

Segmentasi Gaya Hidup Mahasiswa Menggunakan Algoritma *K-Means Clustering*

Nur Kholidin¹, Nana Suarna², Willy Prihartono³

¹²Teknik Informatika, STMIK IKMI Cirebon
Jl.perjuangan No.10B, Karyamulya kec.Kesambi kota Cirebon, Indonesia

¹Nkholidin657@gmail.com

²St_nana@yahoo.com

³Komputerisasi Akuntansi, STMIK IKMI Cirebon
Jl.Perjuangan No.10B Karyamulya kec.Kesambi kota Cirebon, Indonesia

³Willyprihartono@yahoo.com

Abstract

Gaya hidup merupakan salah satu kunci menjaga agar tubuh tetap sehat dan bugar. Mengabaikan pola hidup sehat bisa membuat tubuh lesud dan mudah terserang penyakit. Masalah yang teridentifikasi termasuk sakit maag akibat makan tidak teratur dan efek samping seperti stres, penambahan berat badan, dan penurunan konsentrasi karena seringnya begadang. Penelitian ini berupaya untuk meningkatkan kesadaran mahasiswa untuk hidup sehat dengan memanfaatkan algoritma *K-means clustering* untuk segmentasinya. Data yang digunakan pada proses segmentasi ini berjumlah 214 data dan 16 atribut. Temuan menunjukkan segmentasi optimal pada $K = 2$ menghasilkan rata-rata jarak pusat *cluster* sebesar 0.815, dengan rata-rata jarak pada *cluster 0* adalah 0.779 yang berisi 51 mahasiswa dan masuk kedalam kategori gaya hidup sehat. Pada *cluster 1* rata-rata jaraknya adalah 0.925 berisi 155 mahasiswa dan masuk kedalam kategori gaya hidup tidak sehat. Kemudian pada evaluasi *cluster* menggunakan metode *Davies Bouldin Index* memperoleh nilai 0,118. Hasil ini memberikan wawasan berharga untuk inisiatif kesehatan kampus yang ditargetkan dan disesuaikan dengan kebutuhan gaya hidup mahasiswa. Dengan meningkatkan kesadaran kesehatan, penelitian ini mengantisipasi pengaruh positif terhadap kesehatan mahasiswa secara keseluruhan dan prestasi akademik.

Keywords: *segmentasi, gaya hidup mahasiswa, K-means, clustering, davies bouldin index*

1. Introduction

Mahasiswa adalah salah satu kelompok yang sangat rentan akan tekanan akademik [1]. Serta gaya hidup tidak baik atau cenderung tidak sehat seperti kualitas tidur, pola makan, dan aktivitas fisik [2]. Untuk mencapai kesadaran kesehatan, diperlukan kekonsistenan yang mencakup aspek fisik, mental, dan emosional [3]. Pengertian gaya hidup menurut Setiadi (2010:148) dalam jurnal "pengaruh literasi keuangan, gaya hidup pada perilaku keuangan pada generasi milenial" menyatakan bahwa gaya hidup didefinisikan sebagai cara hidup yang didefinisikan oleh bagaimana seseorang menghabiskan waktu mereka, serta pemikiran mereka tentang diri sendiri dan juga dunia sekelilingnya [4]. kemudian penelitian yang dilakukan oleh Vebriyani dengan judul "Pengaruh Media Sosial dan Gaya Hidup Hedonis terhadap Perilaku Konsumtif Mahasiswa (Studi Kasus pada Mahasiswa Fakultas Ekonomi dan Bisnis Islam IAIN Kudus Angkatan 2016)" menjelaskan pengertian gaya hidup ialah cara hidup dan tingkah laku seseorang mengenai bagaimana seseorang tersebut menghabiskan waktunya (aktivitas), apa yang dianggap penting dalam lingkungannya (*interes*) serta bagaimana seseorang memikirkan diri sendiri dan dunia sekelilingnya (*opini*) [5].

Pada penelitian ini hal yang penting dilakukan yaitu pengumpulan dan pengelolaan data kesehatan mahasiswa secara efektif merupakan tantangan besar. Tantangan analitisnya adalah menerapkan metode *K-Means clustering* yang tepat dan representatif untuk mengidentifikasi segmen mahasiswa yang berpengetahuan luas. Teknik clustering merupakan proses

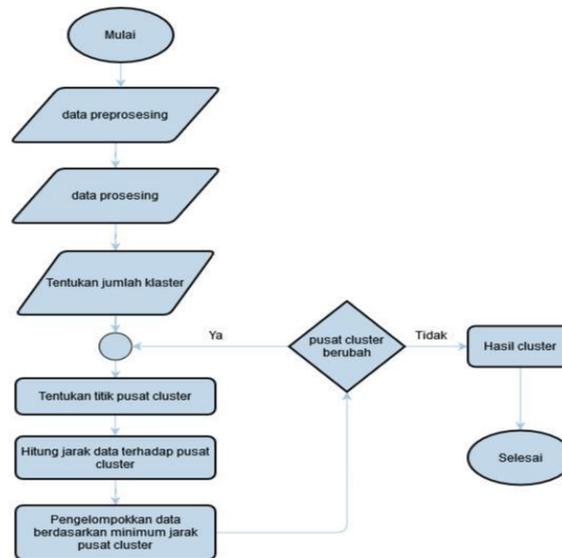
pembentukan kelompok data (*cluster*) dari kumpulan data yang kelompok atau kelasnya tidak diketahui, dan menentukan data mana yang termasuk dalam cluster mana [6]. selain itu teknik *clustering* juga suatu proses untuk menentukan kelas taksonomi atau fikologis untuk analisis topologi data yang ada. Teknik *cluster* mempunyai dua metode dalam pengelompokannya, yaitu hirarki dan non-hirarki. Pengelompokan hirarki adalah teknik pengelompokan data yang bekerja dengan cara mengelompokkan dua atau lebih bagian data yang memiliki persamaan [7]. Sedangkan Algoritma *K-Means* adalah algoritma analisis data *unsupervised* menggunakan sistem partisi. Algoritma ini mempunyai masukan data dan nilai *K* yang artinya data atau objek tersebut tersebar ke dalam *cluster*, dan setiap *cluster* mempunyai nilai titik pusat (*centroid*) yang mewakili cluster tersebut [8]. Pendapat lain juga dikemukakan oleh Ahsina, bahwa Algoritma *K-means* merupakan fungsi yang bertujuan untuk mencapai kesamaan dalam kelompok tinggi dan kesamaan dalam kelompok rendah, pengelompokan menggunakan teknik partisi berbasis titik pusat (*centroid*), dimana setiap *centroid* kelompok mewakili karakteristik setiap kelompok tersebut [9].

Penelitian terkait juga sudah dilakukan oleh beberapa peneliti sebelumnya. Diantaranya oleh [10] membahas mengenai penerapan algoritma *k-means* dalam segmentasi mahasiswa saat penerimaan mahasiswa baru UIN Sultan Syarif Kasim Riau berasal dari sekolah unggulan atau bukan. Temuan analisis ini diharapkan bisa memberikan pandangan dan wawasan yang berharga serta menjadi landasan untuk merancang strategi yang lebih dalam penerimaan mahasiswa baru di tahun berikutnya, pengembangan kurikulum, serta langkah-langkah lainnya yang diperlukan untuk meningkatkan kualitas dan kuantitas pengalaman perkuliahan. Penelitian lain oleh [11] menerangkan bahwa *K-Means clustering* dalam Penerapan metode *Crisp-DM* dengan algoritma *K-means clustering* untuk segmentasi mahasiswa berdasarkan kualitas akademik, juga bisa digunakan untuk menggali serta mendapatkan informasi yang berguna mengenai karakteristik dari mahasiswa dan dapat dijadikan sebagai bahan untuk menyusun program-program yang lebih baik. Penelitian oleh [12] yang membahas mengenai penggunaan algoritma *k-means* untuk segmentasi pelanggan pada aplikasi *alfagift*. Pada penelitian ini diterapkan strategi-strategi penerapan dengan membagi pelanggan ke beberapa kelompok atau segmen yang berbeda. Beberapa karakteristik yang diperiksa antara lain perilaku serta kebutuhan pelanggan yang berbeda-beda. Hal ini dilakukan sebagai data pendukung dalam menentukan dan mengetahui loyalitas serta preferensi untuk pelanggan sekaligus sebagai sarana media promosi bagai pengembang dimasa yang akan datang. Temuan dari analisis ini diharapkan memberikan wawasan yang baru dan berharga untuk merancang strategi dalam berbagai bidang. Algoritma *K-means clustering* terbukti efektif dalam mengelompokkan objek berdasarkan propertinya, menghasilkan informasi yang berguna untuk perbaikan dan pengembangan lebih lanjut.

Sebanyak 214 data telah berhasil dikumpulkan melalui kuesioner dari mahasiswa Program Studi Teknik Informatika angkatan 2020 di STMIK IKMI Cirebon. Kuesioner ini mencakup 16 atribut yang relevan dengan penelitian yang dilakukan. Data yang terkumpul memberikan gambaran yang komprehensif mengenai berbagai aspek yang diteliti dalam konteks penelitian.

2. Research Methods

Metode penelitian yang digunakan dalam penelitian ini adalah algoritma *K-means clustering*. algoritma *K-Means* merupakan metode data *clustering* non hirarki, *clustering* adalah salah satu metode data mining bersifat tanpa pebelan atau bisa disebut *unsupervised* [13]. Yang mempartisi menjadi satu kelompok sehingga data yang memiliki karakteristik serupa dikelompokkan ke dalam satu *cluster* yang sama dan data yang karakteristiknya berbeda dikelompokkan ke dalam kelompok lain [14]. Tujuannya untuk mengelompokkan data ke dalam suatu segmen berdasarkan kemiripan karakteristik data. Tahapan untuk menerapkan metode *k-means clustering* bisa dilihat pada gambar 1.



Gambar 1. Metode penelitian

Tahapan untuk melakukan segmentasi menggunakan algoritma *k-menas clustering*:

1. Menentukan jumlah *cluster* yang diinginkan sesuai dengan kelompok yang akan dibentuk
2. Tetapkan *k* pusat *cluster* awal secara random
3. Menghitung jarak data terdekat terhadap pusat *cluster* menggunakan rumus persamaan *euclidean*.

$$d_{Euclidean}(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2} \quad (1)$$

Dimana $d(x, y)$ yaitu jarak antar variabel data hasil kuesioner dan variabel *centroid*, $x_i = (x_1, x_2, \dots, x_i)$ yaitu variabel data hasil kuesioner, $y_i = (y_1, y_2, \dots, y_i)$ yaitu variabel pada *centroid*.

4. Pengelompokan data berdasarkan minimum jarak pusat *cluster* terdekat
5. Melakukan perhitungan *centroid* baru

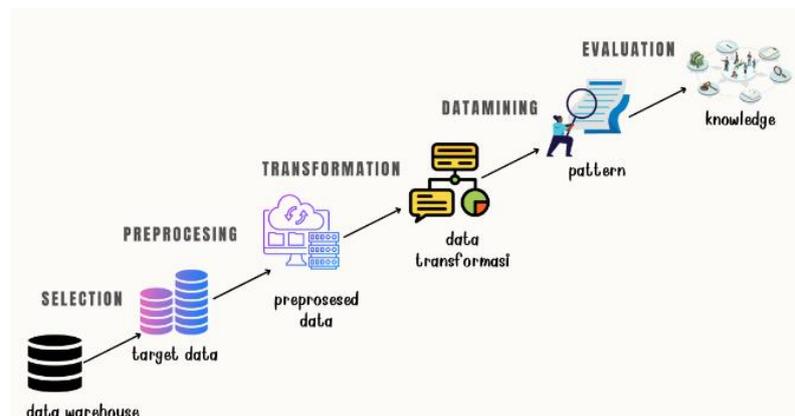
$$\mu_k = \frac{1}{N_k} \sum_{q=1}^{N_k} x_q \quad (2)$$

Dimana μ_k = Titik *centroid* dari *cluster* ke-*K*, N_k = Banyaknya data pada *cluster* ke-*K*, x_q = Data ke-*q* pada *cluster* ke-*k*

6. Mengecek kualitas *cluster* yang dihasilkan oleh algoritma *K-Means* serta melakukan interpretasi hasil

2.1 Teknik Analisis Data

Teknik analisis data pada penelitian menggunakan teknik KDD (*Knowledge Discovery in Databases*). KDD merupakan proses yang melibatkan ekstraksi informasi berguna, yang sebelumnya tidak diketahui, dan berpotensi berharga dari kumpulan data besar. Melibatkan langkah-langkah seperti pada gambar 2.



Gambar 2. Tahapan KDD

1. *Selection*, Proses mengubah data terpilih ke dalam bentuk prosedur.
2. *Data Transformation*, Proses transformasi data terpilih ke dalam bentuk *mining procedure*.
3. *Data Mining*, Proses mengekstraksi pola laten dan menghasilkan data berguna menggunakan berbagai teknik.
4. *Evaluation*, Proses dimana pola-pola yang telah diidentifikasi berdasarkan ukuran yang diberikan berubah seiring waktu.
 - a. Jarak *cluster* dengan titik pusat
 Jarak *cluster* dengan titik pusat adalah ukuran jarak antara setiap *cluster* dengan pusatnya. Pusat *cluster* di *K-means* disebut *centroid*, dan penghitungan jarak ini akan mempengaruhi pembentukan *cluster* yang optimal.
 - b. Nilai DBI
 DBI dihitung dengan mempertimbangkan nilai rata-rata jarak antara masing-masing *cluster* dengan *cluster* lainnya. Metrik ini mengukur kemiripan *cluster* dalam hal bentuk dan ukuran Rumusnya adalah:

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (3)$$
 Dari rumus tersebut, *k* merupakan jumlah *cluster* yang digunakan. Semakin kecil nilai DBI yang dihasilkan (non-negatif ≥ 0), maka semakin baik *cluster* hasil pengelompokan *K-means* yang digunakan.
5. *Knowledge Presentation*, Proses paling akhir dari proses KDD, Data-data yang sudah diproses divisualisasikan agar lebih mudah dipahami oleh pengguna dan diharapkan bisa diambil tindakan berdasarkan analisis.

2.2. Data

Data adalah kumpulan informasi dasar tentang sesuatu yang diperoleh dari pengamatan dan bisa diolah jadi bentuk menjadi sebuah informasi, database atau solusi dari suatu masalah tertentu[15]. Data yang digunakan dalam penelitian ini adalah data sebanyak 214 responden dan terdiri 16 atribut yang dikumpulkan melalui kuesioner menggunakan google formulir. Data hasil pengisian kuesioner tersebut dalam bentuk dokumen *soft file* format *Microsoft Excel*. Pertanyaan kuesioner pada penelitian ini dapat diakses melalui link berikut ini <https://bit.ly/3tqYoql>. Berikut ditampilkan data hasil pengisian kuesioner pada tabel 1.

Tabel 1. Hasil kuesioner gaya hidup mahasiswa

No	Nama	Jenis Kelamin	Usia	P ₁	P ₂	P ₃	P ₄	P ₅	P ₆	P ₇	P ₈	P ₉	P ₁₀	P ₁₁	P ₁₂	P ₁₃
1	Mela	Perempuan	18-25 tahun	3	3	3	4	4	3	2	2	3	1	2	1	Ya

2	Salsa	Perempuan	18-25 tahun	3	4	3	5	5	3	5	3	5	3	2	3	Ya
3	Hajarah	Perempuan	18-25 tahun	3	4	3	4	4	3	3	3	1	1	1	1	Ya
...
213	Riska	Perempuan	18-25 tahun	2	2	3	2	5	3	5	2	4	1	2	3	Ya
214	Trian	Laki-laki	18-25 tahun	2	3	3	1	5	3	3	1	4	1	3	2	Ya

2.3. Data Cleansing

Data cleansing proses ini berfokus pada identifikasi dan pengecekan masalah yang mungkin ada dalam data mentah sehingga data yang akan diolah nantinya dapat menghasilkan informasi berkualitas. Pada penelitian ini pembersihan data dilakukan mencakup responden yang mengisi kuesioner lebih dari 1 kali, ketidaklengkapan, ketidakbenaran serta ketidakrelevan data. Hasil pembersihan data dapat dilihat pada tabel 2.

Tabel 2. Data Cleansing

No	Nama	Jenis Kelamin	Usia	P_1	P_2	P_3	P_4	P_5	P_6	P_7	P_8	P_9	P_{10}	P_{11}	P_{12}	P_{13}
1	Mela	Perempuan	18-25 tahun	3	3	3	4	4	3	2	2	3	1	2	1	Ya
2	Salsa	Perempuan	18-25 tahun	3	4	3	5	5	3	5	3	5	3	2	3	Ya
...
205	Riska	Perempuan	18-25 tahun	2	2	3	2	5	3	5	2	4	1	2	3	Ya
206	Trian	Laki-laki	18-25 tahun	2	3	3	1	5	3	3	1	4	1	3	2	Ya

2.4. Data Selection

Data yang dipakai pada penelitian ini adalah data kuesioner gaya hidup mahasiswa. Yang sebelumnya telah disebarkan kepada responden dan dari 16 atribut yang ada pada kuesioner penelitian tersebut, dilakukan proses *selection* hasil dari data kuesioner yang dilakukan digunakan 14 atribut. Karena 2 atribut memiliki nilai yang tidak relevan untuk dilakukan analisa lebih lanjut. Atribut yang tidak digunakan terdiri dari usia dan P_{13} . Hasil pemilihan atribut dapat dilihat pada tabel 3 berikut.

Tabel 3. Data Selection

No	Nama Atribut	Tipe
1	Nama	polynomial
2	Jenis kelamin	polynomial

3	P_1	integer
4	P_2	integer
5	P_3	integer
6	P_4	integer
7	P_5	integer
8	P_6	integer
9	P_7	integer
10	P_8	integer
11	P_9	integer
12	P_{10}	integer
13	P_{11}	integer
14	P_{12}	integer

2.5. *Data Transformation*

Pada tahap ini transformasi data dilakukan pada atribut nama yang dijadikan sebagai ID serta atribut jenis kelamin akan diubah menjadi numerik. Atribut kelamin akan diinisialisasikan menjadi numerik seperti dalam tabel 4 berikut.

Tabel 1. Tranformasi atribut jenis kelamin

Jenis Kelamin	inisialisasi
Laki-laki	0
Perempuan	1

untuk merubah data bertipe nominal menjadi numerik dapat dilakukan menggunakan operator *Nominal to Numeric*. Proses ini dapat dilihat pada gambar 3.

Row No.	Nama	Jenis kela...	Apakah and...	Apakah				
1	Abdul	0	3	4	4	4	2	2
2	Ahmad	0	4	3	3	3	3	2
3	Adi	0	3	4	5	3	5	4
4	Adi Pangestu	0	5	3	4	4	2	3
5	ADI SADEWA	0	2	4	4	5	3	3
6	Adi Sudrajat	0	4	4	3	2	4	4
7	Adinda	1	4	4	5	4	3	4
8	Adisty	1	3	1	3	1	4	4
9	Aditia	0	5	5	3	3	4	4
10	Adrian	0	5	5	4	4	5	4
11	Agustin	1	4	5	4	5	4	4
12	AJENG	1	4	4	2	3	5	2
13	Ajeng Galuh ...	1	3	3	3	4	5	5
14	Aji saputra	0	3	4	4	1	3	3

Gambar 1. Data transformation

2.6. Data Mining

Pada proses data mining penelitian ini menggunakan *algoritma K-means clustering*. *Algoritma* ini digunakan pada proses segmentasi, mengetahui jumlah *cluster* yang dihasilkan serta menghitung jarak antar *cluster* dengan pusat *cluster* dan mengukur besarnya *Davies Bouldin index* (DBI) menggunakan operator *cluster distance performance*. Rumus yang digunakan untuk perhitungannya adalah

$$d_{Euclidean}(x, y) = \sqrt{\sum_i^n (x_i - y_i)^2} \tag{1}$$

Parameter untuk mengukur besarnya *Davies Bouldin index* (DBI) sesuai pada tabel 5.

Tabel 2. Parameter

Parameter	Keterangan
K-min	2
Max runs	10
Measure type	NumericalMeasurement
Numerical measure	EuclideanDistance
Clustering algorithm	K-means clustering
Max Optimization Steps	100
Main criterion	Davies bouldin

Setelah pengisian parameter, dilakukan 9 kali percobaan untuk mencari *cluster* terbaik dan paling optimal. Percobaan *cluster* yang dilakukan secara berulang membantu memastikan bahwa hasil segmentasi yang diambil adalah hasil yang paling stabil dan sesuai dengan karakteristik data. Diperoleh hasil *cluster* optimal yaitu 2 *cluster*, dimana *cluster* 0 sebanyak 51 items dan *cluster* 1 sebanyak 155 items sesuai pada gambar 4.

Cluster Model

```
Cluster 0: 51 items  
Cluster 1: 155 items  
Total number of items: 206
```

Gambar 2. *Cluster model*

Selanjutnya nilai *Davies Bouldin Index* (DBI) yang dihasilkan dari algoritma *K-means clustering* ini sebesar 0.118 sesuai dengan gambar 5.

PerformanceVector

```
PerformanceVector:  
Avg. within centroid distance: 0.815  
Avg. within centroid distance_cluster_0: 0.925  
Avg. within centroid distance_cluster_1: 0.779  
Davies Bouldin: 0.118
```

Gambar 3. *Performance Vector K=2*

Hasil analisis data menjelaskan bahwa *cluster* 0 dengan jumlah anggota sebanyak 51 mahasiswa masuk kedalam kategori gaya hidup sehat dengan rincian laki-laki sebanyak 19 mahasiswa dan perempuan 32 mahasiswa. Pada *cluster* 1 dengan jumlah anggota sebanyak 155 mahasiswa masuk kedalam kategori gaya hidup tidak sehat dengan rincian laki-laki 57 mahasiswa dan perempuan 98 mahasiswa.

2.7. Evaluation

Metode evaluasi yang digunakan dalam penelitian ini adalah *Davies Bouldin Index* (DBI) dengan rumus

$$DBI = \frac{1}{K} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (3)$$

Dimana *cluster* dengan nilai yang mendekati angka nol merupakan *cluster* terbaik. DBI merupakan salah satu metode yang dikenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979 untuk mengevaluasi *cluster*, dimana *cluster* dengan nilai yang mendekati angka nol merupakan *cluster* terbaik [16]. Dalam table 6 terdapat hasil evaluasi menggunakan DBI dengan jumlah percobaan segmentasi data sebanyak 9 kali.

Tabel 6. Hasil evaluasi menggunakan DBI

Cluster	Hasil nilai DBI
K=2	0.118
K=3	0.170
K=4	0.164
K=5	0.178
K=6	0.162
K=7	0.168
K=8	0.167

K=9	0.167
K=10	0.150

Dari tabel 6 disimpulkan bahwa K=2 adalah cluster terbaik. *Cluster 0*: Merupakan mahasiswa dengan kategori gaya hidup sehat yang berjumlah 51 mahasiswa. *Cluster 1*: Merupakan mahasiswa dengan kategori gaya hidup tidak sehat yang berjumlah 155 mahasiswa.

3. Result and Discussion

3.1. Result

Pembahasan ini merupakan interpretasi hasil yang berfokus pada K=2 sebagai *cluster* terbaik, kita dapat melihat dengan jelas perbedaan antara kedua kelompok mahasiswa yang dihasilkan. Persentase persebaran mahasiswa berdasarkan jenis kelamin dapat dilihat pada gambar 6.



Gambar 6. Persebaran mahasiswa berdasarkan jenis kelamin

Hasil *Performance* dari percobaan menggunakan nilai parameter sesuai pada tabel 5, percobaan pertama dilakukan pada K=2 dimana *cluster* ini menghasilkan rata-rata jarak pusat *cluster* sebesar 0.815, rata-rata jarak pada *cluster 0* adalah 0.779, jarak rata-rata pada *cluster 1* adalah 0.925. Sedangkan hasil evaluasi menggunakan metode DBI memperoleh hasil 0.118 dengan hasil evaluasi ini K=2 menjadi *cluster* yang paling optimal karena nilainya yang paling mendekati 0 dibanding *cluster* lainnya. Hasil *Performance* K=2 bisa dilihat pada gambar 7.

PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: 0.815
Avg. within centroid distance_cluster_0: 0.779
Avg. within centroid distance_cluster_1: 0.925
Davies Bouldin: 0.118
```

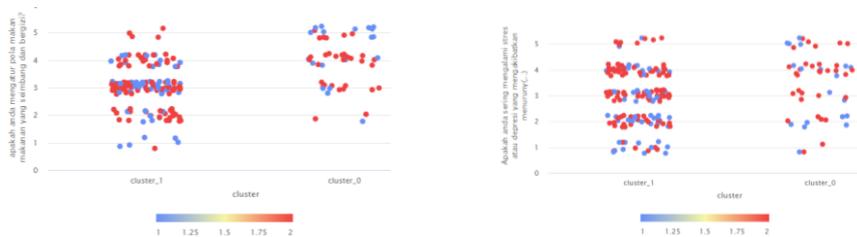
Gambar 7. Performance K=2

Hasil dari visualisasi *cluster* berdasarkan seberapa sering berolahraga (P_1) dan melakukan begadang atau tidur saat larut malam (P_2) bisa dilihat pada gambar 8.



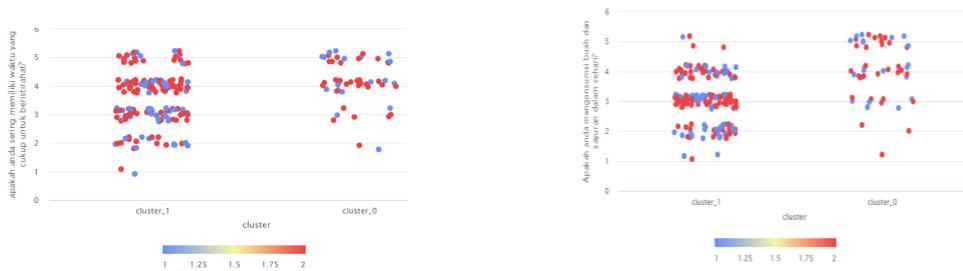
Gambar 8. Visualisasi (P_1) dan (P_2)

Hasil dari visualisasi *cluster* model berdasarkan makan makanan yang seimbang dan bergizi bisa (P_3) dan mengalami stres atau depresi yang mengakibatkan menurunnya kondisi fisik (P_4) dilihat pada gambar 9.



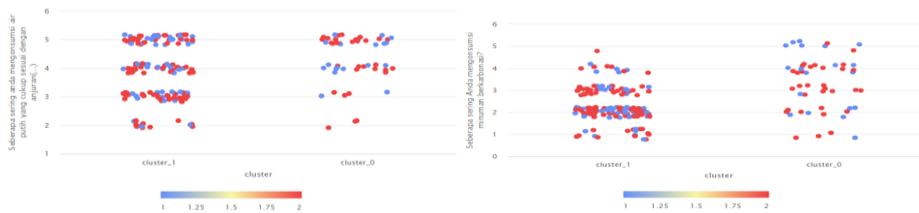
Gambar 9. Visualisasi (P_3) dan (P_4)

Hasil dari visualisasi berdasarkan memiliki waktu yang cukup untuk beristirahat (P_5) dan mengonsumsi buah dan sayuran (P_6) bisa dilihat pada gambar 10.



Gambar 10. Visualisasi (P_5) dan (P_6)

Hasil dari visualisasi berdasarkan mengonsumsi air putih yang cukup (P_7) dan mengonsumsi minuman berkarbonasi (P_8) pada gambar 11.



Gambar 11. Visualisasi (P_7) dan (P_8)

Hasil dari visualisasi mencari informasi mengenai cara menjaga tubuh agar tetap sehat (P_9) dan mengikuti seminar atau kelas untuk menjaga agar tubuh tetap sehat (P_{10}) pada gambar 12.



Gambar 12. Visualisasi (P_9) dan (P_{10})

Hasil visualisasi berdasar mengikuti program olahraga atau aktivitas fisik bersama dengan mahasiswa lain (P_{11}) dan berpartisipasi dalam kampanye atau acara yang berhubungan dengan kesehatan (P_{12}) pada gambar 13.



Gambar 13. Visualisasi (P_{11}) dan (P_{12})

Setelah mengetahui hasil visualisasi dari masing-masing atribut pada penelitian ini ditemukan bahwa perbedaan jelas antara *cluster 0* dan *cluster 1*. Dari visualisasi persebaran *cluster* data gaya hidup mahasiswa, terlihat bahwa sebagian besar mahasiswa cenderung termasuk dalam *cluster 1*. Ini mengindikasikan bahwa mayoritas dari mereka menganut gaya hidup yang tidak sehat. Ada kekurangan dalam aktivitas fisik, kurangnya pola makan yang teratur, kebiasaan tidur larut malam, dan tingginya tingkat depresi serta stres. Situasi ini memprihatinkan karena seharusnya mahasiswa dapat menjaga gaya hidup agar tetap sehat dan bugar, sehingga dapat mencapai prestasi optimal dalam perkuliahan.

3.2. Discussion

Berdasarkan persebaran *cluster* yang divisualisasikan dari data penelitian ini terlihat jelas *cluster* yang dominan berada pada *cluster 1*. Ini menunjukkan bahwa mayoritas mahasiswa menerapkan gaya hidup tidak sehat. Berikut adalah beberapa kegiatan inisiasi kampus yang dapat diterapkan untuk mendukung kesehatan mahasiswanya:

1. Promosi Gaya Hidup Sehat: Mengorganisir kampanye dan acara promosi gaya hidup sehat, termasuk olahraga rutin, seminar nutrisi, dan workshop kesehatan mental.
2. Fasilitas Olahraga: Meningkatkan fasilitas olahraga di kampus untuk mendorong partisipasi dalam aktivitas fisik. Lapangan olahraga atau area terbuka dapat menjadi tempat yang menarik bagi mahasiswa.
3. Klinik Kesehatan Mahasiswa: Menyediakan klinik kesehatan khusus mahasiswa di kampus, yang dapat memberikan layanan kesehatan umum, konseling, dan dukungan kesehatan mental.
4. Kantin Sehat: Menyediakan pilihan makanan sehat dan bergizi ini dapat membantu menciptakan lingkungan yang mendukung gaya hidup sehat sekaligus membentuk kebiasaan makan yang baik di kalangan mahasiswa.
5. Edukasi Kesehatan Mental: Membangun kesadaran tentang kesehatan mental melalui seminar, lokakarya dan kampanye informasi. Menyediakan layanan konseling dan dukungan mental untuk mahasiswa yang membutuhkan.
6. Komunitas Kesehatan: Mendorong pembentukan komunitas kesehatan di kampus yang saling mendukung. Kelompok atau klub yang fokus pada kesehatan dapat menjadi tempat dimana mahasiswa dapat berbagi pengalaman dan pengetahuan.

4. Conclusion

Hasil penelitian segmentasi gaya hidup mahasiswa menggunakan algoritma *K-means clustering* maka hasilnya dapat disimpulkan sebagai berikut:

1. Dalam penelitian ini, ditemukan bahwa *cluster* terbaik dan paling optimal adalah $K=2$. Hasil ini mencerminkan kecenderungan bahwa mahasiswa dapat dikelompokkan menjadi dua kategori utama berdasarkan gaya hidup mereka.
2. Rata-rata jarak yang dihasilkan pada $K=2$, rata-rata jarak dalam pusat *cluster* sebesar 0.815, menunjukkan sejauh mana mahasiswa dalam suatu kelompok memiliki kesamaan pola gaya hidup sehat. Selain itu, analisis rata-rata jarak dalam pusat *cluster 0* yaitu 0.779 dan *cluster 1* yaitu 0.925 memberikan informasi spesifik tentang karakteristik masing-masing *cluster*. Ini membantu mengidentifikasi gaya hidup yang dominan di setiap

klaster. Pada hasil analisis data menjelaskan bahwa *cluster* 0 dengan jumlah anggota sebanyak 51 mahasiswa atau 24.76% masuk kedalam kategori gaya hidup sehat. Pada *cluster* 1 dengan jumlah anggota sebanyak 155 mahasiswa atau 75.24% masuk kedalam kategori gaya hidup tidak.

3. Penentuan *cluster* terbaik menggunakan *Davies Bouldin Index* (DBI) memberikan pemahaman lebih lanjut tentang validitas dan kehomogenan *cluster*. Pada $K=2$, DBI mendekati nilai 0, yaitu 0.118. Nilai yang mendekati nol menandakan bahwa *cluster* tersebut kompak dan terpisah dengan baik. Oleh karena itu, hasil ini mendukung pemilihan $K=2$ sebagai konfigurasi optimal dalam segmentasi mahasiswa berdasarkan gaya hidup sehat. Dengan demikian, analisis ini memberikan kerangka yang kuat untuk memahami dan meningkatkan kesadaran kesehatan di kalangan mahasiswa.

References

- [1] A. E. Saputri, "HUBUNGAN ANTARA KESEPIAN DENGAN STRES AKADEMIK PADA MAHASISWA RANTAU S1 TAHUN PERTAMA DI UNIVERSITAS ISLAM SULTAN AGUNG SEMARANG," no. 30701900019, 2023.
- [2] A. M. Putri, "GAYA HIDUP SEHAT DAN SUBJECTIVE WELL BEING PADA MAHASISWA KEDOKTERAN," vol. 7, no. 2, pp. 635–642, 2023.
- [3] A. Ilhamsyah, "Jurnal Masyarakat Sehat Indonesia (JMSI)," *J. Masy. Sehat Indones.* 1-4, vol. 012, pp. 1–4, 2023.
- [4] N. S. Azizah, "Pengaruh literasi keuangan, gaya hidup pada perilaku keuangan pada generasi milenial," vol. 01, pp. 92–101, 2020.
- [5] T. Vebriyani, "Pengaruh Media Sosial dan Gaya Hidup Hedonis terhadap Perilaku Konsumtif Mahasiswa (Studi Kasus pada Mahasiswa Fakultas Ekonomi dan Bisnis Islam IAIN Kudus Angkatan 2016)," pp. 11–31, 2021.
- [6] H. Priyatman, F. Sajid, and D. Haldivany, "Klasterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa," pp. 62–66, 2019.
- [7] M. G. Sadewo, A. Eriza, A. P. Windarto, and D. Hartama, "Algoritma K-Means Dalam Mengelompokkan Desa / Kelurahan Menurut Keberadaan Keluarga Pengguna Listrik dan Sumber Penerangan Jalan Utama Berdasarkan Provinsi," pp. 754–761, 2019.
- [8] A. Aranta, "ANALISIS PEMILIHAN CLUSTER OPTIMAL DALAM SEGMENTASI PELANGGAN TOKO RETAIL," vol. 18, no. 2, pp. 152–163, 2021.
- [9] N. H. Ahsina, F. Fatimah, and F. Rachmawati, "ANALISIS SEGMENTASI PELANGGAN BANK BERDASARKAN PENGAMBILAN KREDIT DENGAN MENGGUNAKAN METODE K-MEANS CLUSTERING," vol. 8, no. 3, 2022.
- [10] misra hartati siti monalisa, tengku nurainun, "PENERAPAN ALGORITMA K-MEANS DAN METODE MARKETING MIX DALAM SEGMENTASI MAHASISWA DAN STRATEGI PEMASARAN," 2021.
- [11] Y. Suhanda, I. Kurniati, and S. Norma, "Penerapan Metode Crisp-DM Dengan Algoritma K-Means Clustering Untuk Segmentasi Mahasiswa Berdasarkan Kualitas Akademik," *J. Teknol. Inform. dan Komput.*, vol. 6, no. 2, pp. 12–20, 2020, doi: 10.37012/jtik.v6i2.299.
- [12] S. A. Perdana, S. F. Florentin, and A. Santoso, "ANALISIS SEGMENTASI PELANGGAN MENGGUNAKAN K-MEANS CLUSTERING STUDI KASUS APLIKASI ALFAGIFT," *Sebatik*, 2022, [Online]. Available: <https://jurnal.wicida.ac.id/index.php/sebatik/article/view/1991>
- [13] L. A. A. R. P. A. A Sagung Prami Apsari Kumala, "MEASURE COMPARISON DISTANCE ON K-MEANS CLUSTERING FOR GROUPING MUSIC ON MOOD," *J. Elektron. Ilmu Komput. Udayana*, vol. 11, no. 4, pp. 663–674, 2023.
- [14] M. H. Adiya and Y. Desnelita, "Jurnal Nasional Teknologi dan Sistem Informasi Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan Pada RSUD Pekanbaru," vol. 01, pp. 17–24, 2019.
- [15] F. Handayani, "Aplikasi Data Mining Menggunakan Algoritma K-Means Clustering untuk Mengelompokkan Mahasiswa Berdasarkan Gaya Belajar," vol. 12, 2022, doi: 10.34010/jati.v12i1.
- [16] Z. Nabila, "ANALISIS DATA MINING UNTUK CLUSTERING KASUS COVID-19 DI PROVINSI LAMPUNG DENGAN ALGORITMA K-MEANS," vol. 2, no. 2, pp. 100–108, 2021.