

Implementation of the Fuzzy C-Means Clustering Method in Grouping Provinces in Indonesia based on the Types of Goods Sold in E-commerce Businesses in 2022

Bimbim Oktaviandi, Tessy Octavia Mukhti*, Yenni Kurniawati, dan Zamahsary Martha

Departemen Statistika, Universitas Negeri Padang, Padang, Indonesia

*Corresponding author: tessyoctaviam@fmipa.unp.ac.id

Submitted : 01 Agustus 2024

Revised : 12 Agustus 2024

Accepted : 12 Agustus 2024

ABSTRACT

The internet facilitates e-commerce by enabling efficient transactions and building consumer trust. With internet users in Indonesia reaching 204 million in 2022, it is crucial to Cluster provinces based on the types of goods and services sold online to design effective marketing strategies. The Fuzzy C-Means (FCM) method is used for Cluster analysis, allowing objects to have different membership degrees in multiple Clusters and providing accurate Cluster center placement. This study applies Fuzzy C-Means to Cluster 34 provinces in Indonesia based on the sale of goods/services in e-commerce in 2022, aiming to provide insights into market preferences and assist companies in developing more effective strategies. The results show that the method forms two Clusters. By evaluating standard deviation values and ratios, Fuzzy C-Means proves effective in Clustering provinces in Indonesia based on e-commerce sales data. Cluster validation reveals a standard deviation ratio of 0.14, indicating clear and significant Cluster separation.

Keywords: Cluster analysis, E-commerce, Fuzzy C-Means



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Internet telah menjadi teknologi yang penting bagi banyak orang, organisasi, dan perusahaan, khususnya dalam e-commerce. *E-commerce* memungkinkan konsumen dan perusahaan untuk mengakses layanan dan melakukan transaksi dengan lebih efisien. Hal ini sangat membantu konsumen yang memiliki jadwal padat, karena mereka tidak perlu lagi mengunjungi toko fisik untuk membeli produk. Keberhasilan *e-commerce* tidak hanya bergantung pada pembuatan situs web toko online, tetapi juga pada kemampuan perusahaan untuk membangun kepercayaan konsumen. Kepercayaan ini penting karena banyak konsumen yang khawatir tentang kecurangan online. Evaluasi kredibilitas situs oleh pengunjung dimulai sejak mereka memasuki situs tersebut (Sunnyoto, 2013).

Pemasaran melalui internet menjadi aspek penting dalam bisnis, dengan *e-commerce* memotong biaya operasional dan memperluas jangkauan pasar (Supardi dan Dores, 2009). *E-commerce* di Indonesia berkembang pesat, didukung oleh banyaknya pengguna internet yang mencapai 204 juta pada tahun 2022 (Zulfikar, 2023). Ini mendorong banyak perusahaan untuk mengembangkan layanan online yang berkualitas dan tepercaya. Namun, *e-commerce* yang menawarkan banyak kemudahan belum tentu berhasil tanpa kepercayaan dari konsumen. Membangun kepercayaan adalah kunci untuk menarik konsumen agar mau berbagi informasi pribadi dan melakukan transaksi (Sunnyoto, 2013). Setiap provinsi di Indonesia memiliki karakteristik demografis, sosial, dan ekonomi yang berbeda, yang mempengaruhi preferensi konsumen terhadap jenis barang yang dijual secara online.

Terdapat tantangan dalam memahami dinamika pasar di berbagai provinsi di Indonesia. Setiap provinsi memiliki karakteristik demografis, sosial, dan ekonomi yang berbeda, yang mempengaruhi preferensi konsumen terhadap jenis barang yang dijual secara online. Oleh karena itu, pengelompokan provinsi berdasarkan jenis barang dan jasa yang dijual menjadi penting untuk membantu perusahaan *e-commerce* dalam merancang strategi pemasaran, distribusi, dan pengembangan produk yang lebih efektif dan efisien. Dengan memahami pola penjualan di setiap provinsi, perusahaan *e-commerce* dapat menargetkan upaya mereka secara lebih tepat, meningkatkan kepuasan pelanggan, dan mengoptimalkan operasi logistik mereka.

Pengelompokan provinsi berdasarkan jenis barang dan jasa yang dijual sangat penting untuk membantu perusahaan *e-commerce* dalam merancang strategi pemasaran, distribusi, dan pengembangan produk yang lebih efektif

dan efisien. Salah satu analisis yang bisa digunakan dalam pengelompokan adalah analisis *cluster*. Analisis *cluster* adalah proses pengelompokan objek berdasarkan kemiripan untuk mengetahui kemiripan di antara mereka (Han dkk., 2011). Analisis *cluster* secara umum terbagi menjadi dua yaitu hirarki dan non hirarki. Metode hirarki digunakan untuk mengelompokkan objek secara terstruktur berdasarkan kemiripan sifat dan jumlah *cluster* yang diharapkan belum diketahui banyaknya. Sedangkan metode non hirarki menurut Johnson dan Wichern (2002) mengelompokkan objek kedalam suatu kelompok dimana jumlah kelompok ditentukan terlebih dahulu. Analisis *Fuzzy C-Means* merupakan analisis pengelompokan non-hirarki yang menggunakan model *fuzzy*, memungkinkan objek untuk menjadi anggota lebih dari satu *cluster* dengan derajat keanggotaan yang berbeda (Kusumadewi dkk., 2006). Penelitian yang dilakukan oleh Suciati (2021) ini membahas tentang analisis *cluster* dengan memakai analisis *Fuzzy C-Means*. Analisis tersebut menggunakan ukuran jarak *Euclidean* dan memiliki keunggulan dalam menempatkan pusat *cluster* lebih akurat dibanding analisis lainnya. *Fuzzy C-Means* bermanfaat dalam pengelompokan data berdasarkan lebih dari satu variabel sekaligus dan telah banyak diterapkan dalam berbagai penelitian. Penelitian lainnya dari Dauni dkk (2023) disimpulkan bahwa algoritma *Fuzzy C-Mean* memiliki potensi yang signifikan sebagai pendekatan yang efektif dalam melaksanakan proses *clustering* dan *Fuzzy C-Mean* digunakan sebagai alat yang dapat membantu dalam mengelompokkan data untuk memahami pola dan karakteristik serta *Fuzzy C-Mean* dapat menjadi pendekatan yang efisien dan efektif dalam clustering n.

Penelitian ini bertujuan mengimplementasikan analisis *Fuzzy C-Means* untuk pengelompokan provinsi di Indonesia berdasarkan jenis barang dan jasa yang dijual pada *e-commerce* tahun 2022. Hasil penelitian ini diharapkan memberikan wawasan tentang preferensi pasar di berbagai provinsi dan membantu perusahaan *e-commerce* dalam menyusun strategi yang lebih tepat sasaran.

II. METODE PENELITIAN

Penelitian ini data yang digunakan dari publikasi Badan Pusat Statistik (BPS) melalui survei *e-commerce* mengenai jenis barang/jasa yang dijual pada usaha *e-commerce* di Indonesia menurut 34 provinsi tahun 2022, dari data tersebut akan dilakukan pengelompokan provinsi di Indonesia menurut jenis barang dan jasa yang dijual.

Tabel 1. Variabel Penelitian

Variabel	Keterangan	Satuan
X_1	Alat tulis kantor	%
X_2	Fashion	%
X_3	Gadget dan aksesoris	%
X_4	Hobi dan Olahraga	%
X_5	Kecantikan dan perawatan	%
X_6	Makanan dan minuman	%
X_7	Obat-obatan	%
X_8	Barang-barang elektronik	%
X_9	Kebutuhan rumah tangga	%
X_{10}	Alat transportasi	%
X_{11}	Bahan bangunan dan perlengkapan kelistrikan	%
X_{12}	Jasa transportasi	%
X_{13}	Jasa akomodasi	%
X_{14}	Jasa profesi	%

Analisis dilakukan dengan memanfaatkan *software Rstudio*. Tahap pengelompokan dilakukan dengan menggunakan metode analisis *Fuzzy C-Means* dijalankan sesuai dengan langkah berikut.

1. Menentukan *Cluster* Optimal

Metode *Elbow* digunakan untuk menentukan jumlah *cluster* optimal dengan mengamati penurunan paling besar antara jumlah *cluster* yang diuji. Jika terdapat siku pada grafik antara nilai *cluster* pertama dan kedua, jumlah *cluster* tersebut dianggap yang terbaik untuk analisis data.

$$WSS = \sum_{i=1}^g \sum_{x_{ij} \in V_{kj}} \|x_{ij} - v_{kj}\|^2 \quad (1)$$

Keterangan:

- g : Jumlah *cluster*
- x_{ij} : Objek ke- i pada variabel ke- j
- v_{kj} : Merupakan pusat *cluster* objek ke- k pada variabel ke- j
- $x_{ij} \in V_{kj}$: Nilai keanggotaan titik data x_{ij} ke pusat *cluster* V_{kj}
- g : Jumlah *cluster*

2. Algoritma Fuzzy C-Means

Menurut Everitt dkk, (2011), pada Fuzzy cluster untuk objek memiliki tingkat keanggotaan yang menunjukkan sejauh mana objek tersebut terkait dengan *cluster*, dengan nilai antara nol dan satu, yang memungkinkan interpretasi probabilitas. Berikut adalah langkah Fuzzy C-Means.

- a. Tentukan jumlah *cluster* (g), nilai pembobot keanggotaan w , dan tingkat akurasi ε .
- b. Membentuk matriks awal $\mathbf{U}^{(0)}$ secara acak, yang terdiri dari nilai derajat keanggotaan ditunjukkan pada persamaan berikut.

$$\mu_{ik} = \mathbf{U}_{g \times n} = \begin{bmatrix} \mu_{11}(x_1) & \mu_{21}(x_2) & \dots & \mu_{1n}(x_n) \\ \mu_{21}(x_1) & \mu_{22}(x_2) & \dots & \mu_{2n}(x_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mu_{g1}(x_1) & \mu_{g2}(x_2) & \dots & \mu_{gn}(x_n) \end{bmatrix} \quad (2)$$

Keterangan:

- $\mathbf{U}_{g \times n}^r$: Matriks partisi berukuran $g \times n$ pada iterasi ke- r
- g : Ukuran *cluster* yang digunakan
- n : Banyaknya Objek
- μ_{ik} : adalah derajat keanggotaan objek ke- k pada *cluster* ke- i

- c. Menghitung pusat *cluster* ke- i pada variabel ke- j menggunakan persamaan berikut.

$$v_{ij} = \frac{\sum_{k=1}^n (\mu_{ik})^w \cdot x_{kj}}{\sum_{k=1}^n (\mu_{ik})^w} \quad (3)$$

Keterangan:

- v_{ij} = Pusat *cluster* ke- i variabel ke- j
- μ_{ik} = Derajat keanggotaan objek ke- k pada *cluster* ke- i
- r = Iterasi
- w = Parameter Pembobot keanggotaan

- d. Menghitung jarak setiap objek ke titik pusat (d_{ik}).
- e. Lakukan pembaharuan nilai matriks partisi \mathbf{U}^t menggunakan persamaan berikut.

$$\mu_{ik} = \left[\sum_{l=1}^g \left(\frac{d_{il}}{d_{lk}} \right)^{\frac{2}{w-1}} \right]^{-1} \quad (4)$$

d_{ik} dengan merupakan jarak objek ke- k pada pusat *cluster* ke- i .

- f. Pengecekan kondisi berhenti.
Jika $(|\mathbf{U}^t - \mathbf{U}^{t-1}| < \varepsilon)$, maka itersi selesai.
Namun Jika tidak : maka $t = t + 1$ dapat diulangi pada langkah ke-3.
Dengan ε_L merupakan tingkat akurasi yang ditentukan.

3. Pengecekan Validasi Cluster

Cluster optimal ditandai dengan varians besar antar kelompok dan varians kecil dalam kelompok (Johnson dan Wichern, 2002). Menurut Bunkerrs dan James (1996), untuk mencari nilai rasio simpangan baku dirumuskan dengan persamaan berikut.

$$S_w = \frac{1}{g} \sum_{k=1}^g S_k \quad (5)$$

$$\text{dengan } S_k = \frac{1}{n_k - 1} \sum_{i=1}^{n_k} (x_{ik} - \bar{x}_k) \quad (6)$$

dimana $k = 1, 2, \dots, g$ dan

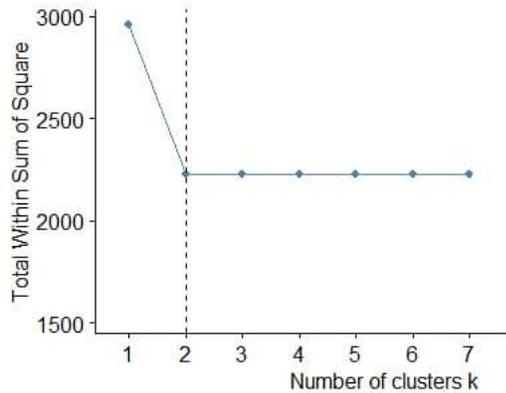
$$S_B = \sqrt{\frac{1}{g-1} \sum_{k=1}^g (\bar{x}_k - \bar{x})^2} \tag{7}$$

Keterangan:

- S_w = Simpangan Baku dalam kelompok
- S_B = Simpangan Baku antar kelompok
- \bar{x}_k = Nilai rata-rata pengamatan di dalam kelompok ke- k
- \bar{x} = Nilai rata-rata seluruh pengamatan
- n_k = Banyaknya pengamatan di dalam kelompok ke- k
- g = Banyaknya kelompok

III. HASIL DAN PEMBAHASAN

Jumlah *cluster* optimal dalam analisis ini ditentukan menggunakan metode *elbow*, yang divisualisasikan dalam Gambar 1 dengan sumbu x menunjukkan jumlah *cluster* dan sumbu y menunjukkan nilai *WSS*. Titik optimal diidentifikasi dari penurunan tajam yang membentuk siku. Hasil analisis menunjukkan terbentuk 2 *cluster* optimal, seperti ditunjukkan Gambar 1.



Gambar 1. Kurva Metode *Elbow*

Setelah diketahui jumlah *cluster* optimal, dilakukan analisis *Fuzzy C-Means* mengidentifikasi 15 provinsi di *cluster* 1 dan 19 provinsi di *cluster* 2, menunjukkan bahwa mayoritas provinsi masuk ke *cluster* 2. Karakteristik setiap *cluster* dapat ditemukan di Tabel 2.

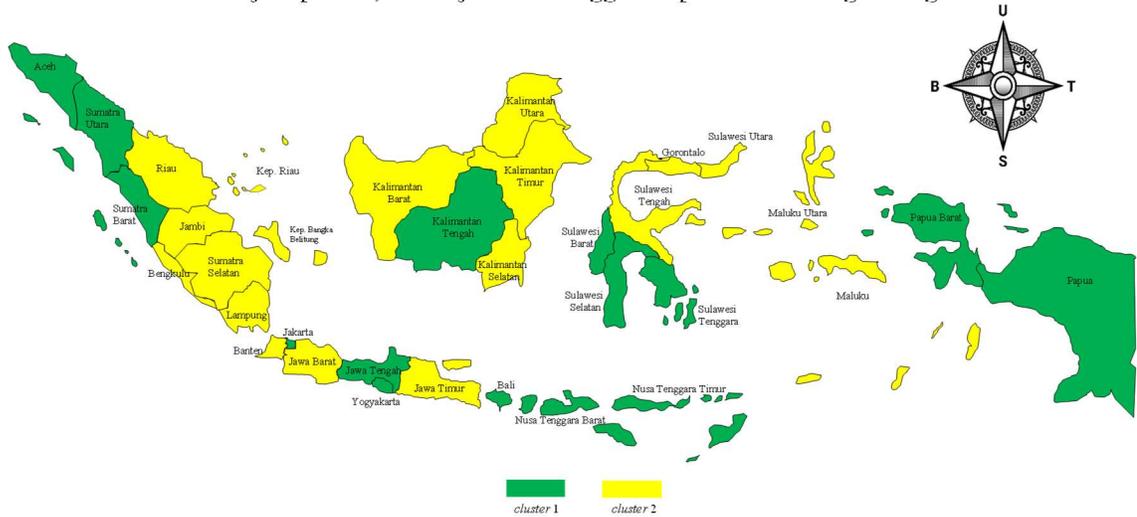
Tabel 2. Karakteristik Untuk Tiap *Cluster*

Cluster	Variabel														Kategori
	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	X ₈	X ₉	X ₁₀	X ₁₁	X ₁₂	X ₁₃	X ₁₄	
1	2.95	16.57	7.62	4.20	7.11	40.41	3.52	1.82	8.69	4.09	1.91	4.78	2.67	3.23	Tinggi
2	2.55	15.36	6.15	3.97	6.52	46.12	3.42	1.84	8.04	4.11	1.91	3.97	1.82	3.36	Rendah

Hasil yang disajikan dalam Tabel 2 menunjukkan karakteristik dari dua *cluster* berdasarkan rata-rata indikator untuk masing-masing *cluster*. Pada *cluster* dengan kategori rendah memiliki rata-rata nilai terendah hampir pada setiap indikator yang diukur. Ini menunjukkan bahwa usaha *e-commerce* dalam *cluster* ini cenderung memiliki performa atau aktivitas yang lebih rendah dalam hal penjualan atau keterlibatan dalam berbagai jenis Barang/Jasa yang dijual. Sebaliknya pada *cluster* dengan kategori tinggi memiliki rata-rata nilai tertinggi pada beberapa indikator penting. Ini menunjukkan bahwa usaha *e-commerce* dalam *cluster* ini cenderung memiliki performa atau aktivitas yang lebih tinggi dalam penjualan berbagai jenis Barang/Jasa.

Dari hasil yang diperoleh, dapat dilihat bahwa *cluster* 1 lebih unggul dalam sebagian besar variabel. Usaha *e-commerce* seperti alat tulis kantor, fashion, gadget dan aksesoris, hobi dan olahraga, kecantikan dan perawatan, obat-obatan, kebutuhan rumah tangga, jasa transportasi, dan jasa akomodasi menunjukkan nilai yang lebih tinggi di *cluster*

1 dibandingkan dengan *cluster* 2. Hal ini mengindikasikan bahwa sektor-sektor tersebut lebih berkembang atau lebih dominan di *cluster* 1. Sebaliknya, *cluster* 2 menunjukkan keunggulan pada variabel makanan dan minuman, barang-barang elektronik, alat transportasi, dan jasa profesi. Ini menandakan bahwa meskipun secara keseluruhan *cluster* 2 memiliki nilai yang lebih rendah, sektor-sektor ini lebih berkembang di *cluster* 2 dibandingkan dengan *cluster* 1. Secara keseluruhan, *cluster* 1 dapat diidentifikasi sebagai cluster dengan usaha *e-commerce* yang lebih kuat di berbagai bidang, terutama di sektor-sektor seperti fashion, gadget, dan jasa transportasi. Sementara itu, *cluster* 2 lebih fokus pada sektor makanan dan minuman serta jasa profesi, menunjukkan keunggulan spesifik di bidang-bidang tersebut.



Gambar 2. Pemetaan *Cluster* untuk Metode *Fuzzy C-Means*

Gambar 2 menunjukkan hasil analisis dengan menggunakan *Fuzzy C-Means* berdasarkan pada data penjualan barang/jasa pada usaha *e-commerce* di Indonesia. Sebagian besar provinsi di Indonesia memiliki indikator penjualan barang/jasa pada rendah. Mayoritas berada di provinsi dekat ibu kota, seperti wilayah yang berada di sekitaran Sumatera, Kalimantan dan Sulawesi cenderung memiliki akses teknologi dan infrastruktur yang kurang memadai, dengan tantangan geografis yang lebih besar, sehingga aktivitas *e-commerce* lebih rendah. Sementara itu, indikator penjualan barang/jasa pada tingkat tinggi, terdapat pada provinsi yang jauh dari ibu kota, didominasi oleh wilayah beberapa provinsi seperti Aceh, Kalimantan Tengah, Bali-Nusa, beberapa provinsi yang ada di pulau Sulawesi dan Papua cenderung memiliki infrastruktur yang lebih baik, pusat ekonomi dan pemerintahan yang kuat, serta akses teknologi yang lebih maju.

Untuk melihat validasi dari *Fuzzy C-Means*, perlu dilakukan uji validitas *cluster* menggunakan nilai rasio simpangan baku dengan hasil pada Tabel 3.

Tabel 3. Hasil Validasi *Cluster*

Nilai	Fuzzy C-Means
Sw	324.69
Sb	2200.45
Sw/Sb	0.14

Pada Tabel 3, hasil validasi *Cluster Fuzzy C-Means* menunjukkan bahwa algoritma ini efektif dalam membentuk *cluster* yang terpisah dengan baik. Dengan nilai simpangan baku dalam kelompok (Sw) sebesar 324.69 dan simpangan baku antar kelompok (Sb) sebesar 2200.45, dan rasio simpangan baku diperoleh sebesar 0.14. Rasio yang kecil ini menandakan bahwa jarak antar-*cluster* relatif besar dibandingkan dengan variasi di dalam *cluster* itu sendiri. Artinya, *cluster* yang terbentuk cukup jelas dan terpisah satu sama lain, sementara variasi di dalam setiap *cluster* relatif kecil. Dengan demikian, hasil ini mengindikasikan bahwa penerapan metode *Fuzzy C-Means* pada data penjualan barang dan jasa di *e-commerce* Indonesia berhasil menghasilkan pembagian *cluster* yang jelas dan berbeda secara signifikan.

IV. KESIMPULAN

Kesimpulan dari analisis ini adalah bahwa metode *Fuzzy C-Means* efektif dalam mengelompokkan provinsi di Indonesia berdasarkan data penjualan barang dan jasa di *e-commerce*. Pengelompokan menghasilkan 2 *cluster*, dimana *cluster* 1 dengan kategori tinggi yang terdiri dari provinsi dengan performa penjualan yang lebih tinggi, dan *cluster* 2 dengan kategori rendah memiliki performa penjualan yang lebih rendah. Validasi *cluster* menunjukkan rasio simpangan baku sebesar 0.14, mengindikasikan pemisahan *cluster* yang jelas dan signifikan. Mayoritas provinsi dengan indikator penjualan rendah terletak di wilayah dengan akses teknologi dan infrastruktur yang kurang memadai, sedangkan provinsi dengan indikator penjualan tinggi umumnya berada di wilayah dengan infrastruktur lebih baik dan akses teknologi yang lebih maju. Untuk penelitian selanjutnya disarankan untuk mengkaji lebih dalam faktor-faktor spesifik yang mempengaruhi performa penjualan *e-commerce* di setiap provinsi. Hal ini penting untuk mengidentifikasi apa yang membuat beberapa provinsi lebih berhasil dalam *e-commerce* dibandingkan dengan yang lain. Faktor-faktor ini bisa mencakup aspek seperti infrastruktur digital, akses internet, tingkat pendidikan, pendapatan per kapita, kebijakan pemerintah, serta preferensi konsumen lokal. Dengan informasi ini, perusahaan dan pembuat kebijakan dapat merumuskan strategi yang lebih efektif.

DAFTAR PUSTAKA

- Badan Pusat Statistika RI. (2022). *Statistik E-Commerce 2022*.
- Everitt, B. S., Landau, S., Leese, M., & Stahl, D. (2011). Cluster analysis. In Isberg (Ed.), *John Wiley & Sons* (5th ed.). John Wiley & Sons,. <https://doi.org/10.1007/BF00154794>
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: concepts and techniques*. Elsevier.
- Johnson, R. A., & Wichern, D. W. (2002). *Applied Multivariate Statistical Analysis (5th Ed)* (5th ed.). Prentice Hall.
- Kusumadewi, S., Hartati, S., Harjoko, S., & Wrdoyo, R. (2006). *Fuzzy Multi-attribute Decision Making*. Graha Ilmu.
- P, D., Pratiwi, & R, T. P. (2023). Pemetaan Keberlangsungan Hidup Umkm Guna Optimalisasi Bantuan Kredit Menggunakan Algoritma Fuzzy C-Means. *Jurnal Responsif*, 5 No1, 61–69. <https://ejurnal.ars.ac.id/index.php/jti>
- Suciati, I. (2021). *Analisis Cluster Menggunakan Metode Fuzzy C-Means pada Data COVID-19 di Provinsi Lampung*.
- Sunyoto. (2013). *Dasar-Dasar Manajemen Pemasaran*. CAPS Center for Academic Publishing Service.
- Supardi, J., & Does, V. (2009). Rancang Bangun Collaborative System Pemasaran Hotel Secara On-Line Dengan Pendekatan Mediator Based. *Jurnal Sistem Informasi (JSI)*, 1 NO.2(pp.55-61).
- Zulfikar, F. (2023). 10 Negara dengan Pengguna Internet Tertinggi di Dunia, Indonesia Nomor Berapa??. *DetikEdu*.