

Optimasi Pelayanan Kapal Penumpang melalui *Clustering* Penumpang dengan Metode *Silhouette Coefficient*

Tayeb Adi Mulyadi ^{1,*}, Dedi Purnomo ¹

¹ Program Studi Ilmu Komputer, Universitas Budi Luhur, Indonesia

* Correspondence: 2011600588@student.budiluhur.ac.id

Copyright: © 2023 by the authors

Received: 26 Juli 2023 | Revised: 29 Juli 2023 | Accepted: 5 Agustus 2023 | Published: 20 Desember 2023

Abstrak

Meningkatkan efisiensi pelayanan kapal penumpang merupakan tantangan besar bagi industri pelayaran. Penelitian ini bertujuan untuk meningkatkan efisiensi dan kualitas layanan kapal penumpang dengan menggunakan metode klastering penumpang berdasarkan metode *Silhouette Coefficient*. Penelitian ini menggunakan algoritma k-means dan metode *Silhouette Coefficient* agar mendapatkan *cluster* penumpang berdasarkan penumpang naik dan penumpang turun. Metode *Silhouette Coefficient* pada penelitian ini digunakan untuk mencari jumlah *cluster*. Dataset yang digunakan adalah data penumpang dari 1 Januari - 31 Desember 2021. Hasil penelitian menunjukkan bahwa didapatkan 2 *cluster* yaitu *cluster* tertinggi (C1) dan *cluster* terendah (C2) dari algoritma k-means dengan hasil nilai centroid yang didapat untuk *cluster* 1 penumpang naik yaitu 53,414 dan penumpang turun yaitu 54,585, nilai centroid yang didapat untuk *cluster* 2 penumpang naik yaitu 596,396 dan penumpang turun yaitu 532,455. Adapun nilai dengan metode *Silhouette Coefficient* didapatkan 0,704 pada *cluster* 1 untuk *cluster* tertinggi ada 2091 data dengan nilai presentase 57,27% dan 0,548 pada *cluster* 2 untuk *cluster* terendah ada 1560 data dengan nilai presentasi 42,73%. Nilai *Silhouette Coefficient* mendekati nilai 1 yang bahwa pelabuhan muara angke sudah cukup baik dalam optimalisasikan pelayanan penumpang.

Kata kunci: *clustering*; algoritma k-means; *silhouette plot*; data mining

Abstract

Improving the efficiency of passenger ship services is a big challenge for the shipping industry. This research aims to optimize ship passenger services at Muara angke port by segmenting passengers using the Silhouette Coefficient method. This research uses the k-means algorithm and the Silhouette Coefficient method in order to get passenger clusters based on boarding passengers and disembarking passengers. The Silhouette Coefficient method in this research is used to find the number of clusters. The dataset used is passenger data from January 1 - December 31, 2021. The results showed that 2 clusters were obtained, namely the highest cluster (C1) and the lowest cluster (C2) from the k-means algorithm with the results of the centroid value obtained for cluster 1 passengers up, namely 53,414 and passengers down, namely 54,585, the centroid value obtained for cluster 2 passengers up, namely 596,396 and passengers down, namely 532,455. The value with the Silhouette Coefficient method is obtained 0.704 in cluster 1 for the highest cluster there are 2091 data with a percentage value of 57.27% and 0.548 in cluster 2 for the lowest cluster there are 1560 data with a presentation value of 42.73%. The Silhouette Coefficient value is close to the value of 1, which means that the Muara angke port is good enough in optimizing passenger services.

Keywords: *clustering*; k-means algorithm; *silhouette plot*; data mining

PENDAHULUAN

Transportasi laut merupakan bagian penting dari perekonomian setiap negara, khususnya bagi negara kepulauan seperti Indonesia (Priadi et al., 2022). Kapal penumpang, yang



mengoperasikan rute antar pulau dan melayani puluhan ribu penumpang per hari, merupakan moda transportasi laut yang penting (Arianto & Sutrisno, 2021). Meningkatkan kualitas layanan kapal penumpang sangat penting untuk memberikan pengalaman terbaik kepada penumpang (La Murdani & Nanlohy, 2022). Namun, untuk menjawab tantangan peningkatan pelayanan kapal penumpang, beberapa kendala harus diatasi (Dirang & Iriani, 2021). Salah satunya adalah administrasi dan pengaturan penumpang di dalam kapal laut (Erwin, 2022). Kadang-kadang, pada periode tertentu, kapal penumpang harus menghadapi kepadatan penumpang dan kebingungan mengenai kursi penumpang yang benar (Handoko et al., 2021). Hal ini dapat membuat penumpang merasa tidak nyaman dan menurunkan kualitas pelayanan (Dewi & Hanty, 2022).

Untuk mengatasi masalah ini, salah satu solusinya adalah dengan menggunakan data mining dan metode pengelompokan penumpang. Data mining merupakan ekstraksi dari informasi atau model pola yang penting untuk menarik data yang ada di database (Rahayu et al., 2022; Takdirillah, 2020; Uska et al., 2020; Wadanur & Sari, 2022). *Clustering* adalah teknik analisis data yang mencoba mengklasifikasikan data ke dalam kelompok-kelompok yang memiliki kesamaan karakteristik (Sinaga & Yang, 2020). Dalam konteks pelayanan kapal penumpang, *clustering* penumpang dapat membantu dalam mengidentifikasi pola dan kelompok penumpang berdasarkan preferensi dan kebutuhan mereka, memungkinkan penumpang untuk duduk sesuai dengan preferensi berbeda (Nugraha et al., 2022). Ada banyak iterasi model yang dibangun untuk analisis pengelompokan, dan beberapa model akan berfungsi lebih efektif daripada yang lain dalam konteks tertentu dan dengan berbagai jenis data (Setyaningtyas et al., 2022). Algoritma K-Means adalah algoritma yang digunakan dalam penelitian ini. Metode *Silhouette* adalah strategi yang diambil pada penelitian ini.

Algoritma K-Means adalah teknik komputasi yang efisien dan salah satu algoritma *clustering* yang biasa digunakan untuk prosedur *clustering* (Fard et al., 2020). K-Means adalah pendekatan tanpa pengawasan yang dapat digunakan untuk mengatasi kesulitan pengelompokan (Ashari et al., 2022; Nugraha et al., 2023). K' adalah jumlah grup atau *cluster* dalam *clustering* K-Means (Haviluddin et al., 2021). Tujuan dari metodologi data mining ini adalah untuk melakukan pengamatan terhadap setiap nilai atribut individu dan membandingkan pengamatan tersebut dengan nilai rata-rata, atau nilai rata-rata (Supardi & Kanedi, 2020). Pengamatan juga dilakukan dengan kelompok lain untuk menemukan kelompok yang mirip satu sama lain. K-Means adalah pendekatan sederhana dan efektif untuk menemukan *cluster* dalam data (Aditya et al., 2020).

Agar implementasi *clustering* algoritma K-Means seefisien mungkin, pendekatan *Silhouette* digunakan untuk menentukan jumlah *cluster* (Aziz & Rifai, 2022). Untuk menentukan seberapa dekat hubungan suatu titik data dengan klasternya sendiri, teknik ini menghitung *Silhouette* setiap titik dan membandingkan hasilnya (Hidayati et al., 2021). Nilai *Silhouette* bervariasi dari satu hingga negatif, dengan nilai yang lebih tinggi menunjukkan lebih banyak kesesuaian untuk klaster objek itu sendiri dan kurang sesuai untuk klaster (Bagirov et al., 2023). Ketika kesamaan dan pengelompokan dalam data, yang mewakili kualitas konsumen diketahui, strategi promosi dapat dibentuk dengan lebih mudah, yang pada gilirannya memungkinkan taktik menjadi lebih akurat dalam hal ke mana mereka harus diarahkan (Fransiska et al., 2022). Orange3 merupakan aplikasi data mining yang bersifat open source. Aplikasi Orange3 dapat digunakan untuk menganalisis dan memvisualisasikan data (Gurunathan et al., 2021). Ini memungkinkan program digunakan untuk pemilihan eksperimen, pemodelan peramalan, dan sistem rekomendasi. Dalam hal inovasi, kualitas, dan keandalan, aplikasi oranye selalu lebih disukai. Program Orange memungkinkan pengguna untuk berinteraksi dengan data dan melakukan analisis data secara intuitif.

Penelitian sebelumnya menerapkan algoritma k-means untuk pemilihan pelanggan potensial (Romadhona et al., 2022). Selain itu, algoritma k-means ini digunakan untuk

mengelompokkan nilai disiplin siswa (Mawarni & Budi, 2022). Penelitian lain juga dilakukan oleh Abdullah et.al (2022) untuk clustering provinsi di Indonesia terhadap risiko pandemi COVID-19 berdasarkan berdasarkan data COVID-19 dengan metode elbow dan menghasilkan 3 cluster provinsi yang diperoleh. Teknik clustering dilakukan oleh Praseptian et al (2022) untuk clustering tingkat kepuasan pengguna lulusan perguruan tinggi. Data yang gunakan dari merupakan data lulusan dari tahun 2019-2021 dengan jumlah data 51 data lulusan dan menghasilkan 3 cluster menggunakan weka yaitu dimana cluster C1 ada 48 alternatif, C2 ada 1 alternatif, dan C3 ada 2 alternatif. Metode *silhouette coefficient* digunakan untuk mengetahui faktor penyebab stunting dan didapatkan 3 cluster dan nilai *silhouette coefficient* = 0.3 (Salsabila et al., 2022).

Beberapa penelitian sebelumnya menggunakan metode *silhouette coefficient* namun hasil yang diperoleh kurang baik dan pada penelitian ini menggunakan metode *silhouette coefficient* dengan teknik *Euclidean* untuk memperoleh hasil temuan yang lebih baik. Pada temuan sebelumnya menggunakan jumlah data yang cukup rendah yaitu 50-150 data, sedangkan pada penelitian ini menggunakan 3651 data penumpang kapal pelabuhan muara angke. Tujuan ini adalah untuk meningkatkan efisiensi dan kualitas layanan kapal penumpang dengan menggunakan metode klastering penumpang berdasarkan metode *Silhouette Coefficient*. Dengan menerapkan klastering tersebut diharapkan penempatan penumpang di atas kapal dapat dioptimalkan untuk meningkatkan pengalaman penumpang dan meningkatkan efisiensi pengelolaan kapal. Sehingga memberikan kontribusi positif bagi industri pelayaran dan pemerintah Indonesia dalam upaya meningkatkan pelayanan transportasi laut.

METODE

Penelitian ini menggunakan pendekatan kuantitatif karena jenis penelitian ini sistematis, terstruktur, dan disusun dari awal hingga akhir. Akibatnya, penelitian ini cenderung menggunakan teknik analisis angka statistik. Pada tahap pertama, dilakukan dengan literature review, dimana diupayakan referensi publikasi ilmiah terkait penelitian dan buku-buku untuk mendukung upaya penelitian ini. Pada tahap kedua adalah mengumpulkan data. Dataset yang digunakan pada langkah ini diperoleh dari Dinas Perhubungan DKI Jakarta. Ada sebanyak 3651 data diambil dari periode 1 Januari dan 31 Desember 2021. File yang sudah diambil tidak bisa langsung digunakan namun harus dibersihkan terlebih dahulu sebelum dapat diproses ketahap selanjutnya.

Pada tahap ketiga adalah pre-processing data dimana ada datayang harus diganti yaitu format dan transformasi data harus diperbarui. Selanjutnya mencari missing value atau nominal yang kosong atau akan diisi atau dihilangkan. Setelah dibersihkan, data disimpan dalam bentuk format csv dapat diproses oleh program aplikasi *Orange3*. Setelah itu tahap keempat yaitu pemilihan algoritma dan pembentukan model data mining menggunakan orange3 seperti pada gambar 1. Setelah pembuatan model data mining selesai selanjutnya data yang telah bersih tadi diimpor ke dalam aplikasi *Orange3*. Kemudian aplikasi Orange melakukan proses *clustering* secara otomatis, menghilangkan atribut yang tidak dibutuhkan untuk mencari *cluster* secara manual. Fitur k-means pada aplikasi *Orange* menentukan jumlah *cluster* dengan melihat nilai *Silhouette Coefficient* tertinggi. Setelah mengaktifkan fitur k-means, terdapat dua *cluster* dengan nilai siluet 0,707, yaitu C1 dengan nominal tinggi dan C2 dengan nominal rendah. Setelah temuan *clustering* muncul, tahap terakhir yaitu evaluasi dimana pada tahapan ini melakukan analisis terhadap hasil *clustering* untuk menghasilkan informasi baru.

HASIL DAN PEMBAHASAN

Hasil

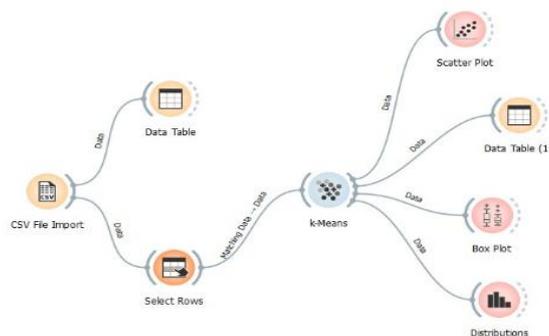
Data yang diambil adalah data pelabuhan muara angke dengan variabel yang digunakan berupa data penumpang kapal naik, penumpang kapal turun, jenis kapal, kapal berangkat,

kapal tiba lalu dilakukan memasukan data dalam bentuk excel. Data yang digunakan merupakan data penumpang periode bulan 1 januari – 31 Desember 2021 sebanyak 3651 data penumpang.

Tabel 1. Dataset pelabuhan muara angke

Tanggal	Pelabuhan	Kapal berangkat	Kapal tiba	Penumpang Naik	Penumpang Turun
2021-01-01	Muara Angke	16	19	896	1276
2021-01-02	Muara Angke	20	16	863	1101
2021-01-03	Muara Angke	11	15	277	1033
2021-01-04	Muara Angke	10	6	340	231
2021-01-05	Muara Angke	6	8	172	204
2021-01-06	Muara Angke	8	6	213	282
2021-01-07	Muara Angke	6	8	121	127
2021-01-08	Muara Angke	6	13	271	176
2021-01-09	Muara Angke	16	5	807	83
2021-01-10	Muara Angke	5	14	172	743

Pada tabel 1 merupakan dataset kapal yang sudah clean atau bersih sehingga bisa dilanjutkan untuk pemrosesan data dan pemodelan data dengan menggunakan algoritma k-means dan *tools orange3*. Sebelum menganalisis data *clustering* maka langkah pertama yaitu membuat pemodelan algoritma k-means menggunakan *orange3* seperti pada gambar 1.

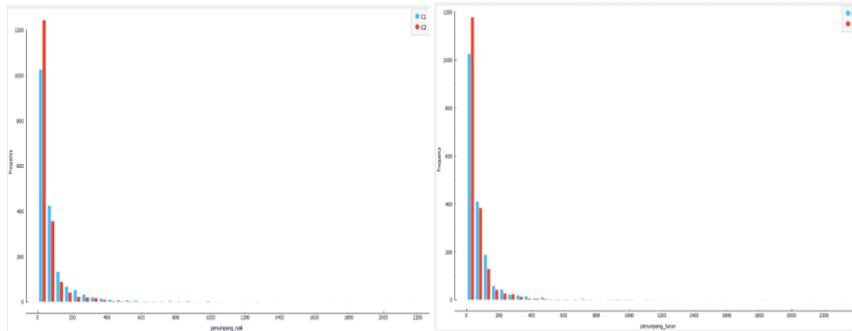


Gambar 1 . Permodelan *clustering* dengan *orange3*

Setelah pemodelan dilakukan oleh *tools orange3* maka selanjutnya adalah mengimport data csv kedalam pemodelan algoritma kemudian diolah menggunakan algoritma k-means, dimana dataset akan dibagi menjadi 2 bagian atau 2 kluster (2 *cluster*) yaitu penumpang naik dan penumpang turun. Setelah di bentuk maka langkah selanjutnya adalah melihat inisialisasi jumlah *cluster* yaitu 2 buah, maka didapatkan hasil *cluster* yang terbentuk yakni 2 *cluster*. Sesuai dengan jumlah dataset pada pembentukan *clusterisasi* ini dataset dibagi menjadi 2 yakni pada *cluster* 1 terdapat 2091 data dan pada *cluster* 2 terdapat 1559 data yang terbentuk. Adapun distribusi data penumpang naik dan penumpang turun dapat dilihat pada gambar 2.

Pada gambar 2 ini merupakan distribusi data kapal untuk penumpang naik dan penumpang turun. Distribusi data ini digunakan untuk memberikan gambaran tentang bagaimana data terdistribusi dalam setiap kluster hasil dari algoritma *clustering* yang digunakan. Ini memungkinkan kita untuk memahami secara visual bagaimana data terbagi dalam kelompok-kelompok yang terbentuk. Setelah melihat gambaran dari data yang akan

digunakan maka dapat lihat hasil *cluster* kapal untuk penumpang naik tabel 2 dan penumpang turun tabel 3.



Gambar 2. Distribusi penumpang naik dan penumpang turun

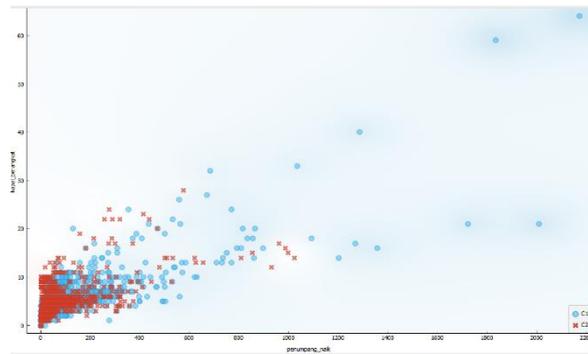
Tabel 2. Hasil *cluster* 1 untuk penumpang naik dan penumpang turun

<i>Cluster</i>	Tanggal	Pelabuhan	Kapal berangkat	Kapal tiba	Penumpang Naik	Penumpang turun
C1	2/6/2021 0:00	Tidung	1	2	0	52
C1	2/7/2021 0:00	Tidung	1	1	0	0
C1	2/8/2021 0:00	Tidung	1	3	0	51
C1	2/9/2021 0:00	Tidung	4	4	18	9
C1	2/10/2021 0:00	Tidung	4	5	7	18
C1	2/11/2021 0:00	Tidung	5	4	24	14
C1	2/12/2021 0:00	Tidung	4	5	12	63
C1	2/13/2021 0:00	Tidung	4	5	17	41
C1	2/14/2021 0:00	Tidung	5	4	55	23

Tabel 3. Hasil *cluster* 2 untuk penumpang naik dan penumpang turun

<i>Cluster</i>	Tanggal	Pelabuhan	Kapal berangkat	Kapal Tiba	Penumpang naik	Penumpang turun
C2	7/3/2021 0:00	Muara angke	6	3	212	47
C2	7/4/2021 0:00	Muara angke	3	6	35	148
C2	7/5/2021 0:00	Muara angke	5	3	90	46
C2	7/6/2021 0:00	Muara angke	3	4	39	36
C2	7/7/2021 0:00	Muara angke	4	3	35	24
C2	7/8/2021 0:00	Muara angke	3	4	30	47
C2	7/9/2021 0:00	Muara angke	3	4	32	39
C2	7/10/2021 0:00	Muara angke	5	2	55	25
C2	7/11/2021 0:00	Muara angke	2	3	17	39

Pada tabel 2 dan tabel 3 merupakan hasil *clustering* menggunakan tools orange3. Kemudian dari hasil *clustering* tersebut bisa dilihat juga hasil *cluster* menggunakan *Scatter plot* seperti pada gambar 3.



Gambar 3. Scatter plot clustering penumpang

Pada gambar 3 merupakan Scatter plot clustering K-Means untuk pengelompokan penumpang naik dan penumpang turun dengan nilai $k=2$, dimana Scatter plot ini membantu untuk memvisualisasikan hasil clustering pada tabel 2 dan tabel 3 memudahkan dalam memahami distribusi data. Setelah dilakukan clustering data menggunakan algoritma k-means maka selanjutnya dilakukan analisis data menggunakan Silhouette plot seperti pada tabel 4.

Tabel 4. Hasil perhitungan silhouette plot untuk cluster 1 dan cluster 2

Cluster	Pelabuhan	Kapal Berangkat	Kapal Tiba	Penumpang Naik	Penumpang Turun	Silhouette
C1	Tidung	1	2	0	52	0,702297888
C1	Tidung	1	1	0	0	0,702588463
C1	Tidung	1	3	0	51	0,702871562
C1	Tidung	4	4	18	9	0,703147124
C1	Tidung	4	5	7	18	0,703415084
C1	Tidung	5	4	24	14	0,703675378
C1	Tidung	4	5	12	63	0,703927935
C1	Tidung	4	5	17	41	0,704172688
C1	Tidung	5	4	55	23	0,704409564
C2	Muara Angke	6	3	212	47	0,505017281
C2	Muara Angke	3	6	35	148	0,511768001
C2	Muara Angke	5	3	90	46	0,518327784
C2	Muara Angke	3	4	39	36	0,524697827
C2	Muara Angke	4	3	35	24	0,530879935
C2	Muara Angke	3	4	30	47	0,536876445
C2	Muara Angke	3	4	32	39	0,542690152
C2	Muara Angke	5	2	55	25	0,548324239

Setelah hasil 2 cluster telah dibentuk maka dapat dilihat 2 perbandingan antara cluster 1 dan cluster 2 seperti pada tabel 5. Bahwa dataset yang digunakan pada penelitian ini 3651 data,

setelah di proses *cluster* dengan algoritma k-means dan metode *Silhouette Coefficient* menghasilkan 2 *cluster* yaitu *cluster* 1 untuk *cluster* tertinggi ada 2091 data dengan nilai presentase 57,27% dan *cluster* 2 untuk *cluster* terendah ada 1560 data dengan nilai presentasi 42,73%.

Tabel 5. Perbandingan *Cluster* dan Jumlah Data

<i>Cluster</i>	Jumlah Data	Presentase
C1	2091	57,27%
C2	1560	42,73%
Total	3651	100%

Pembahasan

Bedasarkan hasil penelitian dan analisa data untuk pengelompokan penumpang naik dan penumpang turun pada kapal di pelabuhan muara angke Jakarta Utara. Pada penelitian ini dibentuk nilai $k=2$ atau 2 *cluster* dengan algoritma k-means, pembentukan *cluster* menggunakan tools orange3 seperti pada gambar 3. Setelah dilakukan pembentukan *cluster* maka dapat dilakukan analisis data dengan menggunakan distribution data, scatter plot, hasil *clustering* dan *Silhouette* plot. Penelitian ini menggunakan metode *Silhouette* ketika penentuan jumlah *cluster* secara otomatis dilakukan oleh aplikasi Orange3 pada fungsi k-means, sehingga lebih cepat dan efisien, sedangkan metode siku tidak dapat langsung mengetahui hasilnya, melainkan harus membandingkan harganya. nilai fokus harus memakan waktu kurang. *Cluster* dibentuk dengan kuantitas 2 yaitu C1 dengan harga nominal rendah dan C2 dengan harga nominal tinggi. Pada tabel 2 dan tabel 3 merupakan hasil dari *clustering* dengan nilai $k=2$ dimana untuk algoritma k-means ini di hitung nilai centroid atau titik pusatnya. Nilai centroid yang didapat untuk *cluster* 1 penumpang naik yaitu 53,414 dan penumpang turun yaitu 54.585. Nilai centroid yang didapat untuk *cluster* 2 penumpang naik yaitu 596.396 dan penumpang turun yaitu 532,455. Nilai centroid ini berguna untuk menghitung jarak antara data point dengan centroidnya agar lebih terstruktur.

Setelah dilakukan perhitungan centroidnya maka dapat dilakukan *Scatter plot* seperti pada gambar 3. *Scatter plot* ini merupakan visualisasi dari hasil *cluster* 1 dan *cluster* 2 yang diberikan warna berbeda antar *cluster* dimana *cluster* 1 berwarna biru dan *cluster* 2 berwarna merah. Kemudian setelah visual dan pembentukan *cluster* langkah selanjutnya yaitu menganalisis data menggunakan *Silhouette* plot. Nilai *Silhouette* plot bisa dilihat di tabel 4. Hasil analisis *Silhouette* plotnya menunjukkan nilai 0,704 dimana nilai *Silhouette* mendekati 1 menunjukkan bahwa objek tersebut cocok dengan *clusternya* dan jarak antara objek dengan objek di *cluster* lainnya relatif lebih besar sehingga pelabuhan muara angke ini dalam pelayanan penumpang sudah cukup baik. Ini menandakan bahwa *cluster* tersebut terpisah dengan baik dari *cluster* lainnya. Kemudian perbandingan antara *cluster* yang terbentuk dapat dilihat pada tabel 5. Setelah melakukan proses *clustering* dengan algoritma k-means dan menggunakan metode *Silhouette Coefficient* untuk mengevaluasi hasilnya, diperoleh hasil sebanyak 2 *cluster* berdasarkan pertimbangan tujuan penelitian ini. *Cluster* pertama disebut *cluster* 1 dan memiliki 2091 data, sedangkan *cluster* kedua disebut *cluster* 2 dan berisi 1560 data.

Hasil *clustering* dari penelitian ini menunjukkan koefisien *silhouette* yang cukup tinggi, yang menunjukkan adanya dua pengelompokan atau kelompok yang berbeda dan bermakna. Hasil analisis *clustering*, dapat diamati bahwa *cluster* 1 berisi volume data yang lebih besar dibandingkan dengan *cluster* 2. Pengamatan ini menunjukkan bahwa titik-titik data dalam *cluster* 1 menunjukkan tingkat kesamaan yang lebih besar dalam atributnya dibandingkan dengan titik data dalam *cluster* 2. Kluster yang memiliki jumlah data lebih besar dapat dianggap sebagai "*cluster*". Istilah "tertinggi" digunakan dalam konteks ini untuk menunjukkan bahwa

variabel tertentu memiliki jumlah pengamatan yang lebih banyak dibandingkan variabel lainnya. Pendekatan koefisien *silhouette* digunakan untuk tujuan memvalidasi hasil pengelompokan dan menentukan jumlah kluster yang paling sesuai. Nilai koefisien *silhouette* yang tinggi untuk $k=2$ menunjukkan bahwa pembagian data menjadi dua kelompok berbeda menunjukkan tingkat kualitas yang menguntungkan.

Sementara itu untuk penelitian sebelumnya menghasilkan 8 *cluster* obat dengan menggunakan metode *elbow*, dan memperoleh nilai *Silhouette coefficient* 0,4854 (Paembonan & Abduh, 2021). Pada penelitian ini, terbentuk 2 *cluster* dengan metode *Euclidean* dan nilai *Silhouette coefficient* 0,704. Perbandingan nilai *Silhouette coefficient* jauh lebih baik dari penelitian sebelumnya. Penelitian sebelumnya menggunakan perhitungan manual dan pembuatan aplikasi untuk rekomendasi obat. Sementara itu, penelitian ini menggunakan aplikasi *Orange3* untuk optimalisasikan pelayanan penumpang kapal. Selanjutnya temuan sebelumnya mengelompokkan data obat berdasarkan rekomendasi penggantian obat untuk pasien, sedangkan penelitian kami menggunakan aplikasi *Orange3*.

SIMPULAN

Hasil temuan kami dengan menggunakan algoritma k-means dan metode *Silhouette coefficient* dapat digunakan untuk mengoptimalkan pelayanan kapal penumpang di pelabuhan Muara Angke. Analisis dan pengujian dengan algoritma k-means pada aplikasi *Orange3* memungkinkan data dikelompokkan berdasarkan penumpang yang naik dan penumpang yang turun dan menghasilkan 2 *cluster* yaitu *cluster* penumpang tinggi (C1) dan *cluster* penumpang rendah (C2). Penerapan algoritma k-means didapatkan hasil bahwa pelayanan penumpang cukup baik di pelabuhan Muara Angke.

REFERENSI

- Abdullah, D., Susilo, S., Ahmar, A. S., Rusli, R., & Hidayat, R. (2022). The application of K-means clustering for province clustering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data. *Quality and Quantity*, 56(3), 1283–1291. <https://doi.org/10.1007/s11135-021-01176-w>
- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 4(1), 51–58. <https://doi.org/10.30865/mib.v4i1.1784>
- Arianto, D., & Sutrisno, A. (2021). Kajian Antisipasi Pelayanan Kapal dan Barang di Pelabuhan Pada Masa Pandemi Covid–19. *Jurnal Penelitian Transportasi Laut*, 22(2), 97–110. <https://doi.org/10.25104/transla.v22i2.1682>
- Ashari, I. A., Negara, I. S. M., & Sumantri, R. B. B. (2022). Evaluasi Pembayaran Keuangan Siswa berdasarkan Penghasilan Wali Siswa menggunakan Metode Clustering K-Means. *Edumatic: Jurnal Pendidikan Informatika*, 6(2), 324–333.
- Aziz, S. M., & Rifai, N. A. K. (2022). Pengelompokan Ekspor Kopi Menurut Negara Tujuan Menggunakan Metode K-Means Clustering dengan Silhouette Coefficient. *Bandung Conference Series: Statistics*, 2(2), 416–424. <https://doi.org/10.29313/bcss.v2i2.4536>
- Bagirov, A. M., Aliguliyev, R. M., & Sultanova, N. (2023). Finding compact and well-separated clusters: Clustering using silhouette coefficients. *Pattern Recognition*, 135. <https://doi.org/10.1016/j.patcog.2022.109144>
- Dewi, A. M., & Hanty, F. (2022). Kualitas Sumber Daya Manusia Transportasi Laut Di Revolusi Industri 4.0 Menuju Era Pelabuhan Pintar. *Majalah Ilmiah Bahari Jogja*, 20(2), 204-210. <https://doi.org/10.33489/mibj.v20i2.298>
- Dirang, M., & Iriani, I. (2021). Analisis Kualitas Pelayanan Terhadap Tingkat Kepuasan Pengguna Ruang Tunggu Penumpang Pelabuhan Tanjung Perak Dengan Metode Servqual Dan Triz. *JUMINTEN*, 2(1), 49–60. <https://doi.org/10.33005/juminten.v2i1.141>

- Erwin, R. (2022). Tanggung Jawab Negara Untuk Mencegah Terjadinya Kecelakaan Kapal Transportasi Laut Menurut Hukum Internasional Dan Hukum Nasional. *SUPREMASI : Jurnal Hukum*, 4(2), 177–199. <https://doi.org/10.36441/supremasi.v4i2.716>
- Fard, M. M., Thonet, T., & Gaussier, E. (2020). Deep k-Means: Jointly clustering with k-Means and learning representations. *Pattern Recognition Letters*, 138. <https://doi.org/10.1016/j.patrec.2020.07.028>
- Fransiska, N. N., Anggraeni, D. S., & Enri, U. (2022). Pengelompokan Data Kemiskinan Provinsi Jawa Barat Menggunakan Algoritma K-Means dengan Silhouette Coefficient. *TEMATIK*, 9(1), 29–35. <https://doi.org/10.38204/tematik.v9i1.901>
- Gurunathan, V., Hamre, J., Klimov, D. K., & Jafri, M. S. (2021). Data mining of molecular simulations suggest key amino acid residues for aggregation, signaling and drug action. *Biomolecules*, 11(10), 1–14. <https://doi.org/10.3390/biom11101541>
- Handoko, Churniawan, E., & Rozak, F. (2021). Analisis Respon Penumpang Terhadap Penerapan New Normal pada Layanan Kereta Api Jarak Jauh di Pulau Jawa. *Jurnal Perkeretaapian Indonesia (Indonesian Railway Journal)*, 5(1), 36–46. <https://doi.org/10.37367/jpi.v5i1.127>
- Haviluddin, H., Patandianan, S. J., Putra, G. M., Puspitasari, N., & Pakpahan, H. S. (2021). Implementasi Metode K-Means Untuk Pengelompokan Rekomendasi Tugas Akhir. *Informatika Mulawarman: Jurnal Ilmiah Ilmu Komputer*, 16(1), 13–18. <https://doi.org/10.30872/jim.v16i1.5182>
- Hidayati, R., Zubair, A., Pratama, A. H., & Indana, L. (2021). Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering Silhouette Coefficient Analysis in 6 Measuring Distances of K-Means Clustering. *Techno.COM*, 20(2), 186–197. <https://doi.org/10.33633/tc.v20i2.4556>
- La Murdani, A. I., & Nanlohy, Y. W. A. (2022). Implementasi Model Autoregressive Integrated Moving Average (Arima) Untuk Peramalan Jumlah Penumpang Kapal Laut Di Pelabuhan Ambon. *VARIANCE: Journal of Statistics and Its Applications*, 3(2), 81–90. <https://doi.org/10.30598/variancevol3iss2page81-90>
- Mawarni, Q. I., & Budi, E. S. (2022). Implementasi Algoritma K-Means Clustering Dalam Penilaian Kedisiplinan Siswa. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(4), 522–528. <https://doi.org/10.30865/json.v3i4.4242>
- Nugraha, H. S., Mutaqin, H., Fathah, A., & Juliane, C. (2023). Mengidentifikasi Strategi Promosi pada Jasa Penjualan Saldo Digital menggunakan Pendekatan Clustering. *Edumatic: Jurnal Pendidikan Informatika*, 7(1), 11–19. <https://doi.org/10.29408/edumatic.v7i1.7385>
- Nugraha, N. B., Alimudin, E., & Bonifacius, V. I. (2022). Implementasi K-Means Clustering Pada Sistem Pakar Penentuan Jenis Sayuran. *Journal of Innovation Information Technology and Application (JINITA)*, 4(2), 133–141. <https://doi.org/10.35970/jinita.v4i2.1627>
- Paembonan, S., & Abduh, H. (2021). Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat. *PENA TEKNIK: Jurnal Ilmiah Ilmu-Ilmu Teknik*, 6(2). https://doi.org/10.51557/pt_jiit.v6i2.659
- Praseptian M, D., Fadlil, A., & Herman, H. (2022). Penerapan Clustering K-Means untuk Pengelompokan Tingkat Kepuasan Pengguna Lulusan Perguruan Tinggi. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 6(3), 1693–1700. <https://doi.org/10.30865/mib.v6i3.4191>
- Priadi, A. A., Ari, B., Sugiarto, R., & Nurullah, P. (2022). Biaya Logistik Sektor Transportasi Laut Dan Pengaruhnya Terhadap PDB Nasional. *Jurnal Transportasi Multimoda*, 19(2), 25–34. <https://doi.org/10.25104/mtm.v19i2.2042>

- Rahayu, S., Yumarlin, M. Z., Bororing, J. E., & Hadiyat, R. (2022). Implementasi Metode K-Nearest Neighbor (K-NN) untuk Analisis Sentimen Kepuasan Pengguna Aplikasi Teknologi Finansial FLIP. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 98–106. <https://doi.org/10.29408/edumatic.v6i1.5433>
- Romadhona, W., Nugroho, B., & Alim Murtopo, A. (2022). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K-Means. *Jurnal Minfo Polgan*, 11(2), 100–104. <https://doi.org/10.33395/jmp.v11i2.11797>
- Salsabila, S., Aina, Widiariyah, T., & Sudarno, S. (2022). Metode K-Harmonic Means Clustering Dengan Validasi Silhouette Coefficient (Studi Kasus : Empat Faktor Utama Penyebab Stunting 34 Provinsi di Indonesia Tahun 2018). *Jurnal Gaussian*, 11(1). <https://doi.org/10.14710/j.gauss.v11i1.34003>
- Setyaningtyas, S., Nugroho, B. I., & Arif, Z. (2022). Tinjauan Pustaka Sistematis: Penerapan Data Mining Teknik Clustering Algoritma K-Means. *Jurnal Teknoif Teknik Informatika Institut Teknologi Padang*, 10(2), 52–61. <https://doi.org/10.21063/jtif.2022.v10.2.52-61>
- Sinaga, K. P., & Yang, M. S. (2020). Unsupervised K-means clustering algorithm. *IEEE Access*, 8, 80716–80727. <https://doi.org/10.1109/ACCESS.2020.2988796>
- Supardi, R., & Kanedi, I. (2020). Implementasi Metode Algoritma K-Means Clustering pada Toko Eidelweis. *Jurnal Teknologi Informasi*, 4(2), 270–277. <https://doi.org/10.36294/jurti.v4i2.1444>
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Penjualan Bisnis Ritel. *Edumatic: Jurnal Pendidikan Informatika*, 4(1), 37–46. <https://doi.org/10.29408/edumatic.v4i1.2081>
- Uska, M., Wirasasmita, R., Usuluddin, U., & Arianti, B. (2020). Evaluation of Rapidminer-Application in Data Mining Learning using PeRSIVA Model. *Edumatic: Jurnal Pendidikan Informatika*, 4(2), 164–171. <https://doi.org/10.29408/edumatic.v4i2.2688>
- Wadanur, A., & Sari, A. A. (2022). Implementasi Algoritma Apriori dan FP-Growth pada Penjualan Spareparts. *Edumatic: Jurnal Pendidikan Informatika*, 6(1), 107–115. <https://doi.org/10.29408/edumatic.v6i1.5470>