

# Penerapan Algoritma K-Means Dan K-Medoid Untuk Mengelompokkan Tingkat Kunjungan Wisatawan Mancanegara Berdasarkan Rute Masuk Ke Indonesia

Rizky Alfian Nugraha  
Universitas Buana Perjuangan Karawang  
Karawang, Indonesia  
[if18.rizkynugraha@mhs.ubpkarawang.ac.id](mailto:if18.rizkynugraha@mhs.ubpkarawang.ac.id)

Yana Cahyana  
Universitas Buana Perjuangan Karawang  
Karawang, Indonesia  
[yana.cahyana@ubpkarawang.ac.id](mailto:yana.cahyana@ubpkarawang.ac.id)

Ayu Ratna Juwita  
Universitas Buana Perjuangan Karawang  
Karawang, Indonesia  
[ayurj@ubpkarawang.ac.id](mailto:ayurj@ubpkarawang.ac.id)

**Abstract**—Jumlah kunjungan wisatawan asing yang berkunjung ke suatu negara merupakan salah satu indikator penting untuk melihat perkembangan pariwisata. Ketertarikan wisatawan asing ke Indonesia dapat meningkatkan devisa negara dan meningkatkan perekonomian masyarakat di daerah wisata. Kurangnya informasi serta pengelompokan menyangkut ketertarikan dan tingkat kunjungan data wisatawan asing berdasarkan jalur masuk ke Indonesia mengakibatkan adanya kesulitan bagi pemerintah dalam melakukan kebijakan dan perencanaan strategi pemasaran, peningkatan infrastruktur serta fasilitas pariwisata untuk menarik minat para wisatawan asing, sehingga menjadi tidak tepat sasaran, menyebabkan wisata-wisata di Indonesia yang kurang dikenal jarang dikunjungi oleh wisatawan asing. Oleh karena itu, tujuan yang ingin dicapai dalam penelitian ini adalah mendapatkan pengelompokan tingkat kunjungan pada data wisatawan asing berdasarkan jalur masuk udara, darat, laut ke Indonesia dari tahun 2017 sampai 2021 dengan menerapkan algoritma k-means dan k-medoid. Algoritma ini dipilih karena mempunyai kemampuan mengelompokkan data dalam jumlah yang cukup besar dengan waktu yang relatif cepat. Proses perhitungan pada penelitian ini memakai 2 cara yaitu perhitungan manual menggunakan excel dan Python yang terbagi menjadi 3 kelompok yaitu C1 tingkat kunjungan tinggi, C2 tingkat kunjungan sedang dan C3 tingkat kunjungan rendah. Hasil pengelompokan algoritma k-means yaitu C1 sebanyak 24, C2 sebanyak 1, dan C3 sebanyak 4. Sedangkan algoritma k-medoid yaitu C1 sebanyak 5, C2 sebanyak 15, dan C3 sebanyak 9. Kemudian hasil evaluasi data menggunakan Silhouette Coefficient menunjukkan bahwa hasil akurasi algoritma k-means bernilai 0,8463 dan k-medoid bernilai 0,4691 yang berarti hasil akurasi algoritma k-means lebih baik daripada algoritma k-medoid karena lebih mendekati nilai 1.

**Kata kunci** —k-means, k-medoid, pengelompokan, wisatawan.

## I. PENDAHULUAN

Indonesia adalah negara arkeologi dengan sumber daya alam pariwisata yang menarik karena keindahan alam dan keanekaragaman budaya, sehingga wisatawan asing sering berkunjung melalui jalur transportasi udara, darat dan laut. Pariwisata dapat dengan mudah didefinisikan sebagai perjalanan individu atau beberapa kelompok manusia ke tempat tujuan, dan direncanakan untuk bersantai, menghibur serta memuaskan keinginan wisatawan asing dalam jangka waktu tertentu [1]. Pariwisata juga merupakan sektor ekonomi penting yang diharapkan dapat meningkatkan pendapatan Indonesia seiring dengan meningkatnya wisatawan asing ke Indonesia. Sektor pariwisata juga memegang peranan penting dalam pembangunan Indonesia sebagai sumber devisa negara. Hal ini sejalan dengan Peraturan Presiden Nomor 38 Tahun 2005 yang menyatakan bahwa semua sektor harus mendukung pembangunan pariwisata Indonesia. Selain itu, kata pemerintah, pariwisata harus menjadi pusat pembangunan Indonesia [2].

Mengetahui dari Badan Pusat Statistik mengenai data kunjungan wisatawan asing berdasarkan jalur masuk udara, darat, laut ke Indonesia dari tahun 2017 sampai 2021 sempat mengalami ketidakstabilan dan pasang surut pengunjung setiap tahun. Pada tahun 2017 jumlah kunjungan wisatawan asing ke Indonesia mulai mengalami peningkatan dan penurunan hingga tahun 2019, namun pada tahun 2020 sampai 2021 terlihat adanya penurunan jumlah pengunjung yang sangat drastis karena efek pandemi covid-19. Kurangnya informasi serta pengelompokan data menyangkut ketertarikan dan tingkat kunjungan pada data wisatawan asing ke Indonesia berdasarkan jalur masuk juga menjadi hal yang mengakibatkan adanya kesulitan bagi pemerintah dalam melakukan kebijakan dan perencanaan strategi pemasaran, peningkatan infrastruktur serta fasilitas pariwisata untuk menarik minat wisatawan asing, sehingga menjadi tidak tepat sasaran, menyebabkan beberapa wisata di Indonesia yang kurang dikenal jarang dikunjungi oleh wisatawan asing [3].

Dalam metode ini, pengelompokan data dilakukan menggunakan 2 cara yaitu perhitungan manual excel dan bahasa pemrograman *Python* untuk mendapatkan hasil pengelompokan yang kemudian akan dievaluasi untuk memperoleh nilai akurasi algoritma k-means dan k-medoid [4]. Hasil pengelompokan algoritma k-means yaitu C1 sebanyak 24, C2 sebanyak 1, dan C3 sebanyak 4. Sedangkan algoritma k-medoid yaitu C1 sebanyak 5, C2 sebanyak 15, dan C3 sebanyak 9. Kemudian hasil evaluasi data menggunakan Silhouette Coefficient menunjukkan bahwa hasil akurasi algoritma k-means bernilai 0,8463 dan k-medoid bernilai 0,4691 yang berarti hasil akurasi algoritma k-means lebih baik daripada algoritma k-medoid karena lebih mendekati nilai 1 [5]. Kelebihannya dalam implementasi, pengelompokan dan teknik data mining sangat mudah untuk proses pengelompokan data yang relatif cepat, efektif dan efisien. Sehingga dengan adanya pengelompokan data seperti ini diharapkan bisa menjadi sumber informasi bagi pemerintahan agar mendapatkan rancangan strategis dalam pengembangan destinasi pariwisata untuk menarik minat wisatawan asing dan menjadikan pariwisata sebagai sektor unggulan.

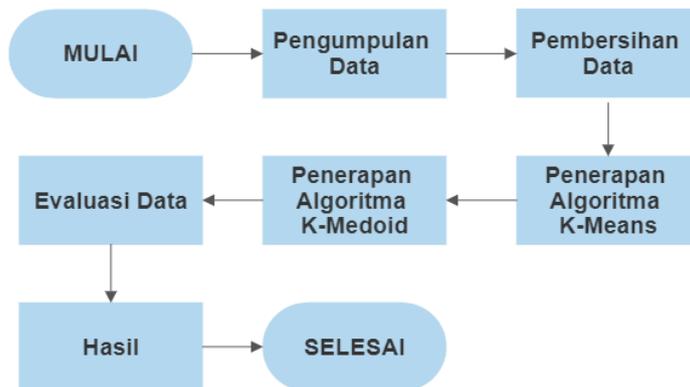
## II. METODE PENELITIAN

### A. Bahan Penelitian

Dalam penelitian ini hal yang penting adalah mempersiapkan bahan penelitian yang didapatkan dari situs resmi pemerintah yaitu badan pusat statistik, data diambil pada hari sabtu 05 februari 2022 sebanyak 1.572 data, kemudian data tersebut dapat dikelompokkan dengan menerapkan algoritma k-means dan k-medoid.

### B. Prosedur Penelitian

Pada Penelitian ini yang digunakan adalah dataset kunjungan wisatawan asing berdasarkan jalur masuk perbulannya, dataset diambil dari situs resmi pemerintah yaitu badan pusat statistik, data akan dikelompokkan dengan menerapkan algoritma k-means dan k-medoid, perhitungannya menggunakan 2 cara yaitu perhitungan manual menggunakan excel dan bahasa pemrograman *python*.



Gambar 1 Prosedur Penelitian

1. Pada tahap pengumpulan data, Dalam proses ini maka dibutuhkan data yang sumbernya valid dan relevan. Maka untuk memperoleh data tersebut, telah dilakukan pencarian data melalui situs resmi pemerintah yaitu Badan Pusat Statistik (BPS) di [www.bps.go.id](http://www.bps.go.id). Data yang diperoleh merupakan data kunjungan wisatawan asing berdasarkan jalur masuk udara, darat, laut ke Indonesia perbulannya pada tahun 2017 sampai 2021.

Table 1 Data Kunjungan Wisatawan Asing Berdasarkan Jalur Masuk

Pintu Masuk	Januari	Februari	Maret	April	Mei	. .	November	Desember
Ngurah Rai	451708	436266	441707	476104	483928	. .	492904	544726
Soekarno-Hatta	174963	196183	214161	196977	156654	. .	183759	186723
Juanda	13792	17389	20497	18431	14529	. .	20780	20546
Kualanamu	16253	21815	21129	21613	15573	. .	20798	22431
Husein Sastranegara	12524	15127	16342	14738	8029	. .	14616	14951
Adi Sucipto	8055	9604	10550	9922	6664	. .	9218	9795
Bandara Int. Lombok	2126	3590	3568	5143	2884	. .	4635	5499
Sam Ratulangi	10905	11305	10761	11574	9755	. .	9261	10743
Minangkabau	4528	5144	6196	5434	3576	. .	5013	5180
Sultan Syarif Kasim II	2450	2989	3071	2745	2364	. .	2917	4042
Sultan Iskandar Muda	1676	1739	2134	2225	1360	. .	2417	4012
Ahmad Yani	2165	2033	2296	1635	996	. .	2296	2627
Supadio	1598	2639	2316	1886	1527	. .	1811	1527
Hasanuddin	951	1379	1825	1312	1105	. .	1595	1821
Sultan Badaruddin II	1038	1573	1419	1409	1396	. .	1218	762
Pintu Udara Lainnya	1972	2742	849	890	889	. .	4006	3593
Batam	134415	159248	172461	154810	145447	. .	167288	190232
Tanjung Uban	40601	49394	61156	48560	58673	. .	52374	59340

Tanjung Pinang	9268	13141	15813	15841	14816	· ·	12378	16426
Tanjung Balai Karimun	8572	11411	10859	10039	6235	· ·	9209	10916
Tanjung Bena	3862	1190	7862	53	37	· ·	5021	7677
Tanjung Mas	1646	2328	4999	1015	50	· ·	2202	5135
Pintu Laut Lainnya	106115	104039	105548	98566	109903	· ·	109792	98769
Jayapura	4418	3889	3871	4702	4743	· ·	5006	6155
Atambua	7835	5973	6741	9618	7457	· ·	9824	12561
Entikong	1563	2158	2366	1333	1290	· ·	1691	3177
Aruk	804	1989	2171	836	939	· ·	1009	1708
Nanga Badau	1187	1128	1185	937	1705	· ·	954	1585
Pintu Darat Lainnya	174745	156591	158058	155883	187062	· ·	126789	124408

2. Pada tahap pembersihan data, dilakukan secara manual dengan memisahkan antara data yang digunakan dan data yang tidak digunakan. Sebelum dibersihkan, terdapat beberapa variabel dan atribut yang tidak diperlukan seperti data kosong, tahunan, jumlah total, dan keterangan yang tidak diperlukan. Kemudian data yang tidak digunakan akan dihapus atau dibersihkan agar data menjadi efektif saat melakukan proses perhitungan. Setelah dibersihkan, hanya terdapat atribut jalur masuk, kunjungan perbulan dan variabel data yang diperlukan. Berikut sample data sebelum dan sesudah dibersihkan:

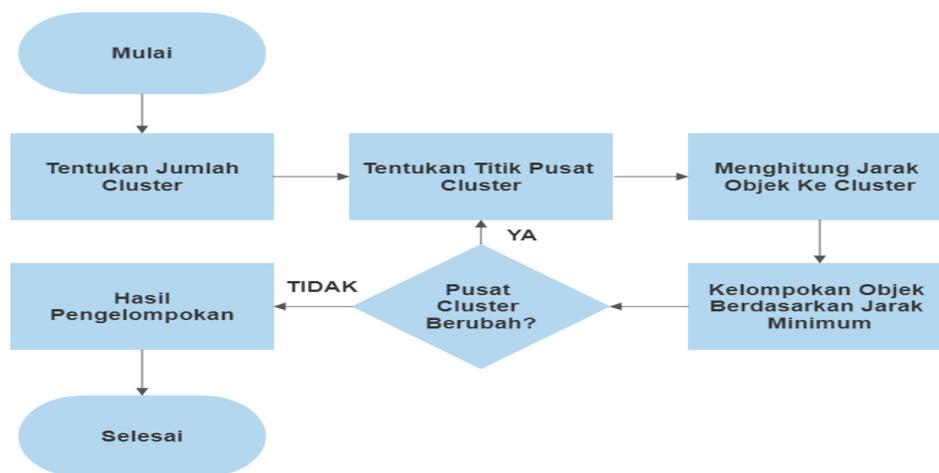
Table 2 Data Sebelum Dibersihkan

Pintu Masuk	Kunjungan Wisatawan Asing perbulan ke Indonesia (2019)				
	Januari	.	.	Desember	Tahunan
<b>A. Pintu Udara</b>	706704	.	.	838978	-
Ngurah Rai	451708	.	.	544726	-
Soekarno-Hatta	174963	.	.	186723	-
·	·	·	·	·	-
·	·	·	·	·	-
Sultan Badaruddin II	1038	.	.	762	-
Pintu Udara Lainnya	1972	.	.	3593	-
<b>B. Pintu Laut</b>	304479	.	.	388495	-
Batam	134415	.	.	190232	-
Tanjung Uban	40601	.	.	59340	-
·	·	·	·	·	-
·	·	·	·	·	-
Tanjung Mas	1646	.	.	5135	-
Pintu Laut Lainnya	106115	.	.	98769	-
<b>C. Pintu Darat</b>	190552	.	.	149594	-
Jayapura	4418	.	.	6155	-
Atambua	7835	.	.	12561	-
·	·	·	·	·	-
·	·	·	·	·	-
Nanga Badau	1187	.	.	1585	-
Pintu Darat Lainnya	174745	.	.	124408	-
<b>Jumlah (A+B+C)</b>	<b>1201735</b>	.	.	<b>1377067</b>	-

Table 3 Data Setelah Dibersihkan

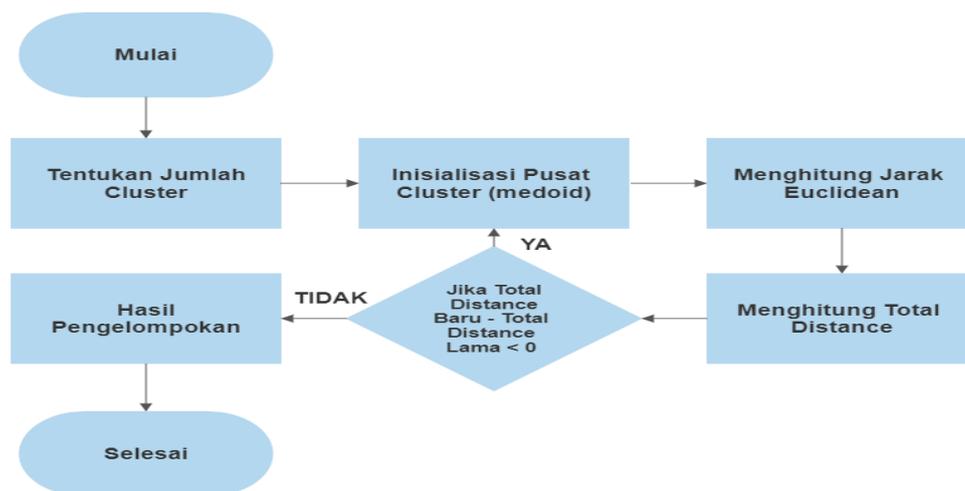
Pintu Masuk	Januari	Februari	.	.	November	Desember
Ngurah Rai	451708	436266	.	.	492904	544726
Soekarno-Hatta	174963	196183	.	.	183759	186723
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Sultan Badaruddin II	1038	1573	.	.	1218	762
Pintu Udara Lainnya	1972	2742	.	.	4006	3593
Batam	134415	159248	.	.	167288	190232
Tanjung Uban	40601	49394	.	.	52374	59340
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Tanjung Mas	1646	2328	.	.	2202	5135
Pintu Laut Lainnya	106115	104039	.	.	109792	98769
Jayapura	4418	3889	.	.	5006	6155
Atambua	7835	5973	.	.	9824	12561
.	.	.	.	.	.	.
.	.	.	.	.	.	.
Nanga Badau	1187	1128	.	.	954	1585
Pintu Darat Lainnya	174745	156591	.	.	126789	124408

3. Pada tahap penerapan algoritma k-means, dilakukan perhitungan menggunakan 2 cara yaitu menggunakan perhitungan manual excel dan perhitungan *python* dengan rumus euclidean distance, tahapannya sebagai berikut:



Gambar 2 Alur Algoritma K-Means

4. Pada tahap penerapan algoritma k-medoid, dataset akan diproses dengan menerapkan algoritma k-medoid menggunakan perhitungan manual dan bahasa pemrograman *python* menggunakan rumus *euclidean distance* dengan tahapan berikut:



5. Pada Tahap evaluasi data, dapat dihitung  $t$  dengan nilai dari -1 hingga 1, dimana jika hasil lebih mendekati nilai 1 maka metode pengelompokan lebih bagus dan jika hasil mendekati nilai -1 maka metode pengelompokan sangat buruk.
6. Kemudian Hasil, setelah melakukan beberapa proses tahapan seperti pengumpulan data, pembersihan data, penerapan algoritma k-means, penerapan algoritma k-medoid, evaluasi data dan hasil. Maka akan keluar hasil dari pengelompokan data kunjungan wisatawan asing berdasarkan jalur masuk udara, darat, laut ke Indonesia yang dibagi menjadi 3 pengelompokan yaitu C1 tingkat kunjungan tinggi, C2 tingkat kunjungan sedang dan C3 tingkat kunjungan rendah.

### III. HASIL DAN PEMBAHASAN

#### A. Algoritma K-Means

Pengujian dataset penelitian ini, setiap perhitungan dilakukan dengan menerapkan algoritma k-means menggunakan rumus *euclidean distance*. Hasilnya akan dievaluasi menggunakan *silhouette coefficient* untuk mendapatkan nilai akurasi dari pengelompokan algoritma k-means. Hasil dari perhitungan manual dan python dapat dilihat pada gambar dan tabel berikut:

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

Table 4 Hasil Perhitungan Manual K-Means

Hasil Perhitungan Manual K-Means					
Data Ke-i	C1	C2	C3	Jarak Terdekat	cluster
Data ke 1	1.791.888	0	1.345.614	0	C2
Data ke 2	684.400	1.116.120	243.875	243.875	C3
Data ke 3	50.308	1.742.310	400.513	50.308	C1
Data ke 4	49.392	1.743.021	400.707	49.392	C1
Data ke 5	24.927	1.768.402	425.978	24.927	C1
Data ke 6	11.709	1.780.804	438.667	11.709	C1
Data ke 7	6.411	1.796.390	454.520	6.411	C1
Data ke 8	16.621	1.775.910	433.696	16.621	C1
Data ke 9	4.250	1.795.874	453.595	4.250	C1
Data ke 10	11.769	1.803.541	461.342	11.769	C1
Data ke 11	13.514	1.805.148	463.103	13.514	C1
Data ke 12	14.807	1.806.537	464.377	14.807	C1
Data ke 13	15.313	1.807.005	464.780	15.313	C1
Data ke 14	16.559	1.808.279	466.127	16.559	C1
Data ke 15	17.261	1.808.881	466.680	17.261	C1
Data ke 16	14.820	1.806.289	464.215	14.820	C1
Data ke 17	543.170	1.255.045	103.495	103.495	C3

Data ke 18	162.960	1.631.458	288.224	162.960	C1
Data ke 19	28.434	1.764.778	422.240	28.434	C1
Data ke 20	11.472	1.781.219	438.721	11.472	C1
Data ke 21	15.745	1.805.385	462.998	15.745	C1
Data ke 22	17.261	1.808.631	466.208	17.261	C1
Data ke 23	339.739	1.457.304	115.189	115.189	C3
Data ke 24	5.985	1.797.151	455.030	5.985	C1
Data ke 25	9.703	1.783.334	441.389	9.703	C1
Data ke 26	14.973	1.806.715	464.531	14.973	C1
Data ke 27	16.967	1.808.695	466.443	16.967	C1
Data ke 28	17.654	1.809.271	467.094	17.654	C1
Data ke 29	531.458	1.274.356	101.613	101.613	C3
		24 data	1 data	4 data	

2. Selanjutnya perhitungan python pada gambar 4 merupakan tampilan untuk *import library* k-means dan menampilkan data menggunakan *python*.

```

+ ✂ 📄 ▶ ■ ⌂ ⏪ Code ▾ Pyth
[1]: #Library
import matplotlib.pyplot as plt # plotting
import numpy as np # dense matrices
import pandas as pd
from pandas import read_csv
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples, silhouette_score

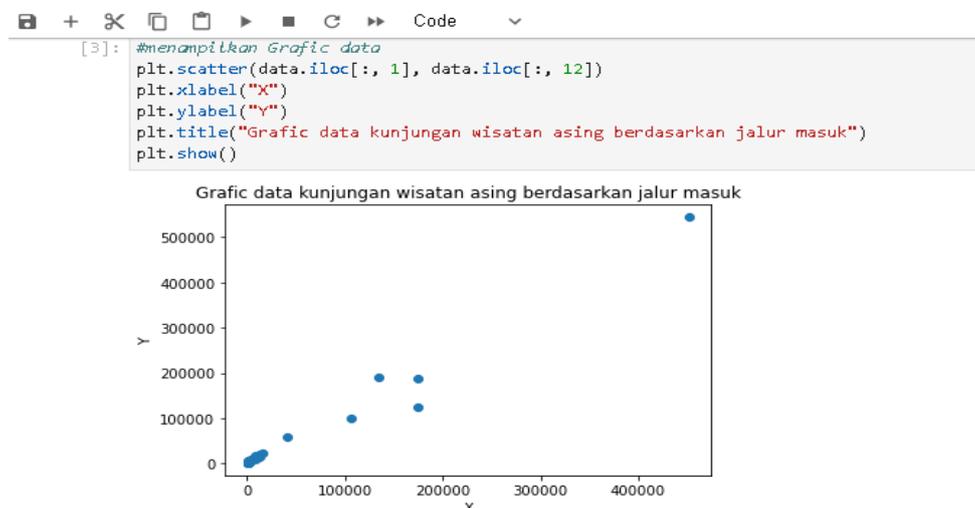
[2]: #Menampilkan dan menyisikkan data
data = read_csv ('rapid python rizky.csv')
data.head()

[2]:
```

	Pintu Masuk	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	September	Oktober	November	Desember
0	Ngurah Rai	451708	436266	441707	476104	483928	549483	604310	602457	589984	565966	492904	544726
1	Soekarno-Hatta	174963	196183	214161	196977	156654	190031	267143	251596	211775	189231	183759	186723
2	Juanda	13792	17389	20497	18431	14529	22485	24913	29180	20462	20895	20780	20546
3	Kualanamu	16253	21815	21129	21613	15573	18935	20929	24623	18913	21518	20798	22431
4	Husein Sastranegara	12524	15127	16342	14738	8029	8652	12581	13984	12869	13420	14616	14951

Gambar 4 Source Code Library K-Means Dan Tampilan Data

3. Pada gambar 5 merupakan tampilan grafik pada data kunjungan wisatawan asing berdasarkan jalur masuk.



Gambar 5 Source Code Dan Hasil Grafik Data K-Means

4. Pada gambar 6 merupakan perhitungan algoritma k-means menggunakan python untuk menampilkan nilai *centroid* yang di dapat dan hasil pengelompokan.

```
[4]: #Algoritma K-means
kmeans = KMeans(n_clusters=3)
kmeans.fit(data.iloc[:, 1 : 12])
label = kmeans.labels_
print (label)
pusat = kmeans.cluster_centers_
print(pusat)

[[1 2 0 0 0 0 0 0 0 0 0 0 0 0 2 0 0 0 0 0 2 0 0 0 0 2]
[[ 6657.875      7986.20833333  9165.66666667  7995.45833333
   6941.33333333  8312.25      8026.125      9458.70833333
   7738.45833333  7811.41666667  8343.70833333]
[451708.      436266.      441707.      476104.
 483928.      549483.      604310.      602457.
 589984.      565966.      492904.      ]
[147559.5      154015.25      162557.      151559.
 149766.5      171281.5      167809.      175215.5
 153253.      148273.5      146907.      ]]]

[5]: #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 = 24
cluster 1 = 1
cluster 2 = 4

[ ]:
```

Gambar 6 Source Code K-Means Dan Hasil Cluster

B. Algoritma K-Medoid

Pengujian dataset penelitian ini setiap perhitungan dilakukan dengan menerapkan algoritma k-medoid menggunakan rumus *euclidean distance*. Hasilnya akan dievaluasi menggunakan *silhouette coefficient* untuk mendapatkan nilai akurasi dari pengelompokan algoritma k-medoid. Hasil dari perhitungan manual dan python dapat dilihat pada gambar dan tabel berikut:

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

Table 5 Hasil Perhitungan Manual K-Medoid

Hasil Perhitungan Manual K-Medoid					
Data Ke-i	C1	C2	C3	Jarak Terdekat	cluster
Data ke 1	0	912.155	893.229	0	C1
Data ke 2	893.564	685.853	704.643	685.853	C2
Data ke 3	342.310	70.523	50.757	50.757	C3
Data ke 4	643.021	50.720	69.942	50.720	C2
Data ke 5	578.402	44.825	26.596	26.596	C3
Data ke 6	480.804	31.718	13.318	13.318	C3
Data ke 7	596.390	16.354	5.172	5.172	C3
Data ke 8	475.910	36.596	18.138	18.138	C3
Data ke 9	675.337	16.450	3.743	3.743	C3

Data ke 10	503.541	8.708	10.479	8.708	C2
Data ke 11	305.148	17.330	12.166	17.330	C2
Data ke 12	306.537	15.685	13.515	15.685	C2
Data ke 13	707.005	15.322	14.048	15.322	C2
Data ke 14	508.279	18.114	15.226	18.114	C2
Data ke 15	308.881	31.718	15.951	31.718	C2
Data ke 16	406.289	16.447	13.748	16.447	C2
Data ke 17	356.647	563.477	544.557	356.647	C1
Data ke 18	124.326	183.124	164.271	124.326	C1
Data ke 19	664.778	48.364	29.640	29.640	C3
Data ke 20	381.219	31.459	13.111	13.111	C2
Data ke 21	405.385	19.364	14.733	19.364	C2
Data ke 22	608.631	26.406	15.992	26.406	C2
Data ke 23	242.312	359.811	341.158	115.189	C1
Data ke 24	497.151	15.180	5.139	15.180	C2
Data ke 25	783.334	29.240	11.051	29.240	C2
Data ke 26	506.715	15.646	13.592	15.646	C2
Data ke 27	308.695	3.963	15.615	15.615	C3
Data ke 28	409.271	3.305	16.282	16.282	C3
Data ke 29	274.356	551.335	532.815	101.613	C1
		5 data	15 data	9 data	

2. Pada Gambar 7 merupakan tampilan untuk *import library* k-medoid dan menampilkan data menggunakan *python*.

```
In [10]: #Library
import numpy as np
from sklearn_extra.cluster import KMedoids
from pandas import read_csv
import pandas as pd
from sklearn.preprocessing import StandardScaler
from sklearn.metrics import silhouette_samples, silhouette_score

In [11]: #Menampilkan dan menyisikkan data
data = pd.read_csv('dataset rizky.csv')
data.head()
```

```
Out[11]:
```

	Pintu Masuk	Januari	Februari	Maret	April	Mei	Juni	Juli	Agustus	September	Oktober	November	Desember
0	Ngurah Rai	451708	436266	441707	476104	483928	549483	604310	602457	589984	565966	492904	544726
1	Soekarno-Hatta	174963	196183	214161	196977	156654	190031	267143	251596	211775	189231	183759	186723
2	Juanda	13792	17389	20497	18431	14529	22485	24913	29180	20462	20895	20780	20546
3	Kualanamu	16253	21815	21129	21613	15573	18935	20929	24623	18913	21518	20798	22431
4	Husein Sastranegara	12524	15127	16342	14738	8029	8652	12581	13984	12869	13420	14616	14951

Gambar 7 Source Code Library K-Medoid Dan Menampilkan Data

3. Pada Gambar 8 merupakan perhitungan algoritma k-medoid menggunakan *python* untuk menampilkan nilai *centroid* yang di dapat dan hasil pengelompokan.

```
In [12]: #Algoritma K-medoid
data_x = data.iloc[:, 1:12]
x_array = np.asarray(data_x)
scaler = StandardScaler().fit(data_x)
x_scaled = scaler.transform(data_x)
kmedoid = KMedoids(n_clusters=3, random_state=123)
kmedoid.fit(x_scaled)
y_kmed = kmedoid.fit_predict(x_scaled)

In [13]: #Menampilkan hasil clustering dan centroid
print(kmedoid.labels_)
print(kmedoid.cluster_centers_)

[[0 0 2 2 2 2 1 2 1 1 1 1 1 1 1 0 2 2 2 1 1 0 1 2 1 1 1 0]
[[ 1.006487    1.27826892  1.36155387  1.14558316  1.04621432  1.13045416
  0.8031675    1.08275279  0.96551463  1.01603227  1.25135664]
[-0.43166918 -0.44756286 -0.45882507 -0.44022978 -0.42723545 -0.42445054
-0.40468661 -0.41861676 -0.40011282 -0.40507696 -0.43168124]
[-0.33053969 -0.34707153 -0.36898069 -0.33441386 -0.34071345 -0.37018298
-0.32621788 -0.31993231 -0.31200336 -0.32638406 -0.35474388]]

In [14]: #Menampilkan data cluster
data['cluster'] = kmedoid.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 = 5
cluster 1 = 15
cluster 2 = 9
```

Gambar 8 Source Code K-Medoid Dan Hasil Cluster

C. Evaluasi Algoritma K-Means Dan K-Medoid

1. Pada Gambar 9 dan 10 merupakan hasil evaluasi algoritma k-means dan k-medoid pada data penelitian ini, hasilnya k-means bernilai 0,8463, sedangkan algoritma k-medoid bernilai 0,4691 yang berarti algoritma k-means lebih baik daripada algoritma k-medoid dan cocok pengelompokan karena lebih mendekati nilai 1.

```
[6]: #Evaluasi silitoutte coefficient k-means
silhouette_avg = silhouette_score(data.iloc[:, 1 : 12], kmeans.labels_)
print(silhouette_avg)

0.8463570739130588
```

Gambar 9 Hasil Akurasi K-Means

```
In [15]: #Evaluasi Silhoutte Coefient k-medoid
silhouette_avg = silhouette_score(x_scaled, y_kmed)
print(silhouette_avg)

0.4691267537728457
```

Gambar 10 Hasil Akurasi K-Medoid

D. Hasil Pengelompokan Algoritma K-Means Dan K-Medoid

1. Pada Tabel 4 merupakan hasil penerapan algoritma k-means menggunakan perhitungan manual excel dan bahasa pemrograman phyton, dapat diketahui hasil pengelompokan k-means dari 29 jalur masuk yaitu C1 sebanyak 24, C2 sebanyak 1 dan C3 sebanyak 4. Sedangkan hasil dari algoritma k-medoid memiliki berbeda, dimana C1 sebanyak 5, C2 sebanyak 15, dan C3 sebanyak 9.

Table 6 Hasil Pengelompokan 2 Algoritma

NO	JALUR MASUK	Manual	Python	Manual	Python
		Algoritma K-Means		Algoritma K-Medoid	
1	Ngurah Rai	C2	C2	C1	C1

2	Soekarno-Hatta	C3	C3	C2	C2
3	Juanda	C1	C1	C3	C3
4	Kualanamu	C1	C1	C2	C2
5	Husein Sastranegara	C1	C1	C3	C3
6	Adi Sucipto	C1	C1	C3	C3
7	Bandara Int. Lombok	C1	C1	C3	C3
8	Sam Ratulangi	C1	C1	C3	C3
9	Minangkabau	C1	C1	C3	C3
10	Sultan Syarif Kasim II	C1	C1	C2	C2
11	Sultan Iskandar Muda	C1	C1	C2	C2
12	Ahmad Yani	C1	C1	C2	C2
13	Supadio	C1	C1	C2	C2
14	Hasanuddin	C1	C1	C2	C2
15	Sultan Badaruddin II	C1	C1	C2	C2
16	Pintu Udara Lainnya	C1	C1	C2	C2
17	Batam	C3	C3	C1	C1
18	Tanjung Uban	C1	C1	C1	C1
19	Tanjung Pinang	C1	C1	C3	C3
20	Tanjung Balai Karimun	C1	C1	C2	C2
21	Tanjung Benoa	C1	C1	C2	C2
22	Tanjung Mas	C1	C1	C2	C2
23	Pintu Laut Lainnya	C3	C3	C1	C1
24	Jayapura	C1	C1	C2	C2
25	Atambua	C1	C1	C2	C2
26	Entikong	C1	C1	C2	C2
27	Aruk	C1	C1	C3	C3
28	Nanga Badau	C1	C1	C3	C3
29	Pintu Darat Lainnya	C3	C3	C1	C1

#### IV. KESIMPULAN DAN SARAN

##### A. Kesimpulan

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

- 1) Penerapan algoritma k-means dan k-medoid pada data penelitian ini menggunakan rumus euclidian distance dan memakai 2 cara perhitungan yaitu manual menggunakan excel dan perhitungan python. Hasil pengelompokan dari 29 jalur menunjukkan hasil yang sama yaitu C1 tingkat kunjungan tinggi sebanyak 24 jalur masuk, C2 tingkat kunjungan sedang sebanyak 1 jalur masuk, dan C3 tingkat kunjungan rendah sebanyak 4 jalur masuk. Sedangkan algoritma k-medoid memiliki hasil yang berbeda yaitu C1 sebanyak 5, C2 sebanyak 15, dan C3 sebanyak 9.
- 2) Hasil tingkat akurasi algoritma k-means dan k-medoid dihitung menggunakan metode Silhouette coefficient, jika nilai mendekati angka 1 berarti hasilnya bagus, jika mendekati -1 berarti hasilnya buruk. Hasil Evaluasi algoritma k-means akurasi bernilai 0,8463, sedangkan hasil evaluasi algoritma k-medoid akurasi bernilai 0,4691, sehingga dapat disimpulkan bahwa hasil pengelompokan algoritma k-means lebih baik dibandingkan algoritma k-medoid..

##### B. Saran

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

- 1) Untuk pihak pengelola sektor pariwisata, diperlukan kerjasama dengan pihak lainnya untuk melakukan lebih banyak promosi melalui media sosial dan pada jalur masuk yang sering dikunjungi untuk memberikan informasi terkait wisata lainnya yang jarang dikunjungi. Serta meningkatkan infrastruktur dan fasilitas pada sektor pariwisata agar para wisatawan asing dapat tertarik untuk berkunjung ke tempat wisata-wisata yang lain, sehingga diharapkan pariwisata bisa menjadi sektor unggulan

karena rancangan strategis dan perkembangan destinasi pariwisata. Dapat menggunakan variabel pendukung lainnya, seperti faktor human error, machine problem, maupun trend problem lainnya pada waktu tertentu dan melakukan prediksi dengan menggunakan metode yang lain seperti Support Vector Machine (SVM), Deep Neural Network (DNN), dan Recurrent Neural Network (RNN) maupun metode lainnya, agar dapat dibandingkan untuk mencari metode yang terbaik.

- 2) Untuk pengembangan penelitian, peneliti selanjutnya dapat melakukan perbandingan menggunakan algoritma lain dan menambah metode evaluasi data.

#### PENGAKUAN

Naskah ilmiah ini adalah sebagian dari penelitian Tugas Akhir milik Rizky Alfian Nugraha dengan judul Penerapan Algoritma K-Means Dan K-Medoid Untuk Mengelompokkan Tingkat Kunjungan Wisatawan Asing Berdasarkan Jalur Masuk Ke Indonesia yang dibimbing oleh Yana Cahyana, M.Kom dan Ayu Ratna Juwita, M.Kom.

#### DAFTAR PUSTAKA

- [1] A. M. Siregar, "Pengelompokan Bidang Laju Pertumbuhan Ekonomi Indonesia Menggunakan Algoritma K-Means," *Accounting Information System*, vol. 2, no. 2, pp. 140–151, 2019.
- [2] Anggraeni, R., "Enhancing the Revisit Intention of Nature-Based Tourism in Indonesia: The Role of Memorable Tourism Experience and Satisfaction," *Management and Business Research Quarterly*, vol. 11, no. 11, pp. 9–19, 2019.
- [3] Paembonan, S., & Abduh, H., "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, vol. 6, no. 2, pp. 48, 2021.
- [4] Seimahaira, S., "Implementasi Data Mining Dalam Menentukan Destinasi Unggulan Berdasarkan Online Reviews Tripadvisor Menggunakan Algoritma K-Means," *Technologia: Jurnal Ilmiah*, vol. 12, no. 1, pp. 53, 2021.
- [5] Suntoro, J., "Data Mining Algoritma K-Medoid Dan Implementasi Menggunakan Bahasa Pemrograman PHP," vol. 9, no. 9, pp. 259–278, 2019.