

ABSTRACT

Since the Covid-19 pandemic, Indonesian tourism has experienced a drastic decline. This decline can be seen in the number of foreign tourists visiting Indonesia. The number of foreign tourist arrivals in 2020 and 2021 is far less compared to 2019 before Covid-19 entered. As a result, the Indonesian economy also suffered. Regarding the recovery of Indonesian tourism after the pandemic has been slow down, this study aims to cluster foreign tourists visiting Indonesia based on the amount of their expenditures and length of stays using the K-means algorithm. Secondary data from National Statistics Bureau classified the origin of tourists were 86 countries. Applying k-means algorithm methods to cluster country of origin, result showed they were three clusters formed based on the attributes of visiting, i.e. length of stay in Indonesia and total amount of their expenditures. Each cluster consists of 14, 54 and 18 countries. The first cluster is characterized by countries that have high tourism spending; the second cluster is formed by countries with moderate tourism spending; and the third cluster is characterized by countries with low tourism spending. The accuracy of the three clusters in explaining the variance of tourist spending is 68.8 percent.

Keywords: clustering, Covid-19, foreign tourists, k-means algorithm.

1. PENDAHULUAN

Merujuk pasal 1 ayat (1) Undang-Undang Republik Indonesia Nomor 10 Tahun 2009, wisata didefinisikan sebagai kegiatan perjalanan yang dilakukan seseorang atau sekelompok orang ke tempat tertentu yang bertujuan untuk rekreasi, pengembangan kepribadian, atau untuk mempelajari keunikan daya tarik wisata yang dikunjungi dalam jangka waktu sementara. Ayat (2) dari pasal ini menyatakan yang melakukan wisata disebut wisatawan (RI, 2009).

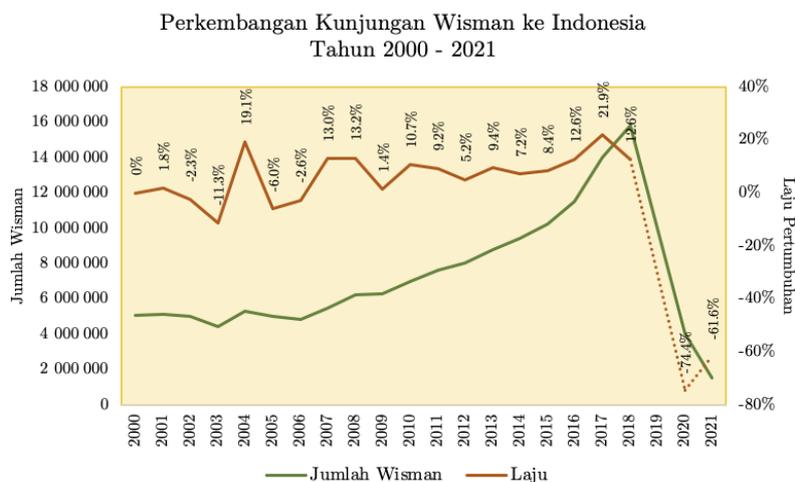
Pariwisata saat ini merupakan salah satu subyek riset menarik bagi peneliti dari berbagai disiplin ilmu. Pariwisata pada umumnya diteliti dari dua sisi pandang, sisi permintaan (*demand side*) dan sisi penawaran berwisata (*supply side*). Pada sisi permintaan, Badan Pusat Statistik (BPS) mendefinisikan tiga kategori wisatawan, yaitu: (a) wisatawan mancanegara (wisman),

yaitu setiap orang yang melakukan perjalanan ke negara di luar negara tempat tinggalnya dalam kurun waktu maksimal 12 bulan dengan tujuan bukan untuk bekerja; (b) wisatawan nusantara (wisnus), didefinisikan sebagai warga negara Indonesia yang melakukan perjalanan di dalam wilayah teritorial Indonesia untuk kurun waktu kurang dari enam bulan dengan tujuan bukan untuk bekerja atau bersekolah; dan (c) wisatawan nasional (wisnas), warga negara Indonesia yang berwisata ke negara lain dalam kurun waktu maksimum 12 bulan (Badan Pusat Statistik (BPS), 2019).

Sebagai salah satu fondasi pembangunan nasional, pariwisata Indonesia berperan penting dalam mendorong perekonomian nasional. Pada Laporan Kinerja Kementerian Pariwisata Tahun 2019, tercatat kontribusi sektor pariwisata pada Produk Domestik Bruto (PDB) sebesar Rp 280

triliun, sekitar 5,5 persen dari PDB nasional. Sayangnya, peran sektor pariwisata mengalami kemunduran sejak Indonesia mulai terdampak pandemi COVID-19. Kasus pertama COVID-19 di Indonesia diumumkan pada tanggal 2 Maret 2020, 10 hari sebelum Badan Kesehatan Dunia (WHO) menetapkan COVID-19 sebagai kasus pandemik (Meriani et al., 2022; Susilo et al., 2020). Pada akhir tahun 2019, total kunjungan wisman ke Indonesia tercatat mencapai 16,1 juta

wisatawan, meningkat dari jumlah kunjungan pada tahun 2018 sebesar 15,8 juta. COVID-19 menyebabkan Indonesia serta negara-negara sumber utama wisatawan dan tujuan berwisata menutup perbatasan negara masing-masing, dan berdampak pada jatuhnya jumlah kunjungan wisman pada tahun 2020 yang tercatat hanya sekitar 4 juta orang. Gambar 1 menunjukkan perkembangan jumlah kunjungan wisman ke Indonesia pada periode tahun 2000 – 2021.



Gambar 1. Perkembangan Jumlah Kunjungan Wisman, Tahun 2000 – 2021
Sumber: Kementerian Pariwisata dan Ekonomi Kreatif (2022)

Gambar 1 menunjukkan, pada periode tahun 2019 – 2021, kunjungan wisman ke Indonesia ‘terjun bebas’ hingga tercapai laju pertumbuhan negatif tertinggi dalam sejarah kepariwisataan Indonesia. Pada periode tahun 2019 – 2020 laju pertumbuhan kunjungan tercatat -74,4 persen, lebih rendah dibandingkan laju yang tercatat saat peristiwa bom Bali I yang terjadi pada tanggal 12 Oktober 2002 yang berdampak pada turunnya jumlah kunjungan wisman di tahun berikutnya.

Sebagai salah satu fondasi perekonomian nasional, menurunnya kunjungan wisman juga dirasakan di berbagai daerah, khususnya yang menitikberatkan sektor pariwisata sebagai sektor unggulan daerah. Sebagai misal, pertumbuhan Produk Domestik Regional Bruto (PDRB) Bali tercatat negatif sebesar -9,33 persen, penurunan terdalam dibandingkan dengan penurunan yang dialami oleh 33 provinsi lain di Indonesia (BPS Provinsi Bali, 2022).

Seiring berkurangnya dampak dari pandemi COVID-19, triwulan II tahun 2022 merupakan titik awal dari pemulihan kunjungan wisman ke berbagai destinasi di Indonesia. Sebagai sentra pariwisata Indonesia, jumlah kunjungan wisman

ke Bali pada April 2022 tercatat 58335 orang, meningkat hampir 400 persen dari kunjungan pada Maret 2022 sejumlah 14620 orang (BPS Provinsi Bali, 2022). Pemulihan kepariwisataan Indonesia seyogyanya diikuti dengan tindakan penguatan destinasi, meningkatkan mutu industri, mengaktifkan promosi, dan meningkatkan kapasitas kelembagaan pariwisata.

Salah satu tindakan untuk meningkatkan efektivitas promosi kepariwisataan Indonesia adalah dengan melakukan segmentasi yang tepat (*segmenting*) dari pasar wisatawan mancanegara yang berkunjung. Segmentasi pasar wisman di Indonesia, pada ranah Statistika, dapat dilakukan dengan melakukan klasterisasi wisman menurut sejumlah karakteristik atau atribut penciri dari wisman. Hingga saat ini aplikasi analisis klaster pada kepariwisataan Indonesia lebih ditujukan pada perspektif penawaran wisata.

Savitri *et al.* (2021) menerapkan teknik *K-means clustering* untuk mengklasterisasi 42 objek wisata di Kabupaten Kulon Progo yang diduga akan berkembang seiring dibangunnya Bandara New Yogyakarta International Airport (NYIA) di kabupaten ini. Penelitian Savitri *et al.*

menunjukkan terdapat tiga klaster objek wisata, yaitu: (a) klaster jumlah pengunjung rendah beranggotakan 20 objek; (b) klaster jumlah pengunjung sedang beranggotakan 15 objek; dan (c) klaster jumlah pengunjung tinggi sebanyak dua objek, yaitu Banjaroyo dan Pantai Glagah.

Riset yang dilakukan Iswandhani & Muhajir (2018) yang ditujukan untuk mengklasterisasi 126 destinasi Yogyakarta menghasilkan terdapat tiga klaster, masing-masing beranggotakan 52, 9, dan 60 destinasi yang diteliti. Atribut destinasi yang digunakan melakukan klasterisasi adalah kepopuleran destinasi di media sosial instagram dengan teknik diaplikasikan *K-means clustering*. Riset yang dilakukan Iswandhani & Muhajir juga mengidentifikasi 10 destinasi dengan tingkat kepopuleran tertinggi, dua diantaranya, Candi Sewu dan Candi Prambanan, tergolong sebagai anggota dari klaster kepopuleran tinggi.

Salah satu riset sebelumnya yang ditujukan mengelompokkan wisatawan yang berkunjung ke Manado di Provinsi Sulawesi Utara oleh Koyongian *et al.* (2021). Menggunakan analisis klaster, karakteristik demografi dan psikografi dari wisman digunakan Koyongian *et al.* dalam mengidentifikasi empat klaster yang disebutnya sebagai *like-everything tourist, future optimistic adventurer, food lover adventurer, dan socializer adventurer*. Analisis karakteristik demografi wisman menunjukkan tidak terdapat bukti bahwa gender, umur, jenis pekerjaan, dan jumlah kunjungan ke Manado sebagai penyebab dari perbedaan antarklaster. Perbedaan disebabkan oleh karakteristik psikografi, meliputi aktivitas dan minat berkunjung serta opini wisman pada kemenarikan Kota Manado sebagai destinasi wisata.

Memperhatikan terbatasnya publikasi yang diarahkan untuk melakukan segmentasi wisman, artikel ini ditujukan melengkapi keterbatasan tersebut dengan melakukan segmentasi dari pasar wisatawan melalui klasterisasi wisman yang berkunjung ke Indonesia. Sebagai teknik analisis peubah ganda (Hair *et al.*, 2019), analisis klaster (*cluster analysis*) merupakan jenis analisis yang ditujukan memberikan informasi tambahan dari matriks data $X_{n \times p}$ dengan n menyatakan jumlah observasi dan p jumlah atribut. Informasi yang diinginkan adalah melakukan penyekatan n objek ke dalam k kelompok (klaster) menurut p atribut. Kriteria penyekatan adalah memaksimumkan kemiripan n_i objek pada klaster ke- i ; $i = 1, \dots, k$ dan memaksimumkan ketakmiripan objek-objek pada klaster ke- i dengan klaster ke- j ; $i \neq j$ (Li &

Wu, 2012). Menyitir Kencana (2020), klasterisasi diarahkan untuk membentuk k grup ($k \lll n$) sehingga keragaman intragrup (*within group*) minimum; dan keragaman antargrup, disebut separasi eksternal, maksimum. Seperti teknik analisis kuantitatif lain yang tidak pernah sempurna, kelemahan analisis klaster terletak pada tidak adanya teknik untuk memvalidasi relevansi dari atribut yang digunakan untuk membentuk klaster (Kencana & Darmayanti, 2022, p. 56).

Menyitir Izenman (2008) dan Hair *et al.* (2019), terdapat 2 teknik klasterisasi yaitu (a) tak berhirarki, dan (b) berhirarki. Teknik klaster *K-means clustering* adalah teknik klasterisasi tak berhirarki, dan tergolong teknik klasterisasi yang masif diaplikasikan mengingat kesederhanaan algoritmanya. Kassambara (2017) menyatakan tahapan pembentukan klaster pada *K-means clustering* sebagai berikut:

1. Menentukan jumlah klaster k yang diinginkan dengan alasan-alasan yang rasional;
2. Memilih k observasi secara acak sebagai pengawalan dari pusat klaster (*centroid*), juga nilai pusat klaster (Han & Kamber, 2011);
3. Mendistribusikan $(n - k)$ observasi tersisa ke *centroid* terdekat menurut jarak observasi ke *centroid*;
4. Meng-*update centroid* dengan menghitung nilai rata-rata klaster yang baru. Rataan baru dihitung menggunakan $\forall x \in k_j; j = 1, \dots, k$. x merupakan observasi anggota dari k_j ;
5. Mengiterasi (2) hingga (4) hingga jumlah kuadrat intraklaster (*within sum square/WSS*) minimum atau jumlah iterasi maksimum telah tercapai.

Dalam menentukan k pada tahap (1), meski ditetapkan *a priori* melalui sejumlah alasan rasional, terdapat 3 metode yang biasa digunakan sebagai indikator dari jumlah klaster yang layak dibentuk. Ketiga metode ini adalah (a) membuat plot antara jumlah klaster pada sumbu horizontal dengan total WSS pada sumbu vertikal; (b) plot antara jumlah klaster pada sumbu horizontal dengan nilai rata-rata lebar siluet (Rousseeuw, 1986) pada sumbu vertikal; dan (c) menghitung statistik *Gap* (Tibshirani *et al.*, 2001).

2. METODE PENELITIAN

2.1 Sumber Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini bersumber dari publikasi Statistik Pengeluaran Wisatawan Mancanegara 2019 (BPS Indonesia, 2020), dan variabel yang digunakan melakukan klasterisasi wisatawan mancanegara ditunjukkan pada Tabel 1.

Tabel 1. Variabel Penelitian

No	Kode	Deskripsi Ringkas	Satuan	Skala
1	ORG	Negara asal	-	Nominal
2	ACC	Akomodasi	US\$	Rasio
3	FnB	Makan dan minum	US\$	Rasio
4	TRA	Transportasi lokal	US\$	Rasio
5	TOU	Paket wisata lokal	US\$	Rasio
6	OTH	Belanja lain	US\$	Rasio
7	LoS	Lama tinggal di Indonesia	Hari	Rasio

2.2 Tahapan Kegiatan dan Analisis Data

Pembentukan klaster wisatawan mancanegara (wisman) yang berkunjung ke Indonesia pada tahun 2019 menurut karakteristiknya seperti diperlihatkan pada Tabel 1 dilakukan melalui sejumlah tahapan berikut:

1. Menghitung matriks jarak antar negara asal wisman ($D_{86 \times 86}$). Matriks jarak dihitung dari

matriks data asal pada Tabel 1, $X_{86 \times 7}$, dengan jumlah observasi 86 negara. Jarak antardua negara asal, P_1 dan P_2 pada ruang R^6 dihitung menggunakan Jarak Euclid menurut pers. (1).

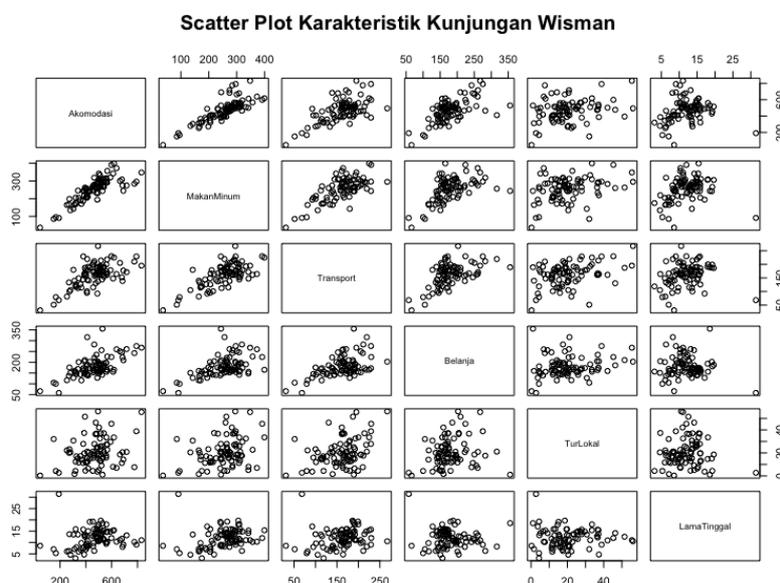
$$d_{P_1, P_2} = \sqrt{\sum_{i=1}^6 (P_{1i} - P_{2i})^2} \quad (1)$$

2. Menduga jumlah klaster awal $k = 3$ sebagai representasi klaster dengan pengeluaran tinggi; klaster dengan pengeluaran menengah; dan klaster dengan pengeluaran rendah;
3. Memvalidasi $k = 3$ dengan membuat *elbow plot* dan menghitung nilai siluet $s(i)$ amatan ke- i (Rousseeuw, 1986) serta menghitung nilai statistik Gap $G_n(k)$ (Tibshirani et al., 2001). Bila terindikasi ada ketaksesuaian antara k dengan hasil validasi, jumlah klaster akhir ditentukan dengan mencermati hasil validasi;
4. Memeriksa dan menginterpretasikan klaster yang terbentuk.

3. HASIL DAN PEMBAHASAN

3.1 Korelasi Antarpasangan Atribut

Seperti ditunjukkan pada Tabel 1, terdapat 6 atribut yang digunakan mengklasterisasi wisman menurut karakteristik kunjungannya. *Scatter plot* antarpasangan atribut diperlihatkan Gambar 2.



Gambar 2. Hubungan Antaratribut Karakteristik Kunjungan Wisman

Gambar 2 memperlihatkan, kecuali dengan atribut TRA dan LoS, antarpasangan atribut lain terlihat memiliki korelasi positif yang bermakna. Pemeriksaan nilai koefisien korelasi Pearson (ρ) antaratribut menunjukkan ρ berkisar antara -0,02 hingga 0,82 seperti diperlihatkan pada Tabel 2.

Tabel 2. Matriks Korelasi Antarpasangan Atribut

	ACC	FnB	TRA	OTH	TOU	LoS
ACC	1,00	0,82	0,64	0,64	0,28	0,02
FnB	0,82	1,00	0,71	0,54	0,34	0,24
TRA	0,64	0,71	1,00	0,56	0,38	0,18
OTH	0,64	0,54	0,56	1,00	0,19	-0,08
TOU	0,28	0,34	0,38	0,19	1,00	-0,02
LoS	0,02	0,24	0,18	-0,08	-0,02	1,00

Sumber: data primer (2023), dianalisis.

Menurut Hinkle *et al.* (2003) yang menyatakan $|\rho| < 0,30$ sebagai korelasi tak bermakna; maka hubungan antara LoS dengan lima atribut lain merupakan hubungan-hubungan yang tak bermakna; TOU dengan ACC dan OTH memiliki koefisien korelasi tak bermakna; dan korelasi antarpasangan lain adalah rendah, moderat, atau korelasi tinggi yang teramati pada $\rho_{ACC;FnB}$ dan $\rho_{FnB;TRA}$. Pemahaman mengenai struktur matriks korelasi antarpasangan atribut yang digunakan sangat membantu dalam memahami kluster yang terbentuk.

3.2 Pemeriksaan Jumlah Kluster Awal

Pemeriksaan kesesuaian jumlah kluster awal $k = 3$ dilakukan dengan membuat *elbow plot*, plot siluet, dan statistik *Gap*. *Elbow plot* memetakan jumlah kluster yang diduga dengan nilai total kuadrat intra kluster (*total within sum-squares* atau Total WSS), dan dihitung menggunakan pers. (2).

$$\text{Total WSS} = \sum_{i=1}^k \sum_{j=1}^{n_k} (x_{ij} - \bar{x}_i)^2 \quad (2)$$

Jumlah kluster yang dianggap layak adalah saat plot mulai ‘melandai’.

Identifikasi jumlah kluster yang dianggap layak menggunakan plot siluet dilakukan dengan memetakan jumlah kluster dengan nilai siluet $s(i)$ dihitung dengan pers. (3) (Reddy & Vinzamuri, 2014).

$$s(i) = \frac{\sum_{i=1}^N \frac{b_i - a_i}{\max(a_i, b_i)}}{N} \quad (3)$$

Pada pers. (3) a_i dan b_i masing-masing menyatakan nilai rata-rata jarak observasi ke- i ke seluruh amatan dalam kluster yang sama, dan b_i menyatakan nilai rata-rata jarak observasi ke- i ke seluruh amatan yang tergabung pada kluster yang lain. Pada metode siluet, jumlah kluster yang dianggap layak adalah saat nilai $s(i)$ maksimum.

Metode ketiga yang digunakan memvalidasi $k = 3$ adalah membuat plot jumlah kluster dengan statistik *Gap*, dan dihitung menggunakan pers. (4) (Reddy & Vinzamuri, 2014):

$$\text{Gap}(k) = \frac{1}{B} \times \sum_b \log(W_b^*(k)) - \log(W(k)) \quad (4)$$

Pada pers. (4), $W_b^*(k)$ menyatakan WSS pada kluster ke- k bila diasumsikan anggota kluster menyebar mengikuti distribusi seragam. Jumlah kluster yang dinilai layak adalah nilai k terkecil yang memenuhi pers. (5):

$$\text{Gap}(k) \geq \text{Gap}(k+1) - s_{k+1} \quad (5)$$

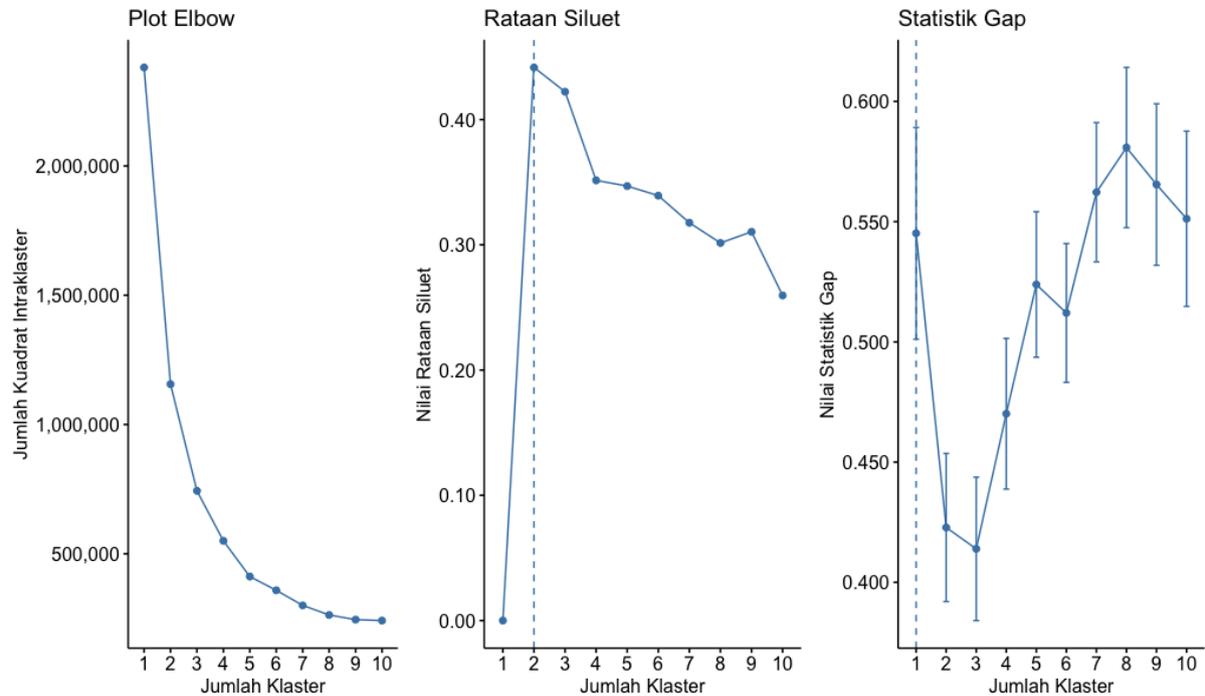
Suku s_{k+1} pada pers. (5) menyatakan simpangan baku dari $W_b^*(k)$. Menggunakan *library factoextra* (Kassambara & Mundt, 2020) pada R 4.1.2, plot dari statistik yang dihitung diperlihatkan pada Gambar 3.

Plot *elbow* mengindikasikan bentuk ‘siku’ mulai terbentuk saat jumlah kluster $k = 4$ dengan Total WSS $\approx 550\,065$. Plot mengindikasikan jumlah kluster lebih banyak dibandingkan dengan dugaan jumlah kluster awal. Pada plot nilai rata-rata siluet, rata-rata tertinggi dijumpai pada $k = 2$; lebih kecil dibandingkan jumlah dugaan awal. Meski demikian, nilai rata-rata siluet untuk $k = 3$ terhitung sebesar 0,422 – lebih kecil 4,5 persen dari nilai $s(2)$ sebesar 0,442. Pemeriksaan statistik *Gap* menghasilkan *Gap* (k) terkecil terdapat pada jumlah kluster $k = 3$ sebesar 0,414. Mencermati hasil verifikasi tentang jumlah kluster yang layak dibentuk pada *K-means clustering* cenderung menuju jumlah kluster $k = 3$, maka klusterisasi karakteristik wisman ke Indonesia pada tahun 2019 dilakukan dengan menetapkan 3 kluster, dengan penciri setiap kluster akan diidentifikasi dengan menggunakan analisis komponen utama pada keenam atribut/karakteristik kunjungan wisman menurut negara asal pada Tabel 1.

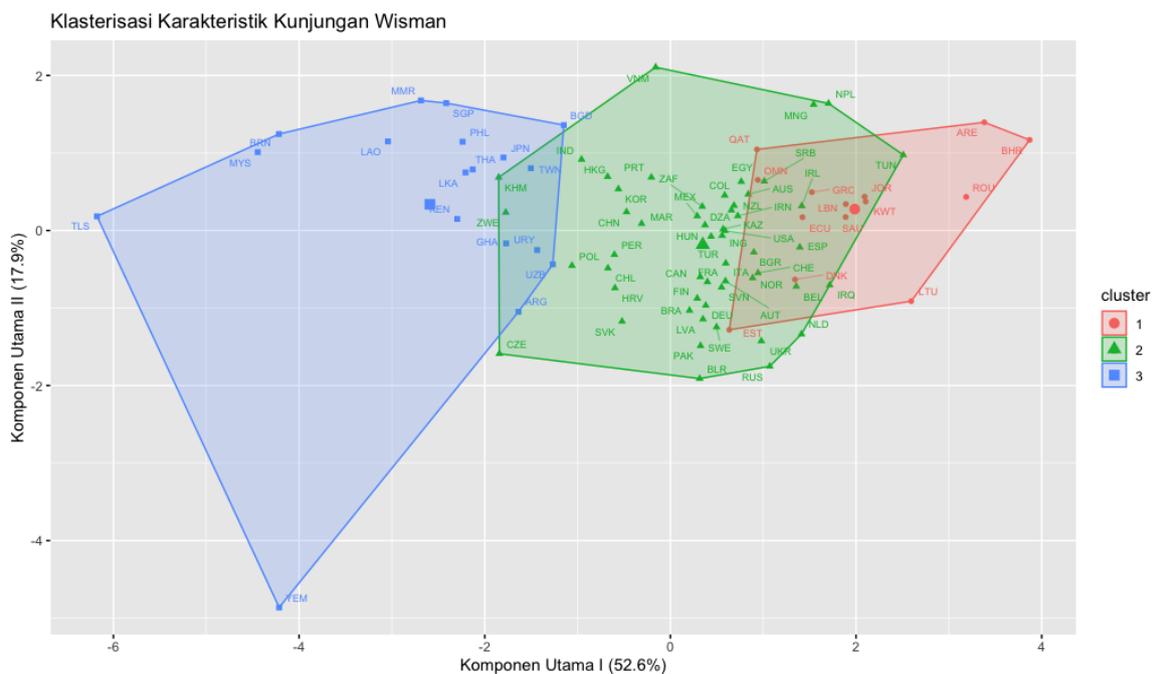
3.3 Klasterisasi Karakteristik Wisman

Ketiga kluster yang dibentuk menggunakan fungsi kmeans pada R 4.1.2 masing-masing beranggotakan 14, 54, dan 18 negara dengan

tingkat akurasi – rasio jumlah kuadrat antarkelas dengan jumlah kudrat total, sebesar 68,8 persen. Gambar 4 memperlihatkan peta ketiga kluster dengan anggota-anggotanya.



Gambar 3. Pemeriksaan Jumlah Kluster yang Layak Dibentuk



Gambar 4. Plot Kluster Menurut Karakteristik Kunjungan Wisman

Plot 2 dimensi pada Gambar 4 memetakan skor faktor masing-masing negara asal wisman dengan sumbu horizontal merupakan komponen

utama (KU) I dan sumbu vertikal merupakan KU II. Kedua komponen utama menjelaskan variansi total atribut yang digunakan pada klasterisasi,

masing-masing 52,6 persen dan 17,9 persen, dan total variansi terjelaskan sebesar 70,5 persen. Plot menunjukkan terjadi *overlapping* pada klaster, mengindikasikan separasi antarklaster belum optimal. Hasil analisis untuk $k = 3$ diperlihatkan pada Tabel 3.

Tabel 3. Centroid dan WSS dari Klaster Terbentuk

	Klaster I	Klaster II	Klaster III
ACC	675,454	477,493	277,449
FnB	322,338	268,696	157,112
TRA	184,697	172,902	109,906
OTH	227,136	183,176	135,546
TOU	23,831	21,983	14,502
LoS	11,544	12,880	9,682
Anggota	14	54	18
WSS	148013	370275	225193

Sumber: data primer (2023), dianalisis.

Hasil telaah Tabel 3 memberikan gambaran Klaster I merupakan klaster yang beranggotakan negara-negara dengan belanja wisata tertinggi; Klaster II beranggotakan negara-negara dengan belanja wisata moderat; dan Klaster III adalah klaster beranggotakan negara-negara dengan belanja wisata rendah.

3.4 Interpretasi Klaster

Pada Klaster I, klaster yang beranggotakan negara asal wisman dengan lima atribut pertama pada Tabel 3 memiliki rata-rata nilai tertinggi. Total jumlah intraklaster (WSS) pada klaster sebesar 148013. Sepuluh besar negara asal wisman yang tergabung ke dalam Klaster I menurut total belanja ditunjukkan pada Tabel 4.

Tabel 4. Sepuluh Besar Negara Asal pada Klaster I

Kode	Negara Asal	Total Belanja*	LoS
BHR	Bahrain	1697,57	11,04
ARE	Uni Emirat Arab	1631,39	9,05
KWT	Kuwait	1520,78	10,23
ROU	Rumania	1489,03	11,98
SAU	Arab Saudi	1443,56	13,17
LTU	Lituania	1440,43	15,42
LBN	Lebanon	1441,52	10,35
JOR	Yordania	1439,23	11,13
ECU	Ekuador	1397,09	11,67
QAT	Qatar	1340,50	8,08

* Total Belanja = ACC + FnB + TRA + OTH + TOU
 Sumber: data primer (2023), dianalisis.

Tabel 4 menunjukkan tujuh dari 10 negara merupakan kelompok negara yang berlokasi di Jazirah Arab. Tiga negara lain berada di kawasan Eropa (Lituania dan Rumania) serta Ekuador berada di kawasan Amerika Latin. Negara asal wisman dengan LoS tertinggi berturut-turut Lituania, Arab Saudi, dan Rumania.

Klaster II yang beranggotakan negara asal wisman dengan belanja wisata moderat memiliki WSS sebesar 370275, dengan 10 besar negara asal wisman pada klaster ini ditunjukkan pada Tabel 5.

Tabel 5. Sepuluh Besar Negara Asal pada Klaster II

Kode	Negara Asal	Total Belanja*	LoS
IRQ	Irak	1318,58	18,55
TUN	Tunisia	1314,86	10,55
NPL	Nepal	1295,32	6,30
BEL	Belgia	1266,26	14,91
NLD	Belanda	1258,60	18,03
IRL	Irlandia	1258,09	11,07
AUS	Australia	1252,94	9,43
BGR	Bulgaria	1234,41	14,14
ESP	Spanyol	1233,53	13,72
NZL	Selandia Baru	1233,22	10,13

* Total Belanja = ACC + FnB + TRA + OTH + TOU
 Sumber: data primer (2023), dianalisis.

Klaster II memperlihatkan Australia sebagai negara asal wisman yang rutin mengunjungi Indonesia memiliki LoS 9,43 hari. Temuan yang menarik adalah posisi Irak di peringkat pertama dari total belanja dan LoS-nya. Sebagai salah satu negara di Asia Tengah yang masih belum pulih sepenuhnya dari dampak perang, posisi Irak di peringkat pertama memerlukan kajian lanjutan. Empat negara lain yang menjadi fokus pemasaran wisata Indonesia – Rusia, Tiongkok, India, dan Korea Selatan bergabung ke dalam klaster ini.

Klaster III, 10 besar negara asal wisman dengan total belanja wisata rendah diperlihatkan pada Tabel 6.

Tabel 6. Sepuluh Besar Negara Asal pada Klaster III

Kode	Negara Asal	Total Belanja*	LoS
BGD	Bangladesh	909,95	5,97
UZB	Uzbekistan	880,51	13,53
TWN	Taiwan	883,40	7,56
JPN	Jepang	836,43	6,94
ARG	Argentina	814,95	15,90
URY	Uruguay	783,94	13,53
KEN	Kenya	779,95	9,77
THA	Thailand	777,14	7,89
LKA	Sri Lanka	777,27	7,12
GHA	Ghana	770,10	13,68

* Total Belanja = ACC + FnB + TRA + OTH + TOU
Sumber: data primer (2023), dianalisis.

Menggunakan data kunjungan wisman tahun 2019, Tabel 6 memperlihatkan Jepang sebagai negara asal wisman yang banyak berkontribusi pada kepariwisataan Indonesia berada di Klaster III. Hal ini diduga tidak terlepas dari menurunnya jumlah kunjungan wisman Jepang dari 573130 pada tahun 2017 menjadi 519623 orang pada tahun 2019 (BPS Indonesia, 2021). Wisman dari negara ASEAN seperti Thailand, Filipina, Brunei Darussalam, Singapura, dan Malaysia teramati juga merupakan anggota pada Klaster III.

4. SIMPULAN DAN REKOMENDASI

4.1 Simpulan

Terkait klasterisasi karakteristik kunjungan wisman ke Indonesia pada tahun 2019 ke dalam tiga klaster, artikel ini menyimpulkan tiga hal berikut:

1. Klaster I yang dicirikan belanja wisata dan lama menginap (LoS) tertinggi beranggotakan 14 negara, yang didominasi oleh wisman yang berasal dari negara-negara di Jazirah Arab;
2. Klaster II yang dicirikan belanja wisata dan LoS moderat beranggotakan 54 negara-negara Eropa, Asia Pasifik, dan Amerika yang selama ini menjadi tujuan promosi pariwisata;
3. Klaster III dengan pencari belanja wisata dan LoS terkecil beranggotakan 18 negara di mana negara anggota ASEAN berada pada klaster ini.

4.2 Rekomendasi

Untuk meningkatkan kualitas klaster dan interpretasi yang diperoleh pada penelitian ini, riset selanjutnya disarankan untuk:

1. Mengelaborasi kemungkinan jumlah klaster yang dibentuk tidak hanya 3 klaster. Mencermati plot *elbow* yang juga melandai untuk $k = 4$ atau $k = 5$; maka kedua nilai k layak untuk diperiksa;
2. Melakukan eksplorasi matriks data dengan menekankan pada pemeriksaan pencilan pada karakteristik wisman. Negara asal wisman yang tergolong pencilan, khususnya pencilan bawah, disarankan untuk dieliminasi sebelum dilakukan klasterisasi;
3. Mengelaborasi teknik klasterisasi yang lain, misalnya K-medoids atau kelas klasterisasi berhirarki seperti klasterisasi aglomeratif.

DAFTAR PUSTAKA

- Badan Pusat Statistik (BPS). (2019). *Neraca Satelit Pariwisata Nasional (NESPARNAR) 2017*.
- BPS Indonesia. (2020). *Statistik Pengeluaran Wisatawan Mancanegara 2019*. BPS Indonesia.
- BPS Indonesia. (2021). *Statistik Indonesia 2020*. In *Statistik Indonesia 2020*. BPS Indonesia.
- BPS Provinsi Bali. (2022). *Provinsi Bali Dalam Angka 2022* (Issue December).
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2019). *Multivariate Data Analysis* (Eighth ed). Cengage Learning.
- Han, J., & Kamber, M. (2011). *Data Mining: Concepts and Techniques* (2nd ed.). Morgan Kaufmann Publishers.
- Hinkle, D. E., Wiersma, W., & Jurs, S. G. (2003). *Applied Statistics for the Behavioral Sciences* (Fifth Ed.). Houghton Mifflin.
- Iswandhani, N., & Muhajir, M. (2018). K-means cluster analysis of tourist destination in special region of Yogyakarta using spatial approach and social network analysis (a case study: Post of @explorejogja instagram account in 2016). *Journal of Physics: Conference Series*, 974(1). <https://doi.org/10.1088/1742-6596/974/1/012033>

- Izenman, A. J. (2008). *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning* (G. Casella, S. Fienberg, & I. Olkin (eds.)). Springer Science+Business Media, LLC.
- Kassambara, A. (2017). Practical Guide To Cluster Analysis in R: Unsupervised Machine Learning. *Journal of Computational and Graphical Statistics*.
- Kassambara, A., & Mundt, F. (2020). *factoextra: Extract and Visualize the Results of Multivariate Data Analyses* (R package version 1.0.7).
- Kencana, E. N. (2020). *Sains Data dengan R: Klusterisasi Menggunakan K-Means Clustering* (Issue May). <https://doi.org/10.13140/RG.2.2.29495.34721>
- Kencana, E. N., & Darmayanti, T. (2022). *Pengantar Sains Data: Analisis Statistika pada Data Besar dengan R* (K. Dharmawan (ed.); Pertama). In Press.
- Koyongian, Y., Katuuk, D. A., Rotty, V. N. J., & Lengkong, J. S. J. (2021). Segmentasi Wisatawan Asing Sulawesi Utara: Analisis pada Faktor Demografis dan Lifestyle Typology. *Jurnal Bahana Manajemen Pendidikan*, 10(1), 71–78. <https://doi.org/10.24036/jbmp.v10i1.112130>
- Li, Y., & Wu, H. (2012). A Clustering Method Based on K-Means Algorithm. *Physics Procedia*, 25, 1104–1109. <https://doi.org/10.1016/j.phpro.2012.03.206>
- Meriani, I. K., Sukarsa, K. G., Jayanegara, K., & Kencana, E. N. (2022). Faktor Kepatuhan Masyarakat Kota Denpasar pada Protokol Kesehatan COVID-19. *E-Jurnal Matematika*, 11(November), 268–277. <https://doi.org/10.24843/MTK.2022.v11.i04.p391>
- Reddy, C., & Vinzamuri, B. (2014). A Survey of Partitional and Hierarchical Clustering Algorithms. In C. Aggarwal & C. K. Reddy (Eds.), *Data Clustering: Algorithms and Applications*. CRC Press.
- RI. (2009). *Undang-undang Nomor 10 Tahun*.
- Rousseeuw, P. J. (1986). Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65.
- Savitri, N., Pranata, R., Clara, A. N. M., & Rahajeng, O. S. (2021). Pengelompokan Kunjungan Wisata Kabupaten Kulon Progo Tahun 2019 Menggunakan K-Means Clustering. *Jurnal Sistem Informasi, Teknologi Informasi Dan Komputer*, 12(1), 38–45. <https://jurnal.umj.ac.id/index.php/just-it/index>
- Susilo, A., Rumende, C. M., Pitoyo, C. W., Santoso, W. D., Yulianti, M., Herikurniawan, H., Sinto, R., Singh, G., Nainggolan, L., Nelwan, E. J., Chen, L. K., Widhani, A., Wijaya, E., Wicaksana, B., Maksum, M., Annisa, F., Jasirwan, C., & Yuniastuti, E. (2020). Coronavirus Disease 2019: Tinjauan Literatur Terkini. *Jurnal Penyakit Dalam Indonesia*, 7(1), 45–67. <https://doi.org/10.25104/transla.v22i2.1682>
- Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of data clusters via the gap statistic. In *Journal of the Royal Statistical Society: Series B* (Vol. 63, Issue Part 2, pp. 411–423).