**JURNAL ILMIAH ELEKTRONIKA DAN KOMPUTER**, Vol.14, No.2, Desember 2021, pp. 368 - 372
p-ISSN : 1907-0012 (print)
e-ISSN : 2714-5417 (online)
http://journal.stekom.ac.id/index.php/elkom · page 368

# Analisa Performa K-Means dan DBSCAN dalam *Clustering* Minat Penggunaan Transportasi Umum

**Ariel Kristianto**[1]
[1]Institut Informatika Indonesia
Pattimura No. 3 (d/h Raya Sukomanunggal Jaya No.3) Surabaya, Jawa Timur. , (031) 734-6375, e-mail: kristiantoariel@gmail.com , ariel@ikado.ac.id

**ARTICLE INFO**

Article history:

Received 25 November 2021
Received in revised form 28 November 2021
Accepted 28 November 2021
Available online 1 Desember 2021

***ABSTRACT***

Public transportation is one of the important modes of transportation and is the backbone of transportation in Indonesia. The development of public transportation is also supported by the government, this government support is evident in the national policy, namely the National Medium Term Development Plan (RPJMN). Although public transportation is an effective mode of transportation, it also has obstacles in its development, namely how to meet customer desires in choosing a mode of transportation. There are several variables that are the focus of this research, namely age, gender, income, cost, speed, comfort, safety, efficiency and flexibility. The search for influential variables will use the K-Means and DBSCAN clustering algorithms, these two algorithms are also compared to their performance to find a better algorithm. The results of the Silhouette Coefficient show that DBSCAN has a better performance with a value of 0.99 than K-Means with a value of 0.86. The variables that affect the interest in using public transportation are the most important ones related to cost, speed, comfort, safety, efficiency and flexibility.

Keywords: K-Means, DBSCAN, Clustering, Public Transportation

## 1. Introduction

Sarana dan prasarana adalah salah satu hal yang penting dalam proses pembangunan, pengembangan dan kemajuan suatu daerah bahkan suatu bangsa. Peningkatan dan pengembangan sarana dan prasarana dilakukan tidak hanya pembangunan infrastruktur seperti jalan namun juga terkait dengan transportasi yang handal dan terintegras. Transportasi memiliki peranan penting bagi berkembangnya suatu daerah, transportasi mendukung berbagai kegiatan, tidak hanya ekonomi namun juga sosial budaya dan lain sebagainya (1).

Transportasi umum adalah salah satu sarana transportasi yang lebih efisien daripada transportasi pribadi atau kendaraan pribadi (2). Penggunaan transportasi umum memiliki manfaat yang tidak hanya terkait ekonomi, namun juga menyangkut manfaat sosial dan juga manfaat lingkungan. Pengembangan transportasi umum juga didukung oleh pemerintah, dukungan pemerintah ini terbukti dalam kebijakan nasional yaitu Rencana Pembangunan Jangka Menengah Nasional (RPJMN) yang menjadikan transportasi umum menjadi tulang punggung trasnportasi di Indonesia (3).

Meskipun transportasi umum adalah sebuah moda transportasi yang efektif, tetapi juga memiliki kendala dalam pengembangannya, yaitu bagaimana memenuhi keinginan pelanggan dalam memilih moda transportasi. Ada beberapa faktor yaitu waktu perjalanan, jarak dan biaya

*Received November 25, 2021; Revised November 28, 2021; Accepted November 28, 2021*

perjalanan(2). Karakteristik pengguna juga mempengaruhi pemilihan dalam penggunaan transportasi umum apa yang mereka inginkan. Karakteristik tersebut antara lain : pendapatan, fasilitas, keamanan, kenyamanan, aksesbilitas dan konektivitas mempengaruhi pemilihan moda transportasi umum (2).

Pengolahan data dapat dilakukan dengan berbagai macam algoritma seperti *Parallelepiped, Minimum Distance, Mahalanobis Distance, Maximum Likelihood, Naive Bayesian, k-Nearest Neighbor, linear regression, Isodata, k-Means, Improved Split and Merge Classification (ISMC), Adaptive Clustering,* dan DBSCAN (4). K-means adalah salah satu algoritma yang mampu mengelompokkan data atau biasa disebut *clustering* ke dalam satu atau beberapa kelompok atau *cluster*, Sehingga terciptalah sebuah kelompok kluster yang berisi data dengan karakteristik yang sama (5). *Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN) juga merupakan salah satu algorita yang digunakan dalam *clustering*, DBSCAN memiliki karalteristik yaitu algoritma ini mengelompokkan data yang berdasar kepada kepadatan data (*density*). Kepadatan dalam DBSCAN menghasilkan tiga macam *state* dari setiap data, yaitu *core* (inti), *border* (batas), dan *noise* (gangguan)(4)(6)(7).

Penelitian ini bertujuan untuk mencari apasaja faktor penentu pemilihan transportasi umum, juga akan membandingkan hasil dari kedua algoritma diatas yaitu K-Means dan DBSCAN.

## 2. Research Method

### Jenis dan Sumber Data

Data yang digunakan adalah data primer, data diperoleh dengan menyebarkan kuisioner kepada beberapa masyarakat dengan berbagai latar belakang pekerjaan, pendidikan, dan tujuan bepergian. Kuisioner disebar dengan beberapa cara, menemui responden dan meminta responden mengisi langsung pada device yang tersedia, juga melalui link yang telah disediakan. Jumlah responden dalam penelitian ini ada lebih kurang 400 responden, yang menghasilkan 388 kuisioner valid. Ada beberapa variabel yang digunakan dalam penelitian ini, dari segi moda transportasi : bus, kereta, pesawat, transportasi online. Dari segi latar belakang konsumen : pendapatan, usia, jenis kelamin. Dari segi kendaraan ada : biaya, kecepatan, kenyamanan, keamanan, efisiensi dan flexibilitas.

### Teknik Pengolahan Data

Teknik pengolahan data dalam penelitian ini akan menggunakan software R yang akan menggunakan metode K-Means dan DBSCAN. Tahapan penelitian akan dijabarkan sebagai berikut :

1. Pembuatan kuisioner
2. Pengumpulan dan *preprocessing* data
3. Pengelompokan data menggunakan metode K-Means.
4. Pengelompokan data menggunakan metode DBSCAN.
5. Pengujian kelompok menggunakan uji *silhouette*.
6. Melakukan analisis dari *output* yang dihasilkan dan memberikan kesimpulan.

Pembuatan kuisioner dilakukan dengan menggunakan beberapa variabel, variabel yang terkait dengan latar belakang konsumen, yaitu usia {1}, jenis kelamin {2}, dan pendapatan {3}. Usia akan dibagi menjadi beberapa kelompok usia antara lain. 11-20 tahun, 21-30 tahun, 31-40 tahun, 41-50 tahun, 51-60 tahun, >61 tahun. Jenis kelamin terbagi menjadi dua yaitu pria dan wanita. Variabel pendapatan adalah pendapatan yang diterima responden baik berupa gaji, tunjangan maupun uang saku yang rutin didapatkan setiap bulannya. Variabel pendapatan ini terbagi menjadi beberapa kelompok yaitu Kurang dari Rp 1.000.000, Rp 1.000.000 - Rp 4.000.000, Rp

p-ISSN: 1907-0012 e-ISSN : 2714-5417

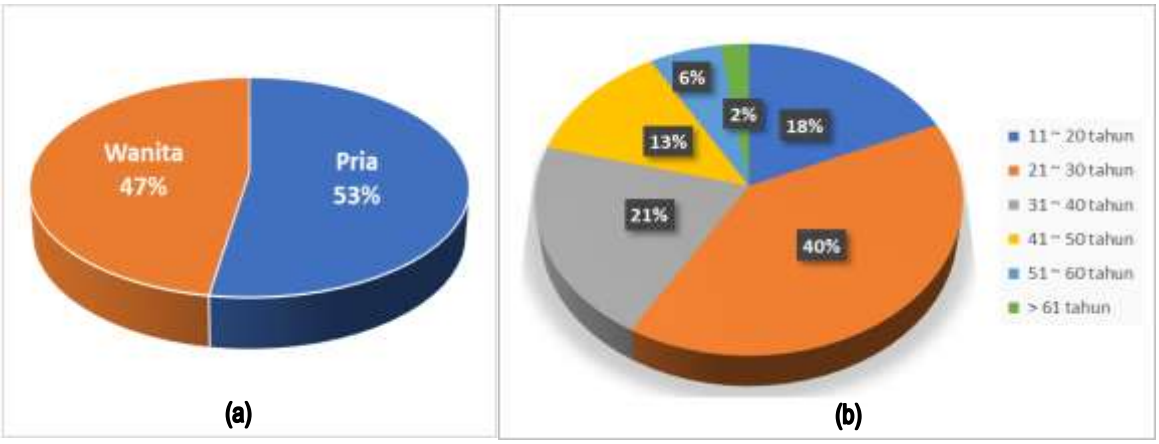
4.000.001 - Rp 8.000.000 Rp 8.000.001 - Rp 12.000.000, Rp 12.000.001 - Rp 16.000.000, Diatas Rp 16.000.001.

Variabel terkait kendaraan ada biaya {4}, kecepatan {5}, kenyamanan {6}, keamanan {7}, efisiensi {8} dan flekibilitas {9}. Biaya adalah terkait harga tiket untuk dapat menggunakan moda transportasi umum tersebut. Kecepatan adalah terkait waktu tempuh yang diperlukan untuk mencapai tujuan, semakin sedikit waktu tempuh yang dibutuhkan artinya semakin cepat moda transportasi tersebut. Kenyamanan adalah terkait fasilitas yang ditawarkan moda transportasi yang ada, mencakup jok atau kursi kendaraan, makanan, minuman, pendingin udara atau AC, dan sebagainya. Keamanan terkait dengan sekuritas yang ditawarkan moda transportasi tersebut, seperti cara mengemudi, kemanan dari tindak kriminal dan sebagainya. Efisiensi terkait ketepatan penggunaan alat transportasi tersebut dengan tujuan yang diinginkan jadi semisal tujuan ada di kota A, apakah alat transportasi tersebut dapat menjangkau kota tersebut atau harus transit dahulu sehingga tidak bisa langsung menuju kota tersebut. Fleksibilitas adalah suatu kondisi apakah alat transportasi tersebut dapat menyesuaikan kemauan penggunanya semisal pengguna ingin berhenti ke ATM, pengguna ingin berhenti ke toilet, apakah alat transportasi tersebut dapat memenuhi kemauan pengguna yang seperti ini.

Pengelompokan data menggunakan algoritma K-Means ataupun DBSCAN dimulai dengan *preprocessing* data, yaitu proses untuk merapikan, menyeragamkan data. Proses penyeragaman data adalah proses untuk menyamakan bentuk data menjadi suatu hal yang homogen sehingga tidak terjadi anomali data dalam proses klastering berikutnya. Pengelompokan menggunakan K-Means dan DBSCAN akan menghasilkan beberapa *cluster* atau beberapa kelompok yang mengelompokkan minat seseorang menggunakan transportasi tersebut berdasarkan variabel yang telah disebutkan sebelumnya. untuk menilai apakah *cluster* yang diciptakan oleh algoritma tersebut baik atau tidak digunakanlah *silhouette coefficient*. Metode ini akan memberitahu tingkat efektivitas penggunaan algoritma tersebut, metode ini akan mengukur sebaran dalam *cluster* tersebut dengan mengukur jarak setiap point dan point lainnya di dalam *cluster* tersebut. Jika nilainya mendekati 1 maka mengindikasikan jarak antar titik pada *cluster* tersebut kecil atau semakin rapat, ini menandakan *cluster* yang baik. Jika nilai menuju negatif 1 (-1), maka menandakan bahwa titik antar *cluster* semakin jauh, dan mengindikasikan *cluster* tersebut kurang baik.

## 3. Results and Analysis

Kuisioner yang disebar memberikan hasil 400 responden mengisi kuisioner, namun hanya 388 yang terisi lengkap. Sehingga penelitian berlanjut dengan 388 kuisioner. Dari keseluruhan kuisioner diketahui beberapa karakteristik, seperti dijelaskan dalam gambar 1a dibawah, berdasarkan jenis kelamin ada 204 responden pria dan 184 responden wanita, dalam bentuk persentase terdapat 53% responden pria dan 47% responden wanita.




Gambar 1. Persentase jenis kelamin dan usia responden

Berdasarkan usia responden terbagi seperti pada gambar 1b, 18% atau 71 responden berusia antara 11-20 tahun, 40% atau 156 responden berusia antara 21-30 tahun, 21% atau 80 responden berusia antara 31-40 tahun, 13% atau 49 responden berusia antara 41-50 tahun, 6% atau 23 responden berusia antara 51-60 tahun, 2% atau 9 responden berusia antara >61 tahun.

Setelah proses *clustering* berlangsung, baik algoritma K-Means atau DBSCAN akan menghasilkan *cluster-cluster* atau kelompok. Kelompok-kelompok ini adalah kaitan antar variabel kepada moda transportasi yang dipilih, sehingga dapat mengetahui hasil dari proses *clustering* itu sendiri untuk menentukan apasaja penentu ketika seseorang memilih moda transportasi tersebut. Setelah proses *clustering* selesai, hasil *cluster* K-Means dan DBSCAN dapat dinilai efektifitasnya dan performanya menggunakan *Silhouette Coefficient. Silhouette Coefficient* akan mengukur rerata jarak antar titik dalam sebuah *cluster*. Semakin dekat maka hasilnya akan mendekati nilai 1, yang juga mendandakan proses *clustering* berhasil dengan baik karena *cluster* yang terbentuk cukup padat dan tidak melebar. Jika nilai *Silhouette Coefficient* mendekati -1, menandakan hasil *cluster* cukup renggang dan canderung melebar, ini menunjukkan hasil proses *clustering* yang kurang baik.

Tabel 1. Nilai *Silhouette Coefficient* K-Means dan DBSCAN

| *Silhouette Coefficient* | K-Means | DBSCAN |
|---|---|---|
| {1}&{2} | 0,53 | 0,91 |
| {1} | 0,86 | 0,98 |
| {3} | 0,79 | 0,98 |
| {4} | 0,89 | 1 |
| {5} | 0,97 | 1 |
| {6} | 0,95 | 1 |
| {7} | 0,87 | 1 |
| {8} | 0,95 | 1 |
| {9} | 0,95 | 1 |
| Rata-rata | 0,86 | 0,99 |

Pada tabel 1 diatas dapat terlihat nilai *Silhouette Coefficient* dari masing-masing algoritma, dapat dilihat bahwa algoritma DBSCAN memiliki performa yang lebih baik daripada algoritma K-Means. Hal ini terlihat dari nilai *Silhouette Coefficient* yang lebih besar dan mendekati 1. Rata-rata nilai *Silhouette Coefficient* DBSCAN juga lebih besar daripada K-Means, ini menandakan bahwa DBSCAN memiliki kemampuan yang baik dalam *clustering* data. Memang dalam penerapannya perlu beberapa kali percobaan untuk menemukan titik optimal dari *cluster* yang ada, seperti mengatur jumlah *cluster* optimal dari K-Means. Pengaturan titik optimal juga didapatkan dari nilai *Silhouette Coefficient* untuk setiap jumlah *cluster* yang terbentuk, mulai dari 2 dan dilakukan secara *incremental*. Pengaturan serupa juga dilakukan pada DBSCAN, nilai *eps* dan *minpts* juga mempengaruhi hasil *clustering*.

Faktor yang mempengaruhi minat penggunaan transportasi umum yang paling utama adalah yang terkait dengan biaya, kecepatan, kenyamanan, keamanan, efisiensi dan flekibilitas. Responden memilih moda transportasi bus karena biaya, efisien dan fleksibel. Responden memilih menggunakan kereta karena kecepatan, kenyamanan dan keamanan. Responden memilih pesawat karena kecepatan dan kenyamanan. Responden memilih menggunakan transportasi online karena efisien dan fleksibel. Dari hasil ini kita dapat menyimpulkan bahwa biaya, kecepatan, kenyamanan, keamanan, efisiensi dan flekibilitas memang menjadi faktor utama dalam pemilihan moda transportasi yang akan digunakan. Usia dan jenis kelamin serta pendapatan tidak mempengaruhi pemilihan moda transportasi tersebut.

p-ISSN: 1907-0012 e-ISSN : 2714-5417

## 4. Conclusion

Kesimpulan dari penelitian ini antara lain, DBSCAN memiliki performa yang baik dalam proses *clustering*, terutama dengan data yang cukup banyak dan berbasis kepada kepadatan data. Hal ini ditunjukkan dengan nilai *Silhouette Coefficient* yang lebih besar dan mendekati 1. Sementara faktor yang mempengaruhi minat penggunaan transportasi umum yang paling utama adalah yang terkait dengan biaya, kecepatan, kenyamanan, keamanan, efisiensi dan flekibilitas. Para penggiat transportasi umum seperti pengusaha, pengelola, dan agen dapat meningkatkan pelayanan dalam hal-hal tersebut.

## References

1. Pratiwi NMF, Indrayani L, Suwena KR. Persepsi Masyarakat terhadap Transportasi Publik Trans Sarbagita di Provinsi Bali. Ekuitas: Jurnal Pendidikan Ekonomi. 2020;8(1):80.
2. Winarno B, Manullang OR. Parameter Penentu Penggunaan Transportasi Umum Di Perkotaan Pati. Tataloka. 2018;20(1):75–86.
3. Prayudyanto MN. Perbandingan Kinerja Buy The Services Angkutan Umum Massal Kota Metropolitan dengan Metode Biaya Operasional Kendaraan dan Indeks Sustainabilitas. Jurnal Penelitian Transportasi Darat. 2021;23(1):55–71.
4. Kristianto A, Sediyono E, Hartomo KD. Implementation dbscan algorithm to clustering satellite surface temperature data in indonesia. Register: Jurnal Ilmiah Teknologi Sistem Informasi. 2020;6(2):109–18.
5. Dinata RK, Safwandi S, Hasdyna N, Azizah N. Analisis K-Means Clustering pada Data Sepeda Motor. INFORMAL: Informatics Journal. 2020;5(1):10–7.
6. Darmi YD, Setiawan A. Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk. Jurnal Media Infotama. 2017;12(2):148–57.
7. Budiman S, Safitri D, Ispriyanti D. Perbandingan Metode K-Means Dan Metode Dbscan Pada Pengelompokan Rumah Kost Mahasiswa Di Kelurahan Tembalang Semarang. JURNAL GAUSSIAN. 2016;5(4):757–62.