

Perbandingan Algoritma K-Means dan DBSCAN Menggunakan Data Kunjungan Wisatawan Asing ke Indonesia di Masa Pandemi COVID-19

Muhamad Irfan Fadillah

Universitas Buana Perjuangan Karawang
Karawang, Indonesiaifi18.muhamadirfan@mhs.ubpkarawang.ac.id

Ayu Ratna Juwita

Universitas Buana Perjuangan Karawang
Karawang, Indonesiaayurj@ubpkarawang.ac.id

Cici Emilia Sukmawati

Universitas Buana Perjuangan Karawang
Karawang, Indonesiacici.emilia@ubpkarawang.ac.id

Abstract— Pandemi *Covid-19* berskala global yang sudah terjadi selama bertahun-tahun berpengaruh secara signifikan pada berbagai sektor. Pariwisata menjadi yang paling terdampak dari pandemi *Covid - 19*, dikarenakan cara untuk menahan peningkatan jumlah orang yang terinfeksi dengan membatasi pergerakan manusia maka wisatawan asing tidak bisa leluasa keluar negeri. Data yang ada di Badan Pusat Statistik (BPS) tentang kunjungan wisatawan mancanegara ke Indonesia di masa pandemi ini hanya berupa data mentah dan belum diolah dengan bahasa pemrograman ataupun berbagai kajian ilmiah. Subyek penelitian ini merupakan permasalahan data pariwisata mancanegara tersebut. Berdasarkan yang telah dirangkum pada penelitian ini, penerapan algoritma K-Means dan *DBSCAN* membutuhkan berbagai tahapan dalam pengolahan data pariwisata mancanegara yaitu pengumpulan data, seleksi data dan implementasi algoritma K-Means dan juga algoritma *DBSCAN*. Setelah didapatkan hasil dilakukan evaluasi data dengan *Silhouette Coefficient*. Hasil Perhitungan evaluasi dari Algoritma *DBSCAN* menghasilkan nilai akurasi sebesar 0,91962 lebih rendah ketimbang Algoritma K-Means yang memiliki nilai *silhouette* 0,96234, bisa disimpulkan bahwa hasil dari penelitian ini Algoritma K-Means lebih baik ketimbang Algoritma *DBSCAN* dalam pengelompokan dataset kunjungan wisatawan asing yang digunakan dalam penelitian ini.

Kata kunci — *pandemi, clustering, covid-19. wisatawan asing.*

I. PENDAHULUAN

Musibah pandemi *Covid-19* yang menimpa dunia selama bertahun-tahun sudah sangat berdampak secara signifikan untuk berbagai sektor. Pariwisata menjadi salah satu hal yang paling terdampak dari pandemi *Covid-19*. Mengingat salah satu cara untuk menahan peningkatan jumlah orang yang terinfeksi adalah dengan membatasi pergerakan manusia. Secara global, sektor pariwisata diperkirakan merugi Rp4.700 triliun pada Juni 2020. Jumlah pengangguran yang diciptakan oleh sektor industri ini di seluruh dunia diperkirakan 108 juta. Selain itu, kedatangan wisatawan pada paruh pertama tahun 2020 diperkirakan akan turun drastis menjadi 65% di seluruh dunia. Di Indonesia, pandemi *Covid-19* juga memukul sektor pariwisata. Jumlah wisatawan asing turun 68,17 persen dari Januari hingga Agustus 2020 dibandingkan tahun 2019 [1]. Banyak pekerja pariwisata harus diberhentikan dan banyak perusahaan di sektor jasa pariwisata bangkrut. Secara umum, teknologi informasi saat ini telah memberikan kontribusi yang signifikan dalam membantu menghadapi pandemi. Dengan menyediakan data yang akurat dan analisis yang lebih baik dapat mengatasi pandemi *Covid-19* dan mempersiapkan masyarakat untuk menyambut era *New Normal* melalui Adaptasi Kebiasaan Baru (AKB). Salah satunya adalah peran otomatisasi dalam pengendalian penyebaran *Covid-19*. Data dan hasil analisisnya bisa membantu memprediksi banyak hal terkait pandemi. Misalnya saja bisa melihat karakteristik kelompok penyebarannya. Data yang diproses akan membantu memvisualisasikan penyebaran pandemi. Data mengenai jumlah dan penyebaran *Covid-19*, berdasarkan kategori kelompok atau wilayah, memudahkan pelacakan kontak dan membantu memprediksi kemungkinan seseorang terkena *Covid-19*.

Penelitian yang dilakukan oleh [2], menunjukkan bahwa 5 tempat wisata teratas di *cluster* terendah adalah catatan pemerintah provinsi DKI Jakarta. Perbaikan infrastruktur objek wisata utama dapat meningkatkan jumlah kunjungan wisatawan, yang akan berdampak pada pengenalan objek wisata dan peningkatan transaksi devisa dalam negeri. Penelitian ini juga dilakukan oleh [3]. Berdasarkan hasil penelitian ini, dengan menerapkan *data mining* pada data *TripAdvisor*, dimungkinkan untuk memberikan informasi wisata dan menghasilkan pengetahuan baru tentang destinasi wisata yang berpengaruh berdasarkan hasil *cluster* yang diperoleh dengan implementasi di dapat dilakukan dengan lebih efektif. Hal ini juga menjadi tolak ukur dalam mengelola pemasaran dan promosi destinasi wisata terbaik di kawasan Asia Timur. Menurut survei, tempat piknik, tempat ibadah, pantai, resort, dan teater menduduki peringkat sebagai tujuan wisata terbaik di Asia Timur. Data diolah dengan *Microsoft Excel* untuk ditentukan nilai *centroid* nya dalam 3 *cluster* yaitu *cluster* tinggi (C1), *cluster* sedang (C2), *cluster* rendah (C3). Sehingga diperoleh penilaian berdasarkan pengelompokan jumlah wisatawan Asing ke Indonesia berdasarkan provinsi dengan 1 provinsi dalam *cluster* C1, 2 provinsi dalam *cluster* C2, dan 31 provinsi di *cluster* C3. Hasil yang didapatkan dari penelitian menunjukan hampir 90% provinsi di Indonesia masih memiliki potensi wisata yang rendah [4]. Hasil *clustering* pada penelitian yang dilakukan oleh [5] memberikan kesimpulan bahwa dengan menggunakan algoritma *DBSCAN*, maka diperoleh 3 *cluster* wilayah beresiko terjadi gempa bumi. Tiap-tiap *cluster* terbentuk dari jumlah kejadian gempa yang berbeda, yaitu 14 kejadian gempa pada *cluster* 1, 1.006 kejadian gempa pada *cluster* 2, dan 8 kejadian gempa pada *cluster* 3. Kombinasi nilai ϵ dan *minPts* yaitu 0,28 dan 3 menghasilkan nilai *Silhouette Coefficient* sebesar 0,8109115 dan Indeks Gamma sebesar 0,98104 yang menggambarkan bahwa *DBSCAN* mampu mengelompokan

wilayah beresiko terjadi gempa bumi dengan cukup baik. Telah dilakukan penelitian juga oleh [6] dapat diambil kesimpulan bahwa penerapan algoritma k-means membagi dataset menjadi tiga kelompok yaitu sangat rendah, rendah dan cukup sesuai dengan kemiripan tingkat pendapatan/kapita/bulan. Hasil pengujian mendapatkan nilai *davies bouldin index* sebesar 0,288 yang memiliki arti kesamaan antar anggota *cluster* yang cukup baik. Penelitian lain juga dilakukan oleh [7] menyimpulkan bahwa banyaknya desa/kelurahan menurut upaya antisipasi/mitigasi bencana alam menurut provinsi dapat diselesaikan dengan teknik *data mining* menggunakan aturan *clustering* untuk mengelompokkan desa/kelurahan menurut upaya antisipasi/mitigasi bencana alam di Indonesia. Data tersebut diolah menggunakan *Microsoft Excel* untuk ditentukan nilai *centroid* dalam 3 *cluster* yaitu *cluster* tingkat tinggi, *cluster* tingkat sedang dan *cluster* tingkat rendah. Sehingga diperoleh penilaian berdasarkan indeks indeks desa/kelurahan menurut upaya antisipasi/mitigasi bencana alam dengan 3 provinsi tingkat tinggi yaitu Jawa Barat, Jawa Tengah, Jawa Timur, 9 provinsi tingkat sedang, dan 22 provinsi lainnya termasuk tingkat rendah.

Data yang ada di Badan Pusat Statistik (BPS) tentang kunjungan wisatawan mancanegara ke Indonesia di masa pandemi ini hanya berupa data mentah dan belum diolah dengan bahasa pemrograman ataupun berbagai kajian ilmiah. Subyek penelitian ini merupakan permasalahan data pariwisata mancanegara tersebut. Penerapan algoritma K-Means dan *DBSCAN* membutuhkan berbagai tahapan dalam pengolahan data pariwisata mancanegara yaitu pengumpulan data, seleksi data dan implementasi algoritma K-Means dan juga algoritma *DBSCAN* dalam mengelompokkan dataset kunjungan wisatawan asing yang digunakan dalam penelitian ini. Mengelompokkan data wisatawan asing yang berkunjung ke Indonesia menggunakan Algoritma K-Means dan *DBSCAN* dengan menggunakan 2 cara perhitungan yaitu menggunakan *Microsoft Excel* atau perhitungan manual dan menggunakan *Python* untuk perhitungan otomatis kemudian membandingkan hasil akurasi.

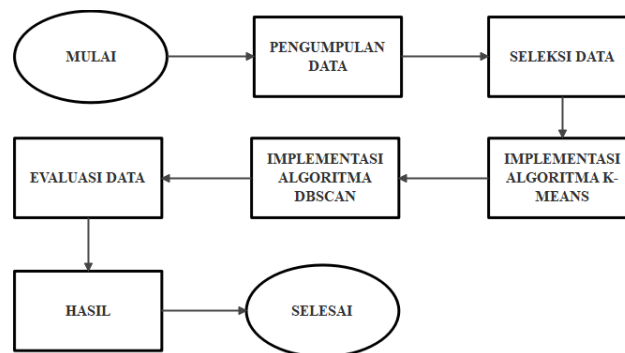
II. METODE PENELITIAN

A. Bahan Penelitian

Bahan atau data yang digunakan dalam penelitian ini adalah *dataset* kunjungan wisatawan asing ke Indonesia yang berasal dari Badan Pusat Statistik (BPS), meliputi data dari bulan Maret 2020 sampai dengan Desember 2021 dimana sedang terjadinya peristiwa besar pandemi *Covid-19*, *dataset* untuk penelitian ini diambil pada jam 14:12 WIB 10 Mei Tahun 2022, kemudian data tersebut dapat dikelompokkan dengan menerapkan Algoritma K-Means dan *DBSCAN*.

B. Prosedur Penelitian

Penelitian ini dilakukan dengan mengambil *dataset* kunjungan wisatawan asing dari Badan Pusat Statistik (BPS), kemudian data akan dikelompokkan dengan menggunakan Algoritma K-Means dan *DBSCAN* dengan 2 cara yaitu perhitungan menggunakan *microsoft excel* dan bahasa pemrograman *python*.



Gambar 1 Prosedur Penelitian

1. Pengumpulan data

Data yang digunakan dalam penelitian ini adalah *dataset* kunjungan wisatawan asing ke Indonesia yang berasal dari Badan Pusat Statistik (BPS), meliputi data dari bulan Maret 2020 sampai dengan Desember 2021 dimana sedang terjadinya peristiwa besar pandemi *Covid-19*, *dataset* untuk penelitian ini diambil pada jam 14:12 WIB 10 Mei Tahun 2022.

Tabel 1 Sample Data Kunjungan Wisatawan Asing Yang Berkunjung Ke Indonesia 2021

Kebangsaan	Januari	Februari	Maret	-	-	-	-	Desember
Brunei Darussalam	4	9	6	-	-	-	-	11
Malaysia	44.001	34.425	37.095	-	-	-	-	48.728
Philippines	779	679	1.014	-	-	-	-	855
Singapore	2.107	1.864	2.396	-	-	-	-	1.195
Thailand	258	301	327	-	-	-	-	435
Vietnam	141	167	155	-	-	-	-	277
Kamboja	2	5	13	-	-	-	-	24
Myanmar	368	293	259	-	-	-	-	316
Indonesia	935	664	1.214	-	-	-	-	1.487

2. Seleksi data

Pada tahap ini dilakukan seleksi secara manual menggunakan *microsoft excel* dan otomatis menggunakan bahasa pemrograman *python* dengan menghapus data yang tidak digunakan atau tidak bisa dihitung, kemudian menghapus atribut dan variable yang tidak perlu. Berikut sample data sebelum dan sesudah dibersihkan:

Tabel 2 Data Sebelum Seleksi Data

Kebangsaan	Jan	Feb	Mar	Apr	Mei	Juni	Juli	Ags	Sept	Okt	Nov	Des	Tahun
Brunei Darussalam	1224	922	444	8	17	17	4	8	13	26	9	9	2701
Malaysia	20752	16404	11445	6149	6641	6203	5805	5786	5337	4533	4329	4624	98011
Philippines	3	3	3	6	4	3	2	4	0	2	2	6	8
Singapore	18443	14188	6830	1511	1955	1670	806	676	713	1003	1275	1343	50413
Thailand	14056	84681	40832	2074	1334	1129	1166	1328	1397	1616	1931	2435	28049
Vietnam	9	7737	7492	2794	324	510	363	315	290	318	404	376	2
Laos	9408	6714	1618	270	293	248	180	190	162	175	169	181	21303
Laos	312	396	38	0	0	0	2	2	0	1	0	2	19608
Laos	312	396	38	0	0	0	2	2	0	1	0	2	753

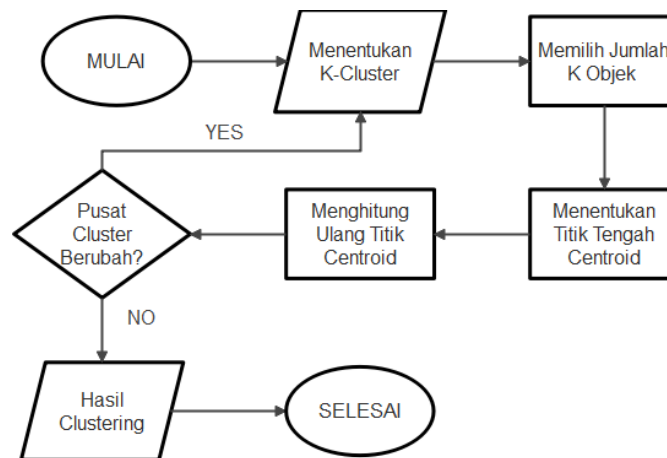
Tabel 3 Data Setelah Seleksi Data

Negara	Jan	Feb	Mar	Apr	Mei	Juni	Juli	Agst	Sept	Okt	Nov	Des
Brunei Darussalam	4	9	6	46	3	13	6	7	11	11	17	11
Malaysia	44.001	34.425	37.095	35.871	45.322	34.345	41.913	37.530	37.946	44.113	39.434	48.728
Philippines	779	679	1.014	995	972	670	483	438	511	831	1.148	855
Singapore	2.107	1.864	2.396	2.541	2.293	1.049	980	946	729	1.142	1.462	1.195
Thailand	258	301	327	418	2.293	299	225	225	265	389	427	435
Australia	154	220	204	300	284	302	181	71	190	261	456	573

New Zealand	9	34	21	63	56	62	24	7	21	59	53	73
Papua New Guinea	3.212	2.364	2.328	2.313	1.913	1.682	2.096	2.318	2.107	3.168	3.322	4.880
South Africa	17	35	28	55	109	104	31	8	27	72	70	16

3. Implementasi Algoritma K-Means

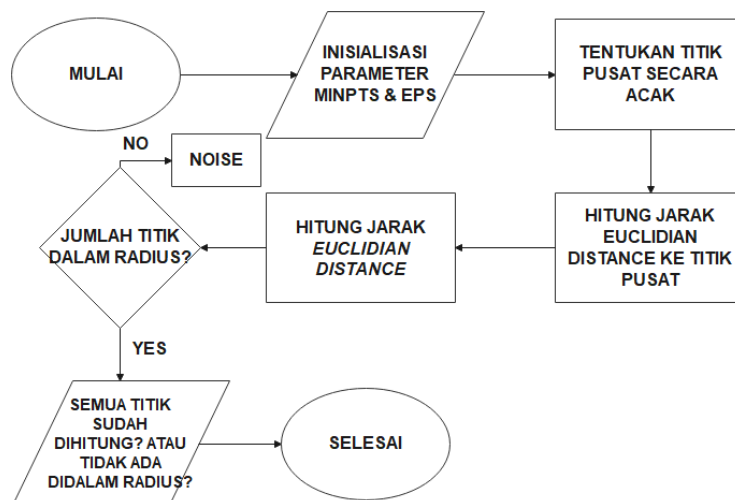
Pada tahap ini dilakukan menggunakan 2 cara yaitu menggunakan *Microsoft Excel* dan perhitungan *python*, tahapannya digambarkan sebagai berikut:



Gambar 2 Tahapan Algoritma K-Means

4. Implementasi algoritma DBSCAN

Pada tahap ini *dataset* akan di hitung menggunakan *Microsoft Excel* dan bahasa pemrograman *python*, dengan tahapannya digambarkan sebagai berikut:



Gambar 3 Alur Algoritma DBSCAN

5. Evaluasi

Pada tahap ini akan dihitung tingkat akurasi menggunakan *Silhouette Coefficient* dengan nilai dari -1 hingga 1, jika hasil lebih mendekati nilai 1 maka metode pengelompokan lebih bagus dan jika hasil mendekati nilai -1 maka metode pengelompokan lebih buruk.

6. Hasil

Setelah melakukan pengumpulan data, seleksi data, dan evaluasi data maka akan diketahui hasil dari pengelompokan data kunjungan wisatawan tersebut dengan menggunakan Algoritma K-Means dan *DBSCAN*. Akan terlihat nilai akurasi dari hasil evaluasi data dan bisa dibandingkan mana nilai akurasi yang lebih baik antara Algoritma K-Means dengan *DBSCAN*

III. HASIL DAN PEMBAHASAN

A. Penerapan Algoritma K-Means

Penerapan Algoritma K-Means dengan 2 cara yaitu perhitungan menggunakan excel dan perhitungan menggunakan bahasa pemrograman *python*. Kemudian hasilnya akan dievaluasi menggunakan *silhouette coefficient* untuk mendapatkan nilai akurasi. Berikut ini *sample* tahapan penerapan Algoritma K-Means:

1. Setelah melakukan perhitungan dengan menggunakan rumus euclidean distance, berikut hasil perhitungan manual k-means:

Tabel 4 Hasil Perhitungan Manual K-Means

Negara	C1	C2	C3	Jarak Terdekat	Cluster
Brunei Darussalam	710447,52	-	32577313304,25	710447,5234	1
Malaysia	13861022078,81	-	2411449051,25	2411449051	3
Philippines	2060572,52	-	31668373842,25	2060572,523	1
Singapore	18158477,52	-	30876062662,25	18158477,52	1
Thailand	3820288,13	-	31975079647,25	3820288,129	1
-	-	-	-	-	-
Maroko	700579,89	-	32572230804,25	700579,8918	1
Mozambik	759235,71	-	32586440730,25	759235,7076	1
Nigeria	601163,63	-	32540798185,25	601163,6286	1
Sudan	731490,29	-	32580272501,25	731490,2865	1
Tanzania	736158,29	-	32581065690,25	736158,2865	1

2. Pertama *import script library* k-means untuk menampilkan data yang akan diolah menggunakan bahasa *python*. *Import library* merupakan kumpulan *script* yang akan digunakan untuk mengolah data kunjungan wisatawan sesuai dengan yang dibutuhkan, kemudian data yang ditampilkan adalah data kunjungan negara asing yang sudah di *import* filenya untuk diolah menggunakan algoritma k-means.
3. Setelah itu menampilkan tampilan grafik pada data kunjungan wisatawan asing ke Indonesia, yang menunjukkan bahwa di titik teratas merupakan data yang termasuk pada *cluster 3* karena datanya memiliki kunjungan tertinggi, titik ditengah merupakan *cluster 2*, dan titik paling bawah merupakan *cluster 1* berdasarkan dari hasil perhitungan data 78 negara asing yang berkunjung ke Indonesia.
4. Kemudian melakukan perhitungan algoritma k-means menggunakan *python* untuk menampilkan perhitungan nilai *centroid* atau titik pusat dan juga nilai dari hasil pengelompokan. Hasilnya *cluster 1* terdapat 76 data, *cluster 2* terdapat 1 data dan *cluster 3* terdapat 1 data .

```
#Menampilkan data Cluster
data['cluster'] = kmeans.labels_
data
print('jumlah anggota setiap cluster')
print('cluster 0 = ', data[data['cluster'] == 0]['cluster'].count())
print('cluster 1 = ', data[data['cluster'] == 1]['cluster'].count())
print('cluster 2 = ', data[data['cluster'] == 2]['cluster'].count())

jumlah anggota setiap cluster
cluster 0 = 76
cluster 1 = 1
cluster 2 = 1
```

Gambar 4 Source Code K-Means Dan Hasil Cluster

B. Algoritma *DBSCAN*

Penerapan algoritma *dbscan* dengan 2 cara yaitu perhitungan menggunakan *Microsoft Excel* dan perhitungan menggunakan bahasa pemrograman *python*. Kemudian hasilnya akan dievaluasi menggunakan *silhouette coefficient* untuk mendapatkan nilai akurasi. Berikut ini *sample* tahapan penerapan Algoritma *dbscan*:

1. Setelah melakukan perhitungan dengan menggunakan rumus *euclidean distance*, berikut hasil perhitungan manual *dbscan*.

Tabel 5 Hasil Perhitungan Manual *DBSCAN*

Negara	Hasil Iterasi Terakhir	Cluster
Brunei Darussalam	1.225	C1
Malaysia	13.759.176.270	C3
Philippines	4.122.098	C1
Singapore	13.845.904	C1
Thailand	6.375.273	C1
-	-	-
Maroko	2.313	C1
Mozambik	68	C1
Nigeria	18.140	C1
Sudan	651	C1
Tanzania	527	C1

2. Pertama-tama input *script* untuk *import library DBSCAN* dan menampilkan data yang akan diolah menggunakan bahasa *python*. *Import library* merupakan kumpulan *script* yang akan digunakan untuk mengolah data kunjungan wisatawan asing sesuai dengan yang dibutuhkan, kemudian data yang ditampilkan adalah data kunjungan negara asing yang sudah di *import file*-nya untuk diolah menggunakan algoritma *DBSCAN*.
3. Kemudian menampilkan kumpulan *dataset* kunjungan wisatawan asing ke Indonesia pada tahun 2021 yang sudah di *import* dan akan dihitung menggunakan *script* algoritma *DBSCAN*
4. Setelah itu menjalankan *script* algoritma *DBSCAN* untuk menentukan nilai *Epsilon* dan *Minpts* dengan menggunakan *test sample* minimal 50 data dan nilai *Epsilon* sebesar 5000
5. Selanjutnya menunjukan bahwa nilai *noise* yang dihasilkan oleh Algoritma *DBSCAN* pada dataset kunjungan wisatawan asing tahun 2021 berjumlah 4 *noise* artinya dari 78 data yang akan diolah hanya 4 data saja yang memiliki nilai *noise* atau data yang akan diolah tergolong cukup baik karena menghasilkan nilai *noise* yang sangat sedikit.
6. Lalu input *script* untuk melakukan *clustering DBSCAN* dengan bahasa pemrograman *python* dan menghasilkan 2 jenis *cluster* yaitu 4 data termasuk di data -1 dan 74 data lain nya termasuk di data 0, sesuai dengan nilai jumlah *noise* terdapat 4 data termasuk kedalam *outliner data*, atau bisa dibilang 4 data tersebut merupakan data yang berjauhan dengan data lain nya.

```
[22] print (dbscan.labels_)

[ 0 -1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 0 0 -1 0 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
  0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 0 0
  0 0 0 0 0]
```

Gambar 5 Hasil *Clustering DBSCAN*

C. Hasil Evaluasi Data

Pada Gambar 13 dan 14 merupakan hasil evaluasi algoritma k-means dan *DBSCAN* pada data penelitian ini, hasilnya k-means mendapatkan nilai akurasi 0,9623, sedangkan algoritma *DBSCAN* mendapatkan nilai akurasi 0,9196 yang berarti algoritma k-means lebih baik daripada algoritma *DBSCAN* dan keduanya sangat bagus akurasi untuk pengelompokan *dataset* pada penelitian ini karena sangat mendekati nilai 1.

```
#Evaluasi silhoutte coefficient k-means
silhouette_avg = silhouette_score(data.iloc[:, 1 : 10], kmeans.labels_)
print(silhouette_avg)
```

0.9623426304323084

Gambar 6 Hasil Akurasi K-Means

```
[24] from sklearn.metrics import silhouette_samples, silhouette_score
MIF = silhouette_score(dataset,dbscan.labels_)
print(MIF)
```

0.9196293828304325

Gambar 7 Hasil Akurasi DBSCAN

D. Hasil Pengelompokan Algoritma K-Means dan DBSCAN

Pada tabel 6 merupakan hasil penerapan Algoritma *K-Means* & *DBSCAN* menggunakan perhitungan manual *excel* dan bahasa pemrograman *python*, dapat diketahui hasil *clustering* pada *dataset* 2021 menggunakan *Microsoft Excel* menunjukkan ada 76 negara berada *cluster* 1, *cluster* ke 2 tidak ada, dan *cluster* 3 terdapat 2 negara. Sedangkan hasil perhitungan dengan *python* menunjukkan ada 76 negara berada *cluster* 1, untuk *cluster* 2 terdapat 1 negara, dan untuk *cluster* 3 terdapat 1 negara. Implementasi menggunakan Algoritma *DBSCAN* menggunakan *excel* menunjukkan ada 74 negara berada *cluster* 1, *cluster* ke 2 tidak ada, dan *cluster* 3 terdapat 4 negara yang berarti hanya ada 2 *cluster*.

Tabel 6 Hasil Pengelompokan Algoritma K-Means dan DBSCAN

NO	NEGARA	Manual	Python	Manual	Python
		Algoritma K-Means		Algoritma DBSCAN	
1	Brunei Darussalam	C1	C1	C1	C1
2	Malaysia	C3	C3	C3	C3
3	Philippines	C1	C1	C1	C1
4	Singapore	C1	C1	C1	C1
5	Thailand	C1	C1	C1	C1
-	-	-	-	-	-
74	Maroko	C1	C1	C1	C1
75	Mozambik	C1	C1	C1	C1
76	Nigeria	C1	C1	C1	C1
77	Sudan	C1	C1	C1	C1
78	Tanzania	C1	C1	C1	C1

IV. KESIMPULAN DAN SARAN

A. Kesimpulan

Hasil dan pembahasan pada penelitian yang telah dilakukan ini, dapat diambil beberapa kesimpulan sebagai berikut:

- 1) Berdasarkan yang sudah dilakukan dalam penelitian ini menerapkan Algoritma *K-Means Clustering* memerlukan berbagai tahap. Untuk memproses data kunjungan wisatawan asing yaitu melalui tahap pengumpulan data, seleksi data setelah itu baru mengimplementasikan Algoritma *K-Means* dengan menggunakan *Microsoft Excel* dan juga *python* dan melanjutkan nya implementasi menggunakan Algoritma *DBSCAN* setelah mendapatkan hasil dilakukan evaluasi data dengan *Silhouette Coefficient*.
- 2) Hasil penelitian yang sudah dilakukan, pada implementasi perhitungan menggunakan Algoritma *K-Means*. Hasil *clustering* pada *dataset* 2021 menggunakan *excel* menunjukkan ada 76 negara berada *cluster* 1, *cluster* ke 2 tidak ada, dan *cluster* 3 terdapat 2 negara. Hasil perhitungan dengan *python* menunjukkan ada 76 negara berada *cluster* 1, untuk *cluster* 2

terdapat 1 negara, dan untuk *cluster* 3 terdapat 1 negara. Implementasi menggunakan Algoritma *DBSCAN* menggunakan *excel* menunjukkan ada 74 negara berada *cluster* 1, *cluster* ke 2 tidak ada, dan *cluster* 3 terdapat 4 negara. dengan negara Malaysia yang menjadi titik pusat pertama. Perhitungan menggunakan *python* menghasilkan 2 *cluster* yaitu -1 dan 0 dengan menggunakan nilai *Epsilon* sebesar 5000 dan nilai *minpts* sebesar 50. Perhitungan evaluasi dari Algoritma *DBSCAN* menghasilkan nilai akurasi sebesar 0.91962 lebih rendah ketimbang Algoritma K-Means yang memiliki nilai *silhouette* 0.96234. Bisa disimpulkan bahwa hasil dari penelitian ini Algoritma K-Means lebih baik ketimbang Algoritma *DBSCAN* dalam pengelompokan *dataset* kunjungan wisatawan asing yang digunakan dalam penelitian ini.

B. Saran

Penelitian ini disarankan menggunakan data serta model penelitian lainnya yang diperinci sebagai berikut :

- 1) Menggunakan data yang tidak memiliki nilai nol lebih disarankan, sehingga perhitungan Algoritma K-Means dan *DBSCAN* akan lebih optimal.
- 2) Diharapkan kepada para peneliti yang lain mengkombinasikan atau mengkomparasi metode lain untuk menaikkan nilai akurasi dari hasil pengelompokan.

PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Muhamad Irfan Fadillah dengan judul Pengelompokan Data Kunjungan Wisatawan Asing Ke Indonesia Di Masa Pandemi *Covid-19* Menggunakan Algoritma K-Means Dan *DBSCAN* yang dibimbing oleh Ibu Ayu Ratna Juwita, M.Kom dan Ibu Cici Emilia Sukmawati, M.Kom.

DAFTAR PUSTAKA

- [1] Kementerian Kesehatan Republik Indonesia, "Pedoman Pencegahan dan Pengendalian Coronavirus Disease (COVID-19)," *Germas*, pp. 0–115, 2020, [Online]. Available: https://infeksiemerging.kemkes.go.id/download/REV-04_Pedoman_P2_COVID-19_27_Maret2020_TTD1.pdf [Diakses 11 Juni 2021].
- [2] L. Maulida, "Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means," *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, p. 167, 2018, doi: 10.14421/jiska.2018.23-06.
- [3] S. Seimahaira, "Implementasi Datamining Dalam Menentukan Destinasi Unggulan Berdasarkan Online Reviews Tripadvisor Menggunakan Algoritma K-Means," *Technol. J. Ilm.*, vol. 12, no. 1, p. 53, 2021, doi: 10.31602/tji.v12i1.4229.
- [4] R. W. Sari and D. Hartama, "Data Mining : Algoritma K-Means Pada Pengelompokkan Wisata Asing ke Indonesia Menurut Provinsi," *Semin. Nas. Sains Teknol. Inf.*, pp. 322–326, 2018.
- [5] R. R. A. Rahman and A. W. Wijayanto, "Pengelompokan Data Gempa Bumi Menggunakan Algoritma Dbscan," *J. Meteorol. dan Geofis.*, vol. 22, no. 1, p. 31, 2021, doi: 10.31172/jmg.v22i1.738.
- [6] B. P. Haryaji, "Penerapan Algoritma K-Means Untuk Memetakan Garis Kemiskinan Menurut Provinsi Di Indonesia," *Skripsi*, pp. 1–65, 2018, [Online]. Available: <https://ecampus.pelitabangsa.ac.id/pb/AmbilLampiran?ref=22949&jurusan=&jenis=Item&usingId=false&download=false&clazz=ais.database.model.file.LampiranLain>.
- [7] M. G. Sadewo, A. P. Windarto, and A. Wanto, "Penerapan Algoritma Clustering Dalam Mengelompokkan Banyaknya Desa/Kelurahan Menurut Upaya Antisipasi/ Mitigasi Bencana Alam Menurut Provinsi Dengan K-Means," *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, vol. 2, no. 1, pp. 311–319, 2018, doi: 10.30865/komik.v2i1.943.