

ISSN : **2987-078X**
E-ISSN : **2987-078X**
DOI : **10.30092/tabayyun** by Crossref

Volume 04 Nomor 1 Juni 2023,
<http://jurnal.radenfatah.ac.id/index.php/tabayyun>

ANALISIS CLUSTER : K-MEANS CLUSTER DALAM PENGELOMPOKKAN PENGGUNAAN MEDIA HIBURAN BERDASARKAN PROFIL PENGGUNA

Jean Evelyn Ilela¹⁾

¹⁾Institut Agama Kristen Negeri Ambon, Indonesia

Email: jeanevelyn111@gmail.com

History	Penerbit: Program Studi Jurnalistik Universitas Islam Negeri Raden Fatah, Palembang, Indonesia
Received : 24 Juni 2023	Berlisensi: Karya ini dilisensikan di bawah
Revised : 25 Juni 2023	ALisensi Creative Commons Atribusi 4.0 Internasional.
Accepted : 25 Juni 2023	
Published : 30 Juni 2023	

Abstract

Entertainment media is no longer just for watching movies and hearing news, but more than that it is a bridge to access broader information and even a business. The purpose of this study is to classify the use of entertainment media using the *Cluster Analysis* approach in terms of user profiles. *Cluster analysis* is a statistical technique that functions in grouping data into certain groups with the understanding that certain data have similarities or certain characteristics that are close together. This study took directly with the number of respondents 70 people. In collecting data, the length of use of mobile phones in accessing entertainment media was taken, the most frequently accessed entertainment media, and the reasons for using access to entertainment media. The results obtained are three *clusters*, namely *Cluster 1*: contains respondents who have an age less than the population average, has an average job as a student and private sector, *Cluster 2*: contains respondents who have an age less than the population average, has an average job of the average student, *Cluster 3*: contains respondents who are older than the population average, have an average job as a civil servant and are not yet working. the use of entertainment media by respondents in the three existing *clusters* is very different from one *cluster* to another. Overall all variables have differences in the three *clusters* because the significance level is not more than 0.05.

Keywords: *cluster analysis; entertainment; media; user profile*

Abstrak

Media hiburan bukan lagi sekedar hanya untuk menonton film dan mendengar berita namun lebih dari itu menjadi jembatan untuk mengakses informasi yang lebih luas bahkan menjadi sebuah bisnis. Tujuan penelitian ini yaitu mengelompokkan penggunaan media hiburan dengan pendekatan Analisis *Cluster* dilihat dari profil pengguna. Analisis *Cluster* adalah sebuah teknik statistika yang berfungsi dalam mengelompokkan data ke dalam beberapa kelompok tertentu dengan pemahaman bahwa sejumlah data tertentu memiliki kemiripan atau karakteristik tertentu yang berdekatan. Penelitian ini mengambil secara langsung dengan jumlah responden 70 orang. Dalam pengumpulan data diambil lamanya penggunaan *handphone* dalam mengakses media hiburan, media hiburan yang paling sering diakses, dan alasan penggunaan akses media

hiburan. Hasil diperoleh terbentuk tiga *cluster*, yaitu *Cluster* 1: berisi responden yang mempunyai usia kurang dari rata-rata populasi, memiliki pekerjaan rata-rata sebagai mahasiswa dan swasta, *Cluster* 2: berisi responden yang mempunyai usia kurang dari rata-rata populasi, memiliki pekerjaan rata-rata sebagai mahasiswa, *Cluster* 3: berisi responden yang mempunyai usia lebih dari rata-rata populasi, memiliki pekerjaan rata-rata sebagai PNS dan belum bekerja. penggunaan media hiburan oleh responden pada ketiga *cluster* yang ada sangat berbeda antar-*cluster* yang satu dengan *cluster* lainnya. Secara keseluruhan semua variabel memiliki perbedaan pada ketiga *cluster* karena tingkat signifikansi tidak lebih dari 0,05.

Kata Kunci: *cluster*; media; hiburan; profil; pengguna

Pendahuluan

Perkembangan media hiburan dewasa ini maju pesat seiring dengan perkembangan teknologi di era globalisasi. Media hiburan sudah menjadi sebuah kebutuhan tersendiri untuk manusia melakukan aktivitasnya, baik untuk bersantai namun juga untuk bekerja. Media hiburan yang tersedia mulai dari layanan televisi, radio, dan media sosial seperti Facebook, Instagram, Whatsapp, Twitter, Tik Tok dan lain sebagainya. Perkembangan media hiburan bukan saja pada layanan atau aplikasi yang disediakan namun media fisikpun turut berkembang, misalnya televisi yang dulu ukuran besar sekarang dibuat slim bahkan dibuat menjadi smart-TV, begitu juga gawai yang dulunya ukuran besar dan tidak bisa menyentuh layar sekarang dibuat tipis dan touch screen. Hal demikian membuktikan bahwa media hiburan telah berkembang secara cepat. Tak terlepas dari perkembangannya, penggunaan media hiburan juga dirasakan mulai dari anak kecil sampai orang lanjut usia. Media hiburan bukan lagi sekedar hanya untuk menonton film dan mendengar berita namun lebih dari itu menjadi jembatan untuk mengakses informasi yang lebih luas bahkan menjadi sebuah bisnis. Menurut Kominfo, berdasarkan data PT Bakrie Telecom, Pengguna Twitter memiliki 19,5 juta pengguna di Indonesia dari total 500 juta pengguna global, sementara jejaring sosial lain yang dikenal di Indonesia adalah Path dengan jumlah pengguna sebesar 700.000 di Indonesia. Line sebesar 10 juta pengguna, Google+ 3,4 juta pengguna dan Linkedlin 1 juta pengguna (kominfo, 2013).

Menurut Jogiyanto (2007) Perilaku adalah tindakan atau kegiatan nyata yang dilakukan karena individual mempunyai keinginan untuk melakukan sesuatu tertentu. Minat perilaku akan menentukan perilakunya (Doni & Faqih, 2017).

Dalam penelitian ini, peneliti ingin mengelompokkan penggunaan media hiburan dengan pendekatan Analisis *Cluster* dilihat dari profil pengguna. Analisis *Cluster* atau kluster adalah sebuah teknik statistika yang berfungsi dalam mengelompokkan data ke dalam beberapa kelompok tertentu dengan pemahaman bahwa sejumlah data tertentu memiliki kemiripan atau karakteristik tertentu yang berdekatan. Analisis *Cluster* terbagi atas dua metode, yaitu hirarki dan non hirarki.

Dalam penelitian ini menggunakan analisis kluster non hirarki, yaitu K-Means *Cluster*. Salah satu algoritma data minim yang cukup populer digunakan baik dalam dunia bisnis, akademik, ataupun industri adalah Algoritma *K-Means*. Algoritma ini bekerja dengan cara membagi data dalam sejumlah *cluster* untuk dianalisis faktor kesamaan (*similarity*) maupun ketidaksamaan (*dissimilarity*) yang melekat pada kumpulan data tersebut (Talakua, Leleury, & Taluta, 2017).

Menurut penelitian yang dilakukan oleh M. W. Talakua dkk mengenai Analisis *Cluster* Dengan Menggunakan Metode *K-Means* Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014 membahas tentang pemanfaatan algoritma *kmeans* untuk mengelompokkan kabupaten/kota di Provinsi Maluku berdasarkan kemiripan karakteristik daerah yang ditinjau dari lima ukuran Indeks Pembangunan Manusia (IPM). Lima ukuran tersebut adalah Angka Harapan Hidup (AHH), Angka Melek Huruf (AMH), Rata-rata Lama Sekolah (RLS), dan Pengeluaran Per Kapita (PPK). Secara umum, *output* penelitian berupa hasil klasterisasi dapat digunakan sebagai sebuah acuan untuk menggambarkan distribusi pengelompokan berdasarkan kondisi profil IPM di Maluku.

Menurut penelitian oleh Safa'at Yulianti dan K. Hidayatulah mengenai Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah, Tujuan dari penelitian ini adalah untuk mengetahui pengelompokan 35 kabupaten/kota

di Provinsi Jawa Tengah dan mengetahui karakteristik setiap kelompok berdasarkan Indikator Kesejahteraan Rakyat Tahun 2010. Meskipun dalam penelitian ini terdapat data outlier (nilai ekstrim) yaitu Kabupaten Kudus dan Kota Surakarta, kedua data outlier tersebut tetap dipertahankan karena tidak bisa dikatakan ada kesalahan pada proses sampling maupun inputing data (Yulianto & Hidayatullah, 2016).

Penelitian serupa juga diterapkan pada tahun 2020 oleh Rizkiana Prima dkk, mengenai Analisis *Cluster* Virus Corona (COVID-19) di Indonesia pada 2 Maret 2020 – 12 April 2020 dengan metode K-Means *Clustering*. Objek yang akan digunakan pada penelitian ini adalah data kasus COVID-19 untuk 34 Provinsi di Indonesia tanggal 2 Maret 2020 sampai dengan 12 April 2020 dengan menggunakan dua metode Elbow dan metode Silhouette. Data yang digunakan dalam praktikum ini merupakan data sekunder yang tersedia di website resmi Gugus Tugas Percepatan Penanganan COVID-19 dan Kawal COVID-19 yaitu data mengenai Kasus Corona Virus (COVID-19) di Indonesia pada 2 Maret 2020 sampai dengan tanggal 12 April 2020 (42 hari). Berdasarkan analisis *cluster* metode Elbow dan metode Silhouette untuk mencari jumlah *cluster* yang optimal, kedua metode tersebut menghasilkan $K = 3$. Analisis K-means *clustering* dengan $K = 3$ diperoleh kesimpulan hasil yang diperoleh ialah good fit karena memiliki variabilitas dalam *cluster* rendah dan antar *cluster* tinggi (Prima, E, & Sutikno, 2020).

Metode Penelitian

Populasi dan Sampel

Populasi yang diambil dari pemuda umur 17-35 tahun pada kader AMGPM Daerah pulau Ambon Utara dan dipilih sampel sebanyak 70 pemuda pada AMGPM Daerah pulau Ambon Utara yang diambil langsung di lapangan.

Metode Pengumpulan Data

Berikut adalah metode pengumpulan data dalam penelitian ini

- Studi Literatur

Pengumpulan data dilakukan dengan cara mempelajari, meneliti dan menelaah berbagai literatur yang bersumber dari buku, situs internet, jurnal ilmiah, dan sumber – sumber lainnya yang berkaitan dengan penelitian yang dilakukan.

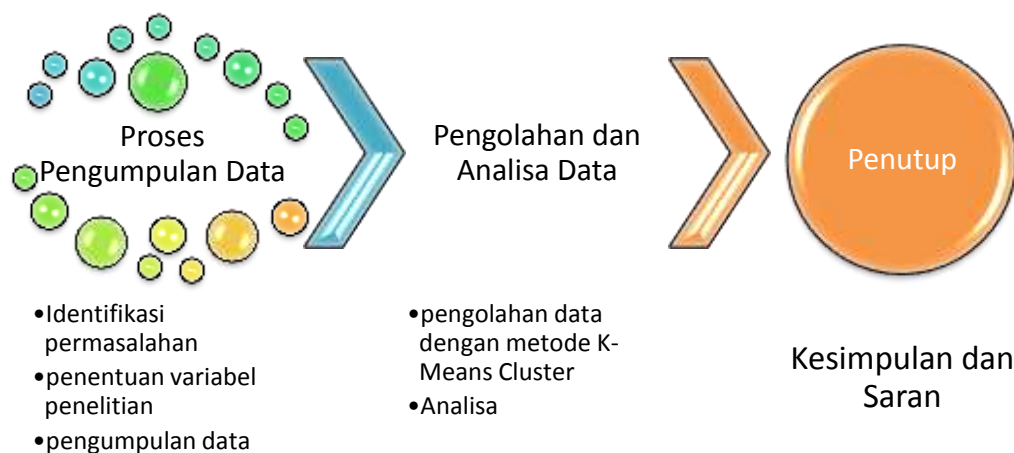
- **Pengumpulan Data Penggunaan Media Hiburan**

Pengumpulan data dengan diambil langsung di lapangan (pengukuran langsung) berdasarkan lama penggunaan gawai pada aplikasi media hiburan, media hiburan yang dipilih, alasan penggunaan, dan profil pengguna.

- **Observasi**

Observasi yang dilakukan ialah mengamati penggunaan media hiburan yang akan dianalisis terkait dengan objek penelitian.

Berikut adalah tahapan–tahapan dalam penelitian:



GAMBAR 1. Diagram Alir Penelitian

Teknik Analisis

Definisi Analisis Cluster

Cluster atau 'Kluster' dapat diartikan "Kelompok", dengan demikian, pada dasarnya analisis *cluster* akan menghasilkan sejumlah kluster (kelompok). Analisis ini diawali dengan pemahaman bahwa sejumlah data tertentu sebenarnya mempunyai kemiripan di antara anggotanya, karena itu, dimungkinkan untuk mengelompokkan anggota-anggota yang 'mirip' atau

mempunyai karakteristik yang serupa tersebut dalam satu atau lebih kluster (Santoso, 2014). Secara sederhana analisis kluster merupakan teknik analisis multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *Cluster* merupakan teknik multivariat yang mempunyai tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya.

Analisis *Cluster* mengklasifikasi objek sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. *Cluster-cluster* yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi. Berbeda dengan teknik multivariat lainnya, analisis ini tidak mengestimasi set variabel secara empiris sebaliknya menggunakan set variabel yang ditentukan oleh peneliti itu sendiri. Fokus dari Analisis *Cluster* adalah membandingkan objek berdasarkan set variabel, hal inilah yang menyebabkan para ahli mendefinisikan set variabel sebagai tahap kritis dalam analisis *cluster*. Set variabel *cluster* adalah suatu set variabel yang merpresentasikan karakteristik yang dipakai objek-objek. Solusi Analisis *Cluster* bersifat tidak unik, anggota *cluster* untuk tiap penyelesaian/solusi tergantung pada beberapa elemen prosedur dan beberapa solusi yang berbeda dapat diperoleh dengan mengubah satu elemen atau lebih. Solusi *cluster* secara keseluruhan bergantung pada variabel-variabel yang digunakan sebagai dasar untuk menilai kesamaan. Penambahan atau pengurangan variabelvariabel yang relevan dapat mempengaruhi substansi hasil analisis *cluster* (Ediyanto, Mara, & Intisari, 2013).

Proses Dasar Analisis Cluster

Pada dasarnya proses pengelompokkan mencari dan mengelompokkan data yang mirip, sehingga kriteri kemiripan (*similarity*) merupakan dasarnya. Proses pengolahan data untuk dikelompokkan menjadi satu atau beberapa kluster adalah sebagai berikut (Santoso, 2014):

- Menetapkan ukuran jarak antar data
- Melakukan proses standarisasi data jika diperlukan

- Melakukan proses *clustering*
- Melakukan penamaan kluster-kluster yang terbentuk
- Melakukan validasi dan profiling kluster

Metode Pengelompokan Analisis Cluster

Adapun metode pengelompokan dalam analisis *cluster* meliputi (Santoso, 2014):

- 1 Metode Hirarki; memulai pengelompokan dengan dua atau lebih obyek yang mempunyai kesamaan paling dekat. Kemudian diteruskan pada obyek yang lain dan seterusnya hingga *cluster* akan membentuk semacam 'pohon' dimana terdapat tingkatan (hirarki) yang jelas antar obyek, dari yang paling mirip hingga yang paling tidak mirip. Alat yang membantu untuk memperjelas proses hirarki ini disebut "dendogram".
- 2 Metode Non-Hirarki; dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan (dua, tiga, atau yang lain). Setelah jumlah *cluster* ditentukan, maka proses *cluster* dilakukan tanpa mengikuti proses hirarki. Metode ini biasa disebut "*K-Means Cluster*".

Asumsi yang harus dipenuhi dalam Analisis *Cluster* yaitu :

- Sampel yang diambil benar-benar dapat mewakili populasi yang ada (*representativeness of the sample*).
- Multikolinieritas, yakni kemungkinan adanya korelasi antar-objek.

Ada lima metode aglomerasi dalam pembentukan *cluster*, yaitu :

1 Pautan Tunggal (*Single Linkage*)

Metode ini didasarkan pada jarak minimum. Dimulai dengan dua objek yang dipisahkan dengan jarak paling pendek maka keduanya akan ditempatkan pada *cluster* pertama, dan seterusnya. Metode ini dikenal pula dengan nama pendekatan tetangga terdekat.

2 Pautan Lengkap (*Complete Linkage*)

Disebut juga pendekatan tetangga terjauh. Dasarnya adalah jarak maksimum. Dalam metode ini seluruh objek dalam suatu *cluster* dikaitkan

satu sama lain pada suatu jarak maksimum atau dengan kesamaan minimum.

3 Metode Ward (*Ward's Method*)

Dalam metode ini jarak antara dua *cluster* adalah jumlah kuadrat antara dua *cluster* untuk seluruh variabel. Metode ini cenderung digunakan untuk mengkombinasi *cluster-cluster* dengan jumlah kecil.

4 Pautan Rata-rata (*Average Linkage*)

Dasarnya adalah jarak rata-rata antar observasi. pengelompokan dimulai daritengan atau pasangan observasi dengan jarak paling mendekati jarak rata-rata.

5 Metode Centroid

Jarak antara dua *cluster* adalah jarak antar centroid *cluster* tersebut. Centroid *cluster* adalah nilai tengah observasi pada variabel dalam suatu set variabel *cluster*.

Pada penelitian ini peneliti menggunakan metode *K-Means Cluster* sebagai solusi untuk pengklasifikasian karakteristik dari data. Alasan penggunaan metode ini ialah karena metode ini memiliki tingkat ketelitian yang cukup tinggi terhadap ukuran data, sehingga relatif lebih terpresisi dan efisien untuk pengolahan data dalam jumlah besar. Selain itu urutan data tidak mempengaruhi algoritmanya.

Analisis Cluster : K-Means Cluster

K-Means merupakan salah satu metode data *clustering* nonhierarki yang berusaha mempartisi data yang ada dalam bentuk satu atau lebih *cluster*, sehingga data dengan karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama pula. Data dengan karakteristik yang berbeda dikelompokkan dalam kelompok yang lain. Metode yang termasuk dalam algoritma *clustering* berbasis jarak yang membagi data kedalam sejumlah *cluster* dan algoritma ini hanya bekerja pada atribut numerik. *K-Means* merupakan salah satu algoritma dalam *data mining* yang biasa digunakan untuk melakukan *clustering* suatu data. Ada banyak pendekatan untuk membuat *cluster*, diantaranya adalah membuat

aturan yang mendikte keanggotaan dalam kelompok yang sama berdasarkan tingkat persamaan di antara anggotaanggotanya. Pendekatan lainnya adalah dengan membuat sekumpulan fungsi yang mengukur beberapa properti dari pengelompokan tersebut sebagai fungsi dari beberapa parameter dari sebuah *clustering* (Anggara, Sujiani, & Helfi, 2016).

Mengelompokan data dengan metode *K-Means* dilakukan dengan algoritma sebagai berikut (Talakua et al., 2017).

- 1) Tentukan jumlah kelompok.
- 2) Alokasikan data ke dalam kelompok secara acak.
- 3) Hitung pusat kelompok dari data yang ada di masing-masing kelompok.

Lokasi *centroid* setiap kelompok diambil dari rata-rata semua nilai data dari setiap fiturnya. Jika M menyatakan jumlah data dalam sebuah kelompok, i menyatakan fitur ke- i dalam sebuah kelompok, dan p menyatakan dimensi data, maka persamaan untuk menghitung *centroid* fitur ke- i digunakan persamaan (1).

$$K = \frac{1}{M} \sum_{j=1}^M X_j \quad (1)$$

Persamaan (1) dilakukan sebanyak p dimensi dari $i = 1$ sampai dengan $i = p$.

- 4) Alokasikan masing-masing data ke *centroid*/rata-rata terdekat. Ada beberapa cara yang dapat dilakukan untuk mengukur jarak data ke pusat kelompok, diantaranya adalah Euclidean pengukuran jarak pada ruang jarak (*distance space*). Euclidean dapat dicari menggunakan persamaan (2).

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \quad (2)$$

Pengalokasian kembali data ke dalam masing-masing kelompok dalam metode *K-Means* didasarkan pada perbandingan jarak antara data dengan *centroid* setiap kelompok yang ada. Data dialokasikan ulang secara tegas ke kelompok yang mempunyai *centroid* dengan jarak terdekat dari data tersebut.

Pengalokasian data ini menurut MacQueen (1967) dapat ditentukan menggunakan persamaan (3).

$$a_{i1} = \begin{cases} 1, & d = \min\{D(X_i, C_l)\} \\ 0, & \text{lainnya} \end{cases} \quad (3)$$

a_{i1} adalah nilai keanggotaan titik x_i ke pusat kelompok C_l , d adalah jarak terpendek dari data x_i ke K kelompok setelah dibandingkan, dan C_l adalah *centroid* ke-1. Fungsi objektif yang digunakan untuk metode *K-Means* ditentukan berdasarkan jarak dan nilai keanggotaan data dalam kelompok. Fungsi objektif menurut MacQueen (1967) dapat ditentukan menggunakan persamaan (4).

$$J = \sum_{i=0}^n \sum_{l=1}^k a_{il} D(x_i, c_l)^2 \quad (4)$$

dengan n adalah jumlah data, k adalah jumlah kelompok, a_{i1} adalah nilai keanggotaan titik data x_i ke kelompok c_1 yang diikuti. a mempunyai nilai 0 atau 1. Apabila data merupakan anggota suatu kelompok, nilai $a_{i1} = 1$. Jika tidak, nilai $a_{i1} = 0$. Kembali ke langkah 3, apabila masih ada data yang berpindah kelompok atau apabila ada perubahan nilai *centroid* di atas nilai ambang yang ditentukan, atau apabila perubahan nilai pada fungsi objektif yang digunakan masih di atas nilai ambang yang ditentukan.

Hasil dan Pembahasan

1. Hasil

Penelitian ini mengambil secara langsung dengan jumlah responden 70 orang dengan rentang umur 17-35 tahun dengan latar belakang pekerjaan yang berbeda. Dalam pengumpulan data diambil lamanya penggunaan handphone dalam mengakses media hiburan, media hiburan yang paling sering diakses, dan alasan penggunaan akses media hiburan.

Pada hasil yang ditunjukkan dari SPSS terlihat bahwa semua data lengkap dan tidak ada yang mising.

TABEL 1. *Descriptive Statistics*

	N	Minimum	Maximum	Mean	Std. Deviation
Alasan Penggunaan	70	1	9	3,33	2,076

Umur	70	17	35	24,57	4,754
Pekerjaan	70	1	4	2,26	1,212
Lama Penggunaan Media Hiburan (Jam)	70	3	24	13,23	5,256
Media Hiburan yang paling sering digunakan	70	1	9	3,34	2,132
Valid N (listwise)	70				

Data Quick Cluster

TABEL 2. Initial Cluster Centers

	Cluster		
	1	2	3
Zscore: Alasan Penggunaan	2,24995	-,15825	-,15825
Zscore(Umur)	1,98348	-,75132	2,19385
Zscore(Pekerjaan)	-,21215	1,43790	,61288
Zscore: Lama Penggunaan Media Hiburan (Jam)	1,28826	2,04926	-1,94598
Zscore: Media Hiburan yang paling sering digunakan	,77710	-1,09866	-1,09866

Output ini adalah tampilan pertama (*initial*) proses *clustering* data sebelum melakukan iterasi. Karena nanti akan dihasilkan proses *clustering* sesudah iterasi yang justru adalah hasil akhir *cluster*, maka *output* ini tidak akan dianalisis.

Proses Iterasi

Setelah melalui tahapan iterasi maka diperoleh hasil *final cluster* berikut ini.

TABEL 3. Final Cluster Centers

	Cluster		
	1	2	3
Zscore: Alasan Penggunaan	1,17199	-,35091	-,56906
Zscore(Umur)	-,50088	-,87754	,69652
Zscore(Pekerjaan)	-,60502	-,76217	,70994
Zscore: Lama Penggunaan Media Hiburan (Jam)	,17394	,42580	-,29529

Zscore: Media Hiburan yang paling sering digunakan	1,22371	-,91108	-,35387
--	---------	---------	---------

Angka di atas terkait proses standarisasi data sebelumnya, yang mengacu pada angka z, dengan ketentuan:

- Angka negatif berarti data di bawah rata-rata total
- Angka positif berarti data di atas rata-rata total

Dengan rata-rata sampel diperoleh dari:

$$X = \mu + z.\sigma \quad (5)$$

Dimana:

X = Rata-rata sampel (dalam hal ini rata-rata variabel pada *cluster* tertentu)

μ = Rata-rata populasi

σ = Standar deviasi

Z = Nilai standarisasi yang diperoleh pada SPSS

Maka diperoleh hasil berikut ini:

TABEL 4. Mean Cluster

	Mean X Cluster		
	1	2	3
Zscore: Alasan Penggunaan	5,763051	2,601511	2,148631
Zscore(Umur)	22,18882	20,39817	27,88126
Zscore(Pekerjaan)	1,526716	1,33625	3,120447
Zscore: Lama Penggunaan Media Hiburan (Jam)	14,14423	15,468	11,67796
Zscore: Media Hiburan yang paling sering digunakan	5,94895	1,397577	2,585549

Tafsiran setiap *cluster*:

- *Cluster* 1: berisi responden yang mempunyai usia kurang dari rata-rata populasi, memiliki pekerjaan rata-rata sebagai mahasiswa dan swasta, menggunakan media hiburan lebih lama yaitu 14 jam dari rata-rata populasi, media hiburan yang paling sering digunakan rata-rata adalah

Instagram dan Google, dan memiliki alasan penggunaan lebih banyak mencari hiburan dan mencari tugas.

- *Cluster 2*: berisi responden yang mempunyai usia kurang dari rata-rata populasi, memiliki pekerjaan rata-rata sebagai mahasiswa, menggunakan media hiburan lebih lama yakni sekitar 15 jam dari rata-rata populasi, media hiburan yang paling sering digunakan rata-rata adalah game online dan Whatsapp, dan memiliki alasan penggunaan lebih banyak mencari informasi dan bermain game.
- *Cluster 3*: berisi responden yang mempunyai usia lebih dari rata-rata populasi, memiliki pekerjaan rata-rata sebagai PNS dan belum bekerja, menggunakan media hiburan kurang lama yaitu 11 jam dari rata-rata populasi, media hiburan yang paling sering digunakan rata-rata adalah Facebook dan Whatsapp, dan memiliki alasan penggunaan lebih banyak mencari informasi.

Perbedaan Variabel Pada *Cluster* Yang Terbentuk

Setelah membentuk 3 *cluster* maka selanjutnya adalah melihat apakah variabel-variabel yang telah terbentuk dalam *cluster* memiliki perbedaan pada setiap *cluster*. Hal ini ditunjukkan dari *F* dan nilai signifikansi (probabilitas) masing-masing variabel. Ditunjukkan melalui output ANOVA berikut ini.

TABEL 5. ANOVA

	<i>Cluster</i>		Error			
	Mean Square	df	Mean Square	df	F	Sig.
Zscore: Alasan Penggunaan	20,851	2	,407	67	51,177	,000
Zscore(Umur)	16,657	2	,533	67	31,274	,000
Zscore(Pekerjaan)	16,768	2	,529	67	31,680	,000
Zscore: Lama Penggunaan Media Hiburan (Jam)	3,160	2	,936	67	3,378	,040
Zscore: Media Hiburan yang paling sering digunakan	24,078	2	,311	67	77,394	,000

The F tests should be used only for descriptive purposes because the *clusters* have been chosen to maximize the differences among cases in different *clusters*. The observed significance levels are not corrected for this and thus cannot be interpreted as tests of the hypothesis that the *cluster* means are equal.

Kolom *cluster* menunjukkan besaran *between cluster mean*, sedangkan kolom error menunjukkan besaran *within cluster mean*, sehingga kolom F didapat dari:

$$F = \frac{\text{Between Means}}{\text{Within Means}} \quad (6)$$

Interpretasi dari tabel ANOVA di atas pada prinsipnya, semakin besar angka F suatu variabel, dan angka signifikansinya adalah dibawah 0,05, maka semakin besar pula perbedaan variabel tersebut pada ketiga *cluster* yang terbentuk. Maka berdasarkan tabel ANOVA di atas maka Zscore: Media Hiburan Yang Paling Sering Digunakan yang memiliki nilai F terbesar yaitu 77, 394 dengan signifikansi 0,000, yang berarti media hiburan yang paling sering digunakan membedakan karakteristik ketiga *cluster*. Atau dengan kata lain penggunaan media hiburan oleh responden pada ketiga *cluster* yang ada sangat berbeda antar-*cluster* yang satu dengan *cluster* lainnya. Secara keseluruhan semua variabel memiliki perbedaan pada ketiga *cluster* karena tingkat signifikansi tidak lebih dari 0,05.

2. Diskusi

Dalam penelitian ini ditemukan beberapa hal terkait penggunaan media sosial yang terbagi dalam tiga *cluster*. Pada ketiga *cluster* terdapat perbedaan minat dalam penggunaan media sosial berdasarkan profil pengguna, yaitu pada profil pengguna cluster 1 sebagai mahasiswa dan swasta yang berusia sekitar 22 tahun lebih banyak menghabiskan waktunya untuk mencari hiburan dan tugas dengan waktu rata penggunaan 14 jam, sementara profil pengguna cluster 2 rata-rata sebagai mahasiswa lebih banyak menghabiskan waktunya untuk bermain game dan mencari informasi dengan durasi waktu rata-rata 15 jam, sedangkan profil pengguna cluster 3 yang rata-rata PNS dan orang masih menganggur lebih banyak menggunakan media sosial untuk mencari informasi dengan durasi waktu rata-rata 11 jam. Penelitian ini justru menarik karena pada penelitian Prasetio dan Hariyani tahun 2020 hasil penelitiannya secara simultan

menunjukkan bahwa motivasi dalam penggunaan media sosial dan pengetahuan kewirausahaan berpengaruh secara signifikan terhadap minat berwirausaha (Prasetio & Hariyani, 2020). Namun, pada penelitian ini justru diperoleh hasil rata-rata profil mahasiswa justru menghabiskan waktunya untuk bermain game, mengerjakan tugas dan mencari informasi, tidak ada alasan penggunaan untuk berwirausaha. Maka ke depannya penelitian ini memungkinkan untuk meneliti lebih dalam terkait hasil penelitian ini.

Kesimpulan dan saran

Berdasarkan tabel ANOVA di atas maka Zscore: Media Hiburan Yang Paling Sering Digunakan yang memiliki nilai F terbesar yaitu 77,394 dengan signifikansi 0,000, yang berarti media hiburan yang paling sering digunakan membedakan karakteristik ketiga *cluster*. dengan terbentuk tiga *cluster*, yaitu *Cluster 1*: berisi responden yang mempunyai usia kurang dari rata-rata populasi, memiliki pekerjaan rata-rata sebagai mahasiswa dan swasta, *Cluster 2*: berisi responden yang mempunyai usia kurang dari rata-rata populasi, memiliki pekerjaan rata-rata sebagai mahasiswa, *Cluster 3*: berisi responden yang mempunyai usia lebih dari rata-rata populasi, memiliki pekerjaan rata-rata sebagai PNS dan belum bekerja.

Referensi

- Anggara, M., Sujiani, H., & Helfi, N. (2016). Pemilihan Distance Measure Pada K-Means Clustering Untuk Pengelompokan Member Di Alvaro Fitness. *Jurnal Sistem Dan Teknologi Informasi*.
- Doni, F. R., & Faqih, H. (2017). Perilaku Penggunaan Media Sosial Pada Kalangan Remaja. *IJSE-Indonesian Journal on Software Engineering*, 3(2).
- Ediyanto, Mara, N., & Intisari, N. S. (2013). Pengklasifikasian Karakteristik Dengan Metode K-Means Cluster Analysis. *Buletin Ilmiah Mat. Stat. Dan Terapannya (Bimaster)*.
- Prasetio, T., & Hariyani, R. (2020). Pengaruh Motivasi Dalam Penggunaan Media Sosial Dan Pengetahuan Kewirausahaan Terhadap Minat Berwirausaha Pada Mahasiswa. *Jurnal IKRA ITH Ekonomika*, 3(3), 94-101.

- Prima, R., E, Y. A., & Sutikno. (2020). Analisis Cluster Virus Corona (COVID-19) di Indonesia pada 2 Maret 2020 – 12 April 2020 dengan Metode K-Means Clustering. *Public Health*.
- Santoso, S. (2014). Statistik Multivariat Edisi Revisi. In *Jakarta: PT Elex Media Komputindo*.
- Talakua, M. W., Leleury, Z. A., & Taluta, A. W. (2017). ANALISIS CLUSTER DENGAN MENGGUNAKAN METODE K-MEANS UNTUK PENGELOMPOKKAN KABUPATEN/KOTA DI PROVINSI MALUKU BERDASARKAN INDIKATOR INDEKS PEMBANGUNAN MANUSIA TAHUN 2014. *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*. <https://doi.org/10.30598/barekengvol11iss2pp119-128>
- Yulianto, S., & Hidayatullah, K. H. (2016). Analisis Klaster Untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Indikator Kesejahteraan Rakyat. *Statistika*.