

**PENERAPAN METODE *TWO-STEP CLUSTERING* UNTUK ANALISIS
PREFERENSI JENIS *MARKETPLACE* MAHASISWA S1 FMIPA
UNIVERSITAS JENDERAL SOEDIRMAN**

Helmi Farid Ashari

Jurusan Matematika, Universitas Jenderal Soedirman
Helmi.ashari@mhs.unsoed.ac.id

Jajang*

Jurusan Matematika, Universitas Jenderal Soedirman
Jajang@unsoed.ac.id

Agustini Tripena Br. Surbakti

Jurusan Matematika, Universitas Jenderal Soedirman
agustini.tripena@unsoed.ac.id

ABSTRACT. *The Corona Virus Disease 2019 has caused a transformation in societal life, including the phenomenon of online shopping. Cluster Analysis is a multivariate technique used to group objects based on the similarity measures. The applied method for mixed data between continuous data and categorical data is the two-step cluster method. This study applies analysis to the characteristics of college students in the Faculty of Mathematics and Natural Sciences at Jenderal Soedirman University based on eight variables: gender, graduation year, e-commerce platform, payment method, product type, monthly online shopping expenses, and monthly allowance. The two-step cluster method comprises an initial clustering stage using a sequential method and a final stage using a log-likelihood distance-based agglomeration method. The determination of the optimal number of clusters is based on the Bayesian Information Criterion (BIC). The analysis results indicate the formation of three optimal clusters out of 100 students: cluster 1 comprising 25 male and 4 female students choosing Shopee, Tokopedia, cluster 2 consisting of 11 male and 35 female students choosing Shopee and Tokopedia, and Lazada, and cluster 3 including 14 male and 11 female students choosing Shopee, Gojek, and Grab.*

Keywords: *Cluster Analysis, Student online shop, Log-likelihood Distance, Two-Step Cluster.*

ABSTRAK. *Corona Virus Disease 2019 memicu transformasi dalam kehidupan masyarakat, salah satunya fenomena berbelanja online. Analisis cluster adalah salah satu teknik multivariat yang digunakan untuk mengelompokkan objek-objek menjadi beberapa cluster berdasarkan ukuran kemiripan. Metode yang dapat diterapkan untuk data campuran antara data kontinu dan data kategorik adalah metode two-step cluster. Penelitian ini akan menerapkan analisis pada karakteristik mahasiswa S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jenderal Soedirman berdasarkan delapan variabel yaitu: jenis kelamin, tahun angkatan, jurusan, platform e-commerce, jenis pembayaran, jenis produk, biaya belanja online per bulan, dan uang saku per bulan. Metode two-step cluster memiliki dua tahapan yaitu tahap clustering awal dengan metode*

*Penulis Korespondensi

sekuensial dan tahap akhir dengan metode pemusataan berdasarkan ukuran jarak *log-likelihood*. Penentuan banyaknya *cluster* optimal didasarkan pada *Bayesian Information Criterion* (BIC). Hasil analisis menunjukkan terbentuknya tiga *cluster* optimal dari 100 mahasiswa. *Cluster* 1 yang terdiri dari 25 mahasiswa laki-laki dan 4 mahasiswa perempuan dengan pilihan *platform* Shopee, Tokopedia, dan Lazada. *Cluster* 2 yang terdiri dari 11 mahasiswa laki-laki dan 35 mahasiswa perempuan dengan pilihan *platform* Shopee dan Tokopedia. *Cluster* 3 yang terdiri dari 14 mahasiswa laki-laki dan 11 mahasiswa perempuan dengan pilihan *platform* Shopee, Gojek, dan Grab.

Kata kunci: Analisis *Cluster*, Belanja *Online*, Jarak *Log-Likelihood*, *Two-Step Cluster*

1. PENDAHULUAN

Corona Virus Disease 2019 memicu transformasi dalam kehidupan masyarakat. Adanya aturan Pembatasan Sosial Berskala Besar yang dikenal dengan PSBB mendorong perubahan pola dalam berkegiatan sosial yang mewajibkan untuk saling menjaga jarak. Hal ini juga berdampak pada kegiatan bisnis jual beli yang sebelumnya dipenuhi secara *offline* menjadi dipenuhi secara *online*. Merupakan sebuah hal yang menarik untuk dikaji preferensi konsumen dalam memilih jenis *marketplace* yang menarik ketika mereka berbelanja secara *online*. Tentunya banyak hal atau banyak variabel yang terkait dengan preferensi seseorang dalam memilih jenis *marketplace*. Oleh karena itu, analisis yang relevan untuk digunakan adalah analisis multivariat. Analisis multivariat adalah suatu analisis dengan melibatkan lebih dari satu variabel. Salah satu penerapan dari analisis multivariat adalah analisis *cluster*. Menurut Jhonson dan Winchern (2007), analisis *cluster* adalah analisis yang digunakan untuk mengelompokkan objek-objek menjadi beberapa *cluster* berdasarkan ukuran kemiripan.

Analisis *cluster* yang umum diterapkan untuk data dengan variabel kontinu dengan kedekatan antar dua objek adalah dengan menggunakan ukuran jarak atau ukuran *dissimilarity*. Namun demikian, akan terjadi masalah apabila akan mengukur dua objek yang mempunyai variabel kategori. Oleh karena itu, perlu metode lain untuk menentukan kemiripan antar dua objek tersebut. Menurut Mongi (2015), metode *two-step cluster* merupakan metode yang didesain untuk menangani jumlah objek besar, terutama pada masalah objek yang mempunyai variabel campuran, kontinu dan kategorik menggunakan ukuran jarak *Log-Likelihood*, serta dilengkapi dengan penentuan banyaknya *cluster* optimal

berdasarkan nilai *Bayesian information criterion* (BIC). Dua tahapan dalam metode *two-step cluster* adalah *clustering* tahap awal dan akhir. *Clustering* tahap awal menggunakan metode sekuensial dan *clustering* tahap akhir merupakan metode pemusatan (*agglomerative*).

Beberapa penelitian analisis *cluster* telah dilakukan oleh beberapa peneliti, antara lain Rundle dkk (2015) yang mengkaji analisis *cluster* dua tahap, yaitu identifikasi segmen dalam konteks aktivitas fisik dan tahap kedua adalah menawarkan implikasi praktis bagi pemasar sosial yang bekerja di bidang aktivitas fisik.

2. METODE PENELITIAN

Penelitian ini akan menerapkan analisis metode *two-step cluster* untuk melakukan analisis terhadap *cluster* yang terbentuk pada minat mahasiswa S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jenderal Soedirman berbelanja *online* yang ditinjau dari 8 variabel yaitu: jenis kelamin, tahun angkatan, jurusan, *platform e-commerce*, jenis pembayaran, jenis produk, biaya belanja *online* per bulan, dan uang saku per bulan.

2.1 Metode Two-Step Cluster

Metode analisis *two-step cluster* adalah metode analisis yang digunakan untuk melakukan pengelompokan objek dengan dua tahap pengelompokan (Norusis, 2004; Tkaczynski, 2017). Metode *two-step cluster* merupakan metode yang didesain untuk menangani jumlah objek besar, terutama pada masalah objek yang mempunyai variabel campuran kontinu dan kategorik menggunakan ukuran jarak *Log-Likelihood*, serta dilengkapi dengan penentuan banyaknya *cluster* optimal berdasarkan nilai *Bayesian information Criterion* (BIC) (Mongi, 2015).

2.2 Ukuran Jarak *Log-Likelihood*

Asumsi yang melandasi jarak *Log-Likelihood* adalah data kontinu berdistribusi normal, data kategorik berdistribusi multinomial, dan variabel kontinu dan variabel kategorik saling bebas. Menurut Schiopu (2010), jarak *Log-Likelihood* antara *cluster* ke-*i* dan *cluster* ke-*j* didefinisikan sebagai :

$$\begin{aligned}
d(j, s) &= \xi_j + \xi_s - \xi_{(j,s)} \\
d(j, s) &= \left[-N \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) - \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{jkl}}{N_j} \log \left(\frac{N_{jkl}}{N_j} \right) \right) \right] \\
&+ \left[-N \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{sk}^2) - \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{skl}}{N_s} \log \left(\frac{N_{skl}}{N_s} \right) \right) \right] \\
&- \left[-N \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{(j,s)k}^2) - \sum_{k=1}^{K^B} \sum_{l=1}^{L_k} \frac{N_{(j,s)kl}}{N_{(j,s)}} \log \left(\frac{N_{(j,s)kl}}{N_{(j,s)}} \right) \right) \right].
\end{aligned}$$

dengan $d(i, j)$ adalah jarak antara *cluster* ke- i atau *cluster* ke- j , ξ_i adalah ukuran keragaman *cluster* ke- i , ξ_j adalah ukuran keragaman *cluster* ke- j , $\xi_{i,j}$ adalah ukuran keragaman gabungan *cluster* ke- i dan *cluster* ke- j , K^A adalah banyaknya variabel kontinu, K^B adalah banyaknya variabel kategorik, L_k adalah banyaknya kategori pada variabel kategorik ke- k , N adalah banyaknya objek pada *cluster*, N_j adalah banyaknya objek *cluster* ke- j , N_s adalah banyaknya objek *cluster* ke- s , N_{jkl} adalah banyaknya objek *cluster* ke- i , *cluster* ke- j atau gabungan *cluster* ke- i dan *cluster* ke- j pada variabel kategorik ke- k yang diambil dari kategori ke- l , $\hat{\sigma}_k^2$ adalah estimasi variansi variabel kontinu ke- k , $\hat{\sigma}_{sk}^2$ adalah estimasi variansi *cluster* ke- i , *cluster* ke- j atau gabungan *cluster* ke- i dan *cluster* ke- j pada variabel kontinu ke- k

2.3 Tahap Pre-clustering

Pre-clustering dilakukan dengan menerapkan metode sekuensial, dimana setiap objek harus bergabung dengan anak *cluster* lama atau membentuk anak *cluster* baru berdasarkan ukuran jarak. Metode pengelompokan sekuensial dengan membangun *cluster feature* (CF) menurut Chiu dkk, (2001) :

- mendeklarasikan obyek pertama sebagai *subcluster* pertama yang dibentuk;
- mencari jarak Log-Likelihood antara *subcluster* pertama dengan obyek kedua;

- c. apabila jarak *Log-Likelihood* antara *subcluster* pertama dengan obyek kedua lebih besar dari *threshold* (θ) maka dibentuk suatu *subcluster* baru. *Subcluster* baru tersebut merupakan obyek kedua itu sendiri;
- d. sebaliknya, apabila syarat pada bagian (c) tidak terpenuhi, maka terjadi penggabungan obyek kedua dengan *subcluster* pertama.

2.4 Tahap *Hierrarchical Cluster of Subcluster*

Hierrarchical Cluster of Subcluster menerapkan metode pemusatan dilakukan untuk membentuk kumpulan *cluster* yang terbentuk pada tahap *pre-clustering*. Metode pemusatan (*agglomerative*) pada pengelompokan *subcluster-subcluster* adalah (Chiu dkk, 2001):

- a. membuat matriks jarak antar *subcluster* menggunakan jarak *Log-Likelihood*;
- b. mencari nilai jarak antara dua *subcluster* berdasarkan ukuran jarak *Log-Likelihood* terbesar. Misalkan *subcluster* **U** dan **V** memiliki ukuran jarak *Log-Likelihood* terbesar maka digabungkan ke dalam kelompok dengan label yang baru yaitu kelompok **UV**;
- c. mencari jarak antara kelompok **UV** dengan beberapa *subcluster* lain dengan ukuran jarak *Log-Likelihood*;
- d. menyusun kembali matriks jarak antara *subcluster* yang baru.

2.5 Penentuan *Cluster Optimal*

Penentuan banyaknya *cluster* optimal didasarkan pada BIC yang didefinisikan sebagai :

$$BIC(J) = -2 \sum_{j=1}^J \xi_j + m_j \log N$$

$$BIC(J) = -2 \sum_{j=1}^J -N_j \left(\sum_{k=1}^{K^A} \frac{1}{2} \log(\hat{\sigma}_k^2 + \hat{\sigma}_{jk}^2) + \sum_{g=1}^{K^B} - \left(\sum_{l=1}^{L_k} \frac{N_{jgl}}{N_j} \log \frac{N_{jgl}}{N_j} \right) \right)$$

$$+ j \left(2K^A + \sum_{k=1}^{K^B} L_k - 1 \right) \log(N)$$

dengan m_j adalah ukuran jarak pada *cluster-j* yang terbentuk dari tahap *Hierarchical Cluster*, N_{jgl} adalah banyaknya objek *cluster* ke- j pada variabel kategorik ke- g yang diambil kategori ke- l

Banyaknya *cluster* optimum dapat dilihat dari nilai BIC yang paling kecil. Beberapa kasus dalam pengklusteran, nilai BIC akan bertambah besar dikarenakan jumlah *cluster* semakin banyak. Maka dalam situasi tersebut, pilih *ratio of BIC changes* terbesar atau dapat memilih nilai *ratio of distance measures* yang terbesar dibandingkan dengan nilai batas $c2 = 1,15$.

2.6 Prosedur Analisis

Langkah-langkah yang digunakan pada penelitian ini adalah sebagai berikut:

1. menentukan variabel penelitian;
2. memisahkan data kategorik dan data kontinu;
3. melakukan *preclustering* atau tahap pengelompokan awal untuk membentuk *subcluster-subcluster* dari data kontinu dan data katgorik dengan menggunakan jarak *Log-Likelihood*;
4. melakukan *Hierrarchical Cluster of Subcluster* dengan metode pemusatan yang berupa hasil-hasil kumpulan *cluster* pada tahap *pre-clustering*;
5. setelah *cluster* terbentuk, dilakukan pemilihan banyaknya *cluster-cluster* optimal dengan metode BIC;
6. Interpretasi pada *cluster* optimal yang terbentuk sesuai anggota *cluster* tersebut.

3. HASIL DAN PEMBAHASAN

3.1 Pre-Clustering

Tahapan awal dalam penelitian *clustering* dua tahap ini adalah menentukan banyaknya *subcluster* yang terbentuk. Dari hasil output setiap *subcluster* yang terbentuk dan banyaknya objek dalam setiap *subcluster* yang terbentuk disajikan pada Tabel 1.

Tabel 1. Tahap Pre-Clustering

| <i>Cluster</i> | N | <i>Cluster</i> | N |
|----------------|---|----------------|---|
| 1 | 4 | 16 | 4 |
| 2 | 5 | 17 | 4 |
| 3 | 3 | 18 | 4 |
| 4 | 2 | 19 | 3 |
| 5 | 3 | 20 | 5 |
| 6 | 3 | 21 | 2 |
| 7 | 4 | 22 | 3 |
| 8 | 2 | 23 | 2 |
| 9 | 3 | 24 | 2 |
| 10 | 5 | 25 | 3 |
| 11 | 5 | 26 | 4 |
| 12 | 3 | 27 | 2 |
| 13 | 3 | 28 | 3 |
| 14 | 4 | 29 | 3 |
| 15 | 5 | 30 | 2 |

Tabel 1 menunjukkan bahwa pada tahap *pre-clustering* ini diperoleh 30 *subcluster* yang terbentuk, setiap objek harus bergabung dengan anak *cluster* lama atau membentuk anak *cluster* baru berdasarkan ukuran jarak.

3.2 *Hierrarchical Cluster of Subcluster dan Penentuan Cluster Optimal*

Tabel 2. Nilai BIC (*Bayesian Information Criterion*)

| Banyaknya <i>Cluster</i> | <i>Schwarz's</i> | Perubahan Nilai BIC | Rasio | Rasio |
|-----------------------------|--|------------------------|------------------------|-----------------|
| | <i>Bayesian</i> <i>Criterion</i> (BIC) | | Perubahan Nilai BIC | Ukuran Jarak |
| 1 | 1466,898 | | | |
| 2 | 1411,559 | -55,339 | 1 | 1,152 |
| 3 | 1376,902 | -34,657 | 0,626 | 1,358 |
| 4 | 1378,094 | 1,192 | -0,022 | 1,255 |

| | | | | |
|----|----------|--------|--------|-------|
| 5 | 1399,65 | 21,557 | -0,39 | 1,394 |
| 6 | 1443,741 | 44,091 | -0,797 | 1,024 |
| 7 | 1489,152 | 45,411 | -0,821 | 1,146 |
| 8 | 1541,703 | 52,551 | -0,95 | 1,124 |
| 9 | 1599,65 | 57,947 | -1,047 | 1,109 |
| 10 | 1661,858 | 62,208 | -1,124 | 1,314 |
| 11 | 1733,403 | 71,545 | -1,293 | 1,004 |
| 12 | 1805,064 | 71,661 | -1,295 | 1,139 |
| 13 | 1880,355 | 75,291 | -1,361 | 1,009 |
| 14 | 1955,872 | 75,516 | -1,365 | 1,081 |
| 15 | 2033,322 | 77,45 | -1,4 | 1,005 |

Tabel 2 menunjukkan bahwa ada dua metode evaluasi yang memberikan informasi bertentangan. Nilai BIC memberikan nilai yang lebih rendah untuk banyaknya *cluster* 3 (1376,902) dibandingkan dengan banyaknya *cluster* 5 (1399,650), yang berarti jumlah *cluster* 3 cenderung lebih baik. Namun, rasio ukuran jarak antar *cluster* menunjukkan bahwa jumlah *cluster* 5 memiliki rasio ukuran jarak yang lebih besar (1,394) dibandingkan banyaknya *cluster* 3 (1,358). Dari simulasi yang dilakukan (Bacher dkk., 2004) menghasilkan suatu batas $c_2 = 1,15$. Jadi, jika hasil perbandingan antara rasio ukuran jarak lebih besar dari batas $c_2 = 1,15$, maka banyaknya *cluster* tersebut merupakan solusi akhir. Melihat pada rasio perubahan nilai BIC pada banyaknya *cluster* 3 (0,626) jauh lebih besar dibandingkan banyaknya *cluster* 5 (-0,390). Berdasarkan evaluasi nilai BIC, rasio perubahan nilai BIC, dan rasio ukuran jarak, dapat disimpulkan bahwa jumlah *cluster* optimal pada minat mahasiswa berbelanja online adalah 3 *cluster* optimal. Dengan metode *two-step cluster*, dapat diketahui juga pada setiap *cluster* berapa banyak peminat mahasiswa berbelanja *online* atau ciri khas yang membedakan antara setiap *clusternya*.

Tabel 3. Ciri Khas setiap *Cluster* Minat Mahasiswa Berbelanja *Online*

| karakteristik | | <i>Cluster</i> | | |
|-------------------------------|-------------------|----------------|----|----|
| | | 1 | 2 | 3 |
| Jenis Kelamin (X_1) | Laki-laki | 25 | 11 | 14 |
| | Perempuan | 4 | 35 | 11 |
| Tahun Angkatan (X_2) | 2021 | 3 | 4 | 1 |
| | 2020 | 3 | 0 | 2 |
| | 2019 | 18 | 4 | 8 |
| | 2018 | 5 | 38 | 14 |
| | Kimia | 13 | 7 | 8 |
| Jurusan (X_3) | Matematika | 16 | 20 | 12 |
| | Fisika | 0 | 19 | 5 |
| | Tunai | 0 | 5 | 21 |
| Pembayaran (X_4) | Non-Tunai | 29 | 41 | 4 |
| | Tokopedia | 10 | 8 | 0 |
| Platform E-commerce (X_6) | Shopee | 18 | 38 | 11 |
| | Lazada | 1 | 0 | 0 |
| | Grab | 0 | 0 | 4 |
| | Gojek | 0 | 0 | 10 |
| | Kosmetik | 0 | 20 | 1 |
| Jenis Produk (X_7) | Pakaian | 11 | 8 | 4 |
| | Alat Elektronik | 9 | 7 | 3 |
| | Alat Rumah Tangga | 2 | 3 | 0 |
| | Alat Olahraga | 1 | 0 | 0 |
| | <i>Top-up</i> | 6 | 4 | 1 |
| | Makanan | 0 | 4 | 14 |

Tabel 4. Variabel Kontinu *Cluster*

| Variabel Kontinu | <i>Cluster</i> | Rata-rata |
|-----------------------------------|----------------|------------------|
| Biaya Belanja per Bulan (X_8) | 1 | Rp. 186.172,41 |
| | 2 | Rp. 202.565,22 |
| | 3 | Rp. 120.400,00 |
| Variabel Kontinu | <i>Cluster</i> | Rata-rata |
| Uang Saku per Bulan (X_9) | 1 | Rp. 1.234.482,76 |
| | 2 | Rp. 1.311.956,52 |
| | 3 | Rp. 1.168.000,00 |

Berdasarkan Tabel 3 dan 4, informasi yang diperoleh antara lain:

1. *Cluster* 1 terdiri dari 29 orang mahasiswa dengan laki-laki (25 orang) dan perempuan (4 orang). Mereka berasal dari tahun angkatan 2018 (5 orang), tahun 2019 (18 orang), tahun 2020 (3 orang), dengan jurusan kimia (13 orang) dan Matematika (16 orang). *Platform E-commerce* Shopee (18 orang), Tokopedia (10 orang), dan Lazada (1 orang). Semuanya menggunakan pembayaran non-tunai. Jenis produk diminati pakaian (11 orang), alat elektronik (9 orang), *top-up* (6 orang), alat rumah tangga (2 orang), dan alat olahraga (1 orang). Sebagian besar mahasiswa biaya belanja *online* sebesar Rp. 186.172/bulan dan mendapatkan uang saku Rp. 1.234.482/bulan.
2. *Cluster* 2 terdiri dari 46 orang mahasiswa dengan laki-laki (11 orang) dan perempuan (35 orang). Mereka berasal dari tahun angkatan 2018 (38 orang), tahun 2019 (4 orang), dan tahun 2021 (4 orang) dengan jurusan kimia (7 orang), Matematika (20 orang), dan Fisika (19 orang). *Platform E-commerce* Shopee (38 orang), Tokopedia (8 orang), mereka menggunakan pembayaran tunai (5 orang) dan non-tunai (41 orang). Jenis produk yang diminati meliputi kosmetik (20 orang), pakaian (8 orang), alat elektronik (7 orang), *top-up* (4 orang), makanan (4 orang). Sebagian besar mahasiswa biaya belanja *online* sebesar Rp. 202.565/bulan dan mendapatkan uang saku Rp. 1.311.956/bulan.
3. *Cluster* 3 terdiri dari 25 orang mahasiswa dengan laki-laki (14 orang) dan perempuan (11 orang). Mereka berasal dari tahun angkatan 2018 (14 orang),

tahun 2019 (8 orang), tahun 2020 (2 orang), tahun 2021 (1 orang) dengan jurusan kimia (8 orang), Matematika (12 orang), dan Fisika (5 orang). Platform *E-commerce* Shopee(11), Grab (4 orang), dan Gojek (10 orang). mereka menggunakan pembayaran dengan tunai (21 orang) dan non-tunai (4 orang). Jenis produk yang diminati meliputi kosmetik (1 orang), pakaian (4 orang), alat elektronik (3 orang), *top-up* (1 orang), dan makanan (14 orang). Sebagian besar mahasiswa biaya belanja *online* sebesar Rp. 120.000/bulan dan mendapatkan uang saku Rp. 1.168.000/bulan.

4. KESIMPULAN

Berdasarkan hasil analisis metode *two-step cluster* pada data mahasiswa aktif S1 Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Jenderal Soedirman dalam minat berbelanja *online*, karakteristik mahasiswa dapat dikelompokkan menjadi 3 *cluster* optimal berdasarkan nilai BIC. *Cluster 1* yang terdiri dari 25 mahasiswa laki-laki dan 4 mahasiswa perempuan dengan pilihan platform Shopee, Tokopedia, dan Lazada. *Cluster 2* yang terdiri dari 11 mahasiswa laki-laki dan 35 mahasiswa perempuan dengan pilihan platform Shopee dan Tokopedia. *Cluster 3* yang terdiri dari 14 mahasiswa laki-laki dan 11 mahasiswa perempuan dengan pilihan platform Shopee, Gojek, dan Grab *Cluster 1* mempunyai persentase belanja online 15% dari uang saku perbulan. *Cluster 2* mempunyai persentase belanja online 15,4% dari uang saku per bulan. *Cluster 3* mempunyai persentase belanja online 10% dari uang saku perbulan.

DAFTAR PUSTAKA

- Bacher, J., Wenzig, K., dan Vogler, M., *SPSS Two Step Cluster- A First Evaluation*, 2024.
- Chiu, T., Fang, D., Chen, J., Wang, Y., dan Jeris, C., *A Robust and Scalable Clustering Algorithm for Mixed Type Attributes in Large Database Environment*, Proceedings of The Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 2021, 263-268).

- Johnson, R. A. dan Wichern, D. W., *Applied Multivariat Statistics*, Ed. ke-5, Prantice Hall, New Jersey, 2002.
- Mongi, C. E., *Penggunaan Analisis Two Step Clustering untuk Data Campuran*, Journal de Cartesian, 2015.
- Norusis, M. J., *SPSS 19.0 Statistical Procedures Companion*, Prentice-Hall, Upper Saddle River, NJ (US), 2004.
- Rundle-Thiele, S., Kubacki, K., Tkaczynski, A., dan Parkinson, J., *Using Two-Step Cluster Analysis to Identify Homogeneous Physical Activity Groups*, Marketing Intelligence & Planning, **33**(4) (2015), 522-537.
- Schiopu, D., *Applying Two Step Cluster Analysis for Identifying Bank Customers' Profile*, EL-TC, **62**(3) (2010), 66-75.
- Tkaczynski, A., *Segmentation Using Two-Step Cluster Analysis*, Segmentation in Social Marketing: Process, Methods and Application, 2017, 109-125.