

Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Data Pasien Rehabilitasi Narkoba

Ega Yolanda*, Suhardi

Fakultas Sains dan Teknologi, Program Studi Ilmu Komputer, Universitas Islam Negeri Sumatera Utara, Medan, Indonesia

Email: ^{1,*}yolandaega29@gmail.com, ²suhardi@uinsu.ac.id

Email Penulis Korespondensi: yolandaega29@gmail.com

Abstrak—Penyalahgunaan narkoba merupakan permasalahan serius yang menyebabkan ketergantungan dan efek berbahaya. Rehabilitasi merupakan upaya untuk menyelamatkan pecandu narkoba agar bisa hidup normal dan sehat secara fisik dan mental, serta meningkatkan keterampilan dan hubungan sosial. Badan Narkotika Nasional Provinsi Sumatera Utara bertugas melakukan pencegahan, pemberantasan, dan rehabilitasi narkoba. Ada perbedaan program rehabilitasi antara remaja dan dewasa, yaitu menerapkan program "parenting" bagi remaja. Namun, proses penentuan program masih manual dan kurang efisien. Penelitian ini menggunakan data mining dengan algoritma K-Means clustering untuk mengelompokkan data pasien rehabilitasi narkoba secara efisien. Hasil clustering menghasilkan tiga kelompok pasien berdasarkan karakteristik mereka, yang memberikan informasi penting bagi Badan Narkotika Nasional Provinsi Sumatera Utara untuk memberikan program rehabilitasi yang lebih sesuai dengan kebutuhan kelompok pasien. Melalui proses pengelompokan data, pola-pola kelompok pengguna narkoba dapat diidentifikasi berdasarkan karakteristik yang mereka miliki. Dengan demikian, pihak Badan Narkotika Nasional Provinsi Sumatera Utara dapat menentukan program yang lebih efektif dan sesuai dengan kebutuhan masing-masing kelompok. Berdasarkan hasil cluster yang telah didapat, bahwa program parenting dapat diterapkan pada kelompok cluster dua. Dari hasil penelitian, disimpulkan bahwa algoritma K-Means clustering menggunakan Python dan tools Jupyter Notebook dapat memberikan hasil clustering yang optimal. Penelitian ini dapat menjadi dasar untuk pengembangan aplikasi dan penelitian lebih lanjut, serta perbandingan dengan algoritma clustering lainnya dalam pengelompokan data pasien rehabilitasi narkoba.

Kata Kunci: Clustering; Data Mining; Narkoba; Rehabilitasi; Pasien; Python; Jupyter Notebook

Abstract—Drug abuse's a serious problem leading to addiction and harmful effects. Rehabilitation aims to save drug addicts and help them lead a normal, physically, and mentally healthy life while improving skills and social relationships. The North Sumatra Province National Narcotics Agency's responsible for drug prevention, eradication, and rehabilitation. There're different rehabilitation programs for adolescents and adults, with a "parenting" program applied for adolescents. However, the manual and inefficient process of determining programs poses challenges. This research utilizes data mining with the K-Means clustering algorithm to efficiently categorize drug rehabilitation patient data. The clustering results in three patient clusters based on their characteristics, providing essential information for North Sumatra Province National Narcotics Agency to tailor rehabilitation programs to each group's needs. Through the data clustering process, drug user patterns can be identified based on their shared attributes. Consequently, The North Sumatra Province National Narcotics Agency can determine more effective and suitable programs for each cluster. The findings show that the parenting program is appropriate for Cluster two. The study concludes that using the K-Means clustering algorithm with Python and Jupyter Notebook tools yields optimal clustering results. This research serves as a foundation for application development, further investigations, and comparisons with other clustering algorithms in drug rehabilitation patient data grouping.

Keywords: Clustering; Data Mining; Drugs; Rehabilitation; Patient; Python; Jupyter Notebooks

1. PENDAHULUAN

Narkoba adalah zat yang jika dikonsumsi secara tidak wajar justru akan menimbulkan ketergantungan dan berbagai kondisi medis. Nilai penggunaan narkoba kronis saat ini berkembang secara bertahap [1]. Seperti yang mungkin kita ketahui, penggunaan narkoba yang kronis dapat memiliki efek yang sangat berbahaya [2]. Rehabilitasi merupakan salah satu upaya para ahli untuk mengalahkan penyalahgunaan narkoba [3]. Rehabilitasi dilakukan untuk menyelamatkan pecandu narkoba dari ketergantungan narkoba dan hidup normal, sehat jasmani dan rohani, sehingga mereka dapat menyesuaikan dan meningkatkan keterampilan, pengetahuan, kecerdasan dan hubungan sosial di lingkungan atau dengan keluarganya.

Badan Narkotika Nasional Provinsi Sumatera Utara adalah unit yang didedikasikan untuk pencegahan, pemberantasan, dan pemberantasan penyalahgunaan dan rehabilitasi pecandu narkoba. Ada perbedaan program yang diberikan oleh BNNP Sumut bagi pasien rehabilitasi remaja atau anak-anak dan dewasa, yaitu program *parenting* yang diberikan untuk pasien rehabilitasi remaja dan anak-anak. Dimana para staf di BNN berperan sebagai orang tua pengganti untuk membantu para remaja selama menjalani program rehabilitasi. Penentuan program yang masih manual seperti harus melakukan wawancara terlebih dahulu, kemudian menentukan SOP penanganannya dan kemudian menentukan program yang akan dijalani oleh pasien, dinilai masih kurang efisien dikarenakan membuang banyak waktu sedangkan pasien perlu segera di beri penanganan.

Data mining sebagai *knowledge discovery in database* (KDD) yang tujuan utamanya adalah membuat dan mengeksplorasi data atau data yang dimilikinya [4]. Salah satu teknik yang digunakan pada data mining adalah *clustering* [5].

Proses pengorganisasian sesuatu ke dalam kelompok (*cluster*) yang dalam kelompoknya memiliki beberapa karakteristik yang sama dikenal dengan *clustering* [6]. *Clustering* bertujuan untuk membuat objek menjadi satu kelompok dengan kemiripan data [7].

Pengelompokan prosedur informasi adalah dasar dari algoritma *K-Means non-hierarkis* [8]. Algoritma *K-Means* dimulai dengan membentuk *cluster* pada awalnya dan kemudian secara iteratif mengoreksi *cluster* tersebut sampai tidak ada perubahan yang signifikan pada *cluster* tersebut. [9].

Mempersiapkan informasi siap pakai untuk digunakan adalah tujuan dari *preprocessing* data [10]. Adapun tahapan dari *preprocessing* data yaitu yaitu transformasi data, pada fase ini dilakukan penyesuaian format data dan pemilihan data, pada fase ini dipilih atribut data yang dianggap penting berkaitan dengan rumusan masalah [11].

Python adalah *scripting language* yang berorientasi objek [12]. *Python* mendukung modul dan paket yang memfasilitasi modularitas program [13]. *Jupyter Notebooks* adalah format rilis untuk alur yang dapat dihitung yang dapat disalin/dijalankan orang lain [14]. *Jupyter Notebook* adalah situs aplikasi yang berjalan localhost [15].

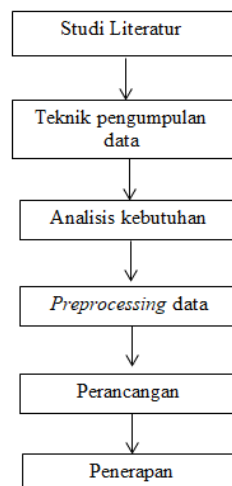
Telah banyak Penelitian terdahulu terkait dengan penelitian yang akan datang, termasuk penelitian dengan judul “Penerapan Algoritma *K-Means Clustering* Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau” tahun 2021 [16], hasil akhir yang didapat pengelompokan penyakit ISPA kedalam 2 *cluster*, sedangkan peneliti melakukan pengelompokan terhadap data pasien rehabilitasi narkoba kedalam 3 *cluster*. Penelitian selanjutnya dengan judul “Data Mining Pengelompokan Pecandu Narkoba Di Masyarakat Berdasarkan Faktor Penyebab Penggunaannya Menggunakan Metode *K-Means Clustering*” [17], hasil akhir yang didapat yaitu algoritma *K-Means clustering* mengelompokan pecandu narkoba di masyarakat dan menghasilkan 3 *centroid*, sedangkan peneliti mengelompokan data pasien yang akan melakukan rehabilitasi narkoba . Penelitian lain dengan judul “Menentukan *Cluster* Yang Tepat Dengan *K-Means* Dalam Rangka Mengukur Efektivitas Pelaksanaan Anggaran Pada Kementerian Agraria Dan Tata Ruang/Badan Pertahanan Nasional [18], hasil yang didapat Menemukan model klaster untuk mengukur efektivitas pelaksanaan anggaran dan dapat digunakan untuk menentukan alokasi anggaran dan menghindari pemborosan dana di anggaran Kementerian Pertanian dan Tata Ruang, sedangkan peneliti akan memodelkan *cluster* data pasien rehabilitasi narkoba menggunakan *Python*. Adapun penelitian lain dengan judul “Implementasi Metode *K-Means* Dalam Pengklasteran Daerah Pungutan Liar di Kabupaten Sukabumi (Studi Kasus : Dinas Kependudukan Dan Pencatatan Sipil)” [19], dengan hasil akhir mendapatkan tingkat pungutan liar di setiap kecamatan menggunakan metode *clustering*, sedangkan peneliti akan menghasilkan jumlah anggota pada masing-masing *cluster* menggunakan *clustering*. Penelitian lain dengan judul “Implementasi *K-Means Clustering* untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model *CRISP-DM*” [20], dengan hasil akhir mendapatkan nilai tertinggi untuk *cluster* daerah rawan bencana, sedangkan peneliti akan menghasilkan jumlah anggota pada masing-masing *cluster* menggunakan *clustering* .

Hasil akhir dari penelitian ini membahas tentang pengelompokan data pasien rehabilitasi narkoba yang di dapatkan dari sumber informasi Badan Narkotika Nasional Provinsi Sumatera Utara (BNNP Sumut) dengan mengambil data pada tahun 2022. Berdasarkan permasalahan yang terjadi bahwa peneliti ingin melakukan sebuah analisa data dari data resmi dari BNNP Sumut untuk dapat diolah menggunakan *K-Means clustering*. Tujuan penelitian ini untuk menggunakan algoritma *K-Means clustering* terhadap data pasien rehabilitasi narkoba dan dapat menghasilkan *cluster* yang optimal sehingga mempermudah pihak BNN untuk memberikan program rehabilitasi yang lebih efisien dan tepat sasaran.

2. METODOLOGI PENELITIAN

2.2 Kerangka Penelitian

Kerangka penelitian adalah struktur dasar atau kerangka acuan yang digunakan dalam penelitian untuk mengatur dan merangkai langkah-langkah proses penelitian. Kerangka penelitian membantu peneliti merancang metode, mengumpulkan data, menganalisis data, dan mempresentasikan hasil penelitian. Tahap penelitian ini terlihat pada gambar 1 yang terdiri dari studi literatur, pengumpulan data, analisis kebutuhan, *preprocessing* data, perancangan, dan penerapan.



Gambar 1. Tahap-Tahap Kerangka Penelitian

Adapun penjelasan dari gambar tahap-tahap kerangka penelitian diatas adalah sebagai berikut:

2.2.1 Studi Literatur

Tujuan penelitian kepustakaan adalah untuk mengumpulkan informasi dari berbagai sumber keilmuan, seperti buku, jurnal, dan karya lain yang berkaitan dengan fokus penulis..

2.2.2 Teknik Pengumpulan Data

Teknik pengumpulan data yang diuji dalam penelitian ini adalah:

a. Wawancara

Wawancara akan dilakukan dengan ibu Rolika Perawati Silalahi selaku Konselor Bidang Rehabilitasi. Wawancara dengan pihak BNN dapat dilakukan untuk mengumpulkan data demografi, riwayat penggunaan narkoba dan informasi yang berkaitan dengan penelitian.

b. Rekam Medis

Rekam medis akan diambil di bidang rehabilitasi. Rekam medis dapat menjadi sumber data penting untuk mendapatkan informasi mengenai riwayat penggunaan narkoba, kondisi kesehatan pasien dan hasil tes medis yang dilakukan selama masa rehabilitasi.

2.2.3 Analisis Kebutuhan

Dalam melakukan penelitian ini dibutuhkan beberapa analisis kebutuhan agar dapat terlaksana dengan baik. Dalam penelitian ini, data pasien rehabilitasi narkoba menjadi fokus utama. Data yang diperlukan meliputi profil pasien, jenis narkoba yang digunakan, lama penggunaan, dan usia.

Selain data pasien, algoritma *K-Means clustering* juga menjadi alat utama dalam penelitian ini. Diperlukan pemahaman yang baik tentang cara kerja algoritma *K-Means clustering* dan kemampuan untuk mengimplementasikan dalam bahasa pemrograman tertentu. Bahasa pemrograman yang dimaksud adalah bahasa pemrograman *Python*.

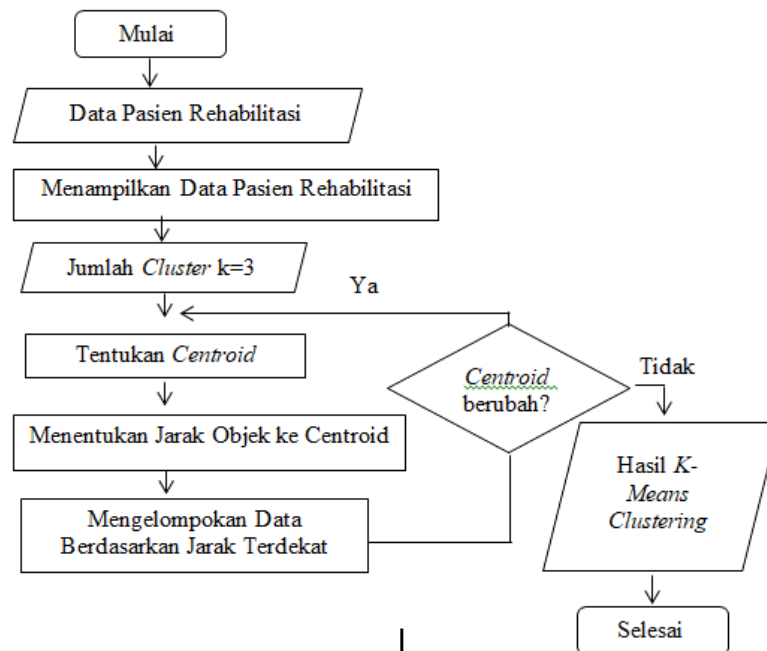
2.2.4 Preprocessing Data

Setelah pengumpulan data dan analisis kebutuhan, kemudian dilakukan tahap *preprocessing* data. Ini terdiri dari beberapa fase: transformasi data dan pemilihan data.

2.2.5 Perancangan

Sehubungan dengan pengujian yang telah dilakukan, peneliti perlu melacak data untuk dikumpulkan dengan langkah-langkah yang telah ditentukan sebelumnya.

Ada juga patokan yang didapat dari penelitian ini, yaitu umur, jenis narkoba yang digunakan dan lama penggunaannya. Data disajikan ke dalam *Microsoft Excel* dan dikelola menggunakan *Python* dengan *tools Jupyter Notebook*. Ada juga *Flowchart* yang sudah direncanakan seperti yang ditampilkan di bawah ini:



Gambar2 Flowchart K-Means Clustering

Keterangan :

1. Masukkan detail pasien rehabilitasi.

2. Menampilkan informasi tentang pasien rehabilitasi.
3. Masukkan jumlah *cluster* $k = 3$.
4. Tentukan pusat *cluster*.
5. Tentukan jarak objek ke pusat *cluster* menggunakan jarak *Eulidean Distance*.
6. Pengelompokan kelompok berdasarkan nilai terdekat.
7. Apakah nilai titik pusat berubah?
 - a. Jika ya, tentukan kembali titik pusatnya
 - b. Kalau tidak, hasilnya dihasilkan.
8. Output hasil perhitungan dengan *K-Means clustering*, selesai..

2.7 Penerapan Algoritma

K-Means clustering adalah teknik analisis data yang dapat diterapkan pada data rehabilitasi narkoba untuk mengklasifikasikan individu dalam rehabilitasi ke dalam beberapa kelompok berdasarkan kesamaan pola atau karakteristik perilaku. Menerapkan Algoritma *K-Means Clustering* pada data pasien rehabilitasi narkoba akan mengklasifikasikan pasien ke dalam kelompok yang sama berdasarkan variabel yaitu usia, jenis narkoba yang digunakan dan lama penggunaannya. Penerapan algoritma ini diharapkan informasi ini membantu untuk memahami karakteristik setiap kelompok baru dan dapat memberikan wawasan penting untuk pengambilan keputusan dalam program rehabilitasi narkoba, sehingga memungkinkan perumusan strategi dan rehabilitasi yang lebih efektif dan tepat sasaran.

3. HASIL DAN PEMBAHASAN

3.1 Implementasi Algoritma *K-Means Clustering*

Dalam menggunakan teknik pengelompokan, cara utama untuk membuat kelompok adalah mengubah data menjadi struktur numerik dengan menggunakan kode yang diberikan, tentukan jumlah grup (K), jumlah *centroid*, dan jarak total dari objek ke *centroid*, lalu hubungkan dengan jarak terpendek, dengan asumsi tidak ada kemungkinan pergerakan..

Untuk menentukan kelompok suatu benda, yang terpenting adalah mengukur jarak *Euclidean*nya ($X Y Z$), yang digambarkan sebagai berikut:

Table 1. Umur (X)

Kode	Umur(X)
1	<16 Tahun
2	17-25 Tahun
3	26-35Tahun
4	36-45Tahun
5	46-55Tahun
6	>55 Tahun

Table 2. Jenis Narkoba (Y)

Kode	JenisNarkoba (Y)
1	Sabu
2	Ganja
3	Inex
4	Extacy

Table 3. Lama Penggunaan(Z)

Kode	Lama Penggunaan(Z)
1	1 Bulan
2	2 Bulan
3	3 Bulan
4	4 Bulan
5	5 Bulan
6	> 5 Bulan

Penjelasan tabel pengkodean diatas adalah :

- a. Table 1 berisi data tentang variabel umur (X) dari subjek penelitian dalam penelitian. Tabel ini menyajikan data dalam bentuk kelompok umur, misalnya, 16-25 tahun, 26-45 tahun, dan 40-55 tahun.
- b. Table 2 berisi data tentang variabel jenis narkoba (Y) yang digunakan oleh subjek penelitian dalam penelitian. Variabel ini mencatat jenis narkoba yang dikonsumsi oleh setiap individu dalam penelitian. Jenis narkoba yang dicatat dalam tabel ini yaitu sabu, ganja, inex, dan *extacy*.

c. Table 3 berisi data tentang variabel lama penggunaan (Z) dari subjek penelitian dalam penelitian tertentu. Variabel ini mencatat durasi atau lama penggunaan narkoba oleh setiap individu dalam sampel penelitian. Tabel ini menyajikan informasi tentang berapa lama seseorang telah menjadi pengguna narkoba, misalnya, 1-5 bulan atau lebih dari 5 bulan.

Pengkodean berikut dilakukan dengan menggunakan model ini, yaitu memodifikasi untuk menentukan pengelompokan. Perubahan informasi di atas tampak dalam tabel di bawah ini.

Table 4. Transformasi Data

No	Inisial	X	Y	Z
1	R	3	1	4
2	AS	3	1	2
3	JN	3	1	2
4	RI	2	3	3
5	MH	2	2	5
6	MI	3	1	6
7	RF	3	1	5
8	AS	3	1	1
9	MJ	4	1	6
10	EF	2	1	2
11	MIL	3	4	2
12	MA	4	1	6
13	ES	6	1	4
14	GBDH	5	2	6
15	MHL	3	1	5
16	BS	3	1	4
17	HSM	3	1	6
18	MIS	2	1	6
19	AH	6	1	1
20	MHN	4	2	4

Kemudian, bagi *cluster* menjadi tiga kelompok ($K = 3$) dan cari pusat *centroid*-nya. Ada juga metode untuk menghitung pengelompokan, seperti yang ditunjukkan di bawah ini.

$K=3$ Centroid

$C_1 = (3, 1, 4)$ diambil dari data 1

$C_2 = (3, 1, 2)$ diambil dari data 2

$C_3 = (6, 1, 4)$ diambil dari data 13

Selanjutnya lakukan perhitungan dengan rumus berikut:

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

Iterasi 1:

1. A (3,1,4)

$$C_1 = (3,1,4) = \sqrt{(3-3)^2 + (1-1)^2 + (4-4)^2} = 0$$

$$C_2 = (3,1,2) = \sqrt{(3-3)^2 + (1-1)^2 + (4-2)^2} = 2$$

$$C_3 = (6,1,4) = \sqrt{(3-6)^2 + (1-1)^2 + (4-4)^2} = 3$$

2. B (3,1,2)

$$C_1 = (3,1,4) = \sqrt{(3-3)^2 + (1-1)^2 + (2-4)^2} = 2$$

$$C_2 = (3,1,2) = \sqrt{(3-3)^2 + (1-1)^2 + (2-2)^2} = 0$$

$$C_3 = (6,1,4) = \sqrt{(3-6)^2 + (1-1)^2 + (2-4)^2} = 3,61$$

3. C (3,1,2)

$$C_1 = (3,1,4) = \sqrt{(3-3)^2 + (1-1)^2 + (2-4)^2} = 2$$

$$C_2 = (3,1,2) = \sqrt{(3-3)^2 + (1-1)^2 + (2-2)^2} = 0$$

$$C_3 = (6,1,4) = \sqrt{(3-6)^2 + (1-1)^2 + (2-4)^2} = 3,61$$

Perhitungan di atas memberikan hasil Iterasi 1, seperti tabel di bawah ini.

Table 5. Hasil Iterasi 1

No	Inisial	X	Y	Z	C1	C2	C3	Group
1	R	3	1	4	0,00	2,00	3,00	1
2	AS	3	1	2	2,00	0,00	3,61	2
3	JN	3	1	2	2,00	0,00	3,61	2
4	RI	2	3	3	2,45	2,45	4,58	1

5	MH	2	2	5	1,73	3,32	4,24	1
6	MI	3	1	6	2,00	4,00	3,61	1
7	RF	3	1	5	1,00	3,00	3,16	1
8	AS	3	1	1	3,00	1,00	4,24	2
9	MJ	4	1	6	2,24	4,12	2,83	1
10	EF	2	1	2	2,24	1,00	4,47	2
11	MIL	3	4	2	3,61	3,00	4,69	2
12	MA	4	1	6	2,24	4,12	2,83	1
13	ES	6	1	4	3,00	3,61	0,00	3
14	GBDH	5	2	6	3,00	4,58	2,45	3
15	MHL	3	1	5	1,00	3,00	3,16	1
16	BS	3	1	4	0,00	2,00	3,00	1
17	HSM	3	1	6	2,00	4,00	3,61	1
18	MIS	2	1	6	2,24	4,12	4,47	1
19	AH	6	1	1	4,24	3,16	3,00	3
20	MHN	4	2	4	1,41	2,45	2,24	1

Setelah dihitung dengan menggunakan rumus pengelompokan yang ada, pengelompokan dilakukan berdasarkan jarak minimum ke *centroid* terdekat adalah:

Group Lama: (0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0)
 Group Baru: (1 2 2 1 1 1 1 2 1 2 2 1 3 3 1 1 1 1 3 1)

Terjadi pergantian kelompok yang dilakukan dengan iterasi sebagai berikut:
 Iterasi 2: K =3

Centroid 1 Group 1

$$C1 = \left(\frac{3+2+2+3+3+4+4+3+3+3+2+4}{12} \right) = 3$$

$$\left(\frac{1+3+2+1+1+1+1+1+1+1+2}{12} \right) = 1,33$$

$$\left(\frac{4+3+5+6+5+6+6+5+4+6+6+4}{12} \right) = 5$$

C1 = (3, 1,33, 5)

Centroid 2 Group 2

$$C2 = \left(\frac{3+3+3+2+3}{5} \right) = 2,8, \left(\frac{1+1+1+1+4}{5} \right) = 1,6, \left(\frac{2+2+1+2+2}{5} \right) = 1,8$$

C2 = (2,8, 1,6, 1,8)

Centroid 3 Group 3

$$C3 = \left(\frac{6+5+6}{3} \right) = 5,67, \left(\frac{1+2+1}{3} \right) = 1,33, \left(\frac{4+6+1}{3} \right) = 3,67$$

C3 = (5,67, 1,33, 3,67)

Jadi K=3 Centroid

C1 = (3, 1,33, 5)
 C2 = (2,8, 1,6, 1,8)
 C3 = (5,67, 1,33, 3,67)

1. A (3,1,4)

$$C1 = (3, 1,33, 5) = \sqrt{(3-3)^2 + (1-1,33)^2 + (4-5)^2} = 1,05$$

$$C2 = (2,8, 1,6, 1,8) = \sqrt{(3-2,8)^2 + (1-1,6)^2 + (4-1,8)^2} = 2,28$$

$$C3 = (5,67, 1,33, 3,67) = \sqrt{(3-5,67)^2 + (1-1,33)^2 + (4-3,67)^2} = 2,70$$

2. B (3,1,2)

$$C1 = (3, 1,33, 5) = \sqrt{(3-3)^2 + (1-1,33)^2 + (2-5)^2} = 3,02$$

$$C2 = (2,8, 1,6, 1,8) = \sqrt{(3-2,8)^2 + (1-1,6)^2 + (2-1,8)^2} = 0,66$$

$$C3 = (5,67, 1,33, 3,67) = \sqrt{(3-5,67)^2 + (1-1,33)^2 + (2-3,67)^2} = 3,16$$

3. C (3,1,2)

$$C1 = (3, 1,33, 5) = \sqrt{(3-3)^2 + (1-1,33)^2 + (2-5)^2} = 3,02$$

$$C2 = (2,8, 1,6, 1,8) = \sqrt{(3-2,8)^2 + (1-1,6)^2 + (2-1,8)^2} = 0,66$$

$$C3 = (5,67, 1,33, 3,67) = \sqrt{(3-5,67)^2 + (1-1,33)^2 + (2-3,67)^2} = 3,16$$

Perhitungan diatas memberikan hasil Iterasi 2, seperti tabel di bawah ini.

Table 6. Hasil Iterasi 2

No	Inisial	X	Y	Z	C1	C2	C3	Group
----	---------	---	---	---	----	----	----	-------

1	R	3	1	4	1,05	2,28	2,70	1
2	AS	3	1	2	3,02	0,66	3,16	2
3	JN	3	1	2	3,02	0,66	3,16	2
4	RI	2	3	3	2,79	2,01	4,08	2
5	MH	2	2	5	1,20	3,32	3,96	1
6	MI	3	1	6	1,05	4,25	3,56	1
7	RF	3	1	5	0,33	3,26	3,00	1
8	AS	3	1	1	4,01	1,02	3,79	2
9	MJ	4	1	6	1,45	4,41	2,89	1
10	EF	2	1	2	3,18	1,02	4,04	2
11	MIL	3	4	2	4,01	2,42	4,12	2
12	MA	4	1	6	1,45	4,41	2,89	1
13	ES	6	1	4	3,18	3,93	0,58	3
14	GBDH	5	2	6	2,33	4,76	2,52	1
15	MHL	3	1	5	0,33	3,26	3,00	1
16	BS	3	1	4	1,05	2,29	2,71	1
17	HSM	3	1	6	1,05	4,25	3,56	1
18	MIS	2	1	6	1,42	4,32	4,36	1
19	AH	6	1	1	5,01	3,35	2,71	3
20	MHN	4	2	4	1,56	2,54	1,83	1

Setelah dihitung menggunakan rumus pengelompokan yang ada pada Iterasi 2, pengelompokan dilakukan berdasarkan jarak minimum ke *centroid* terdekat adalah:

Group Lama: (1 2 2 1 1 1 1 2 1 2 2 1 3 3 1 1 1 1 3 1)
 Group Baru : (1 2 2 2 1 1 1 2 1 2 2 1 3 1 1 1 1 1 3 1)

Setelah dilakukan perhitungan dengan rumus pengelompokan yang ada, terjadi perubahan kelompok. Kemudian lanjutkan dengan iterasi berikutnya hingga anggota kelompok berhenti berganti.

3.2 Implementasi Python (Jupyter Notebook)

3.2.1 Data Preparation (Persiapan Data)

a. Import Data

Mengimpor data ke *Jupyter Notebook* merupakan langkah pertama yang penting dalam analisis data menggunakan bahasa pemrograman *Python*. Untuk impor data, digunakan pustaka *Pandas*, yang sangat populer dalam analisis data. Pertama, pastikan *Panda* diinstal di lingkungan *Jupyter Notebook*. Kemudian harus mengimpor perpustakaan dengan menulis baris kode yang sesuai di *Jupiter Notebook*. Penting untuk memastikan bahwa data yang diimpor berada di direktori yang benar sehingga *Jupyter Notebook* dapat mengaksesnya dengan benar. Ketika fase impor data selesai, data dapat diproses dan dianalisis. Pada langkah ini, dataset diubah ke dalam format CSV yang terdiri dari 445 record dan 4 atribut.

b. Transformasi

Tahapan transformasi data merupakan proses penting dalam analisis data untuk mengubah, menyesuaikan, atau mengolah data agar lebih sesuai dengan kebutuhan analisis dan memenuhi asumsi model yang digunakan. Identifikasi data yang memerlukan transformasi dan memilih untuk menghapus baris atau kolom yang memiliki nilai kosong. Pada proses ini digunakan untuk mengubah nama atribut pada dataset sehingga dapat diolah. Hasil data transformasi dapat dilihat dibawah ini.

Table 7. Hasil Transaformasi Data

No	Inisial	Umur	Jenis Narkoba	Lama Penggunaan (Bulan)
1	R	26	1	4
2	AS	27	1	2
3	JN	26	1	2
4	RI	22	3	3
5	MH	22	2	5
6	MI	32	1	12
7	RF	27	1	5
8	AS	29	1	1
9	MJ	44	1	7
...
441	MS	18	1	4
442	S	34	1	8
443	FH	27	1	5

444	A	50	1	2
445	RGL	33	1	5

c. Seleksi

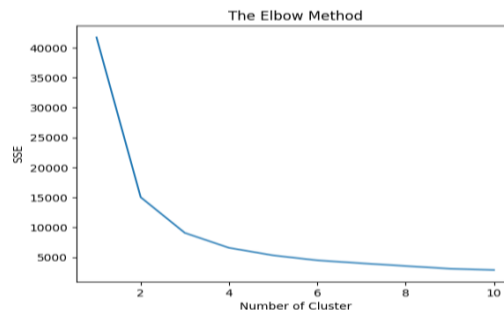
Tahapan seleksi data adalah proses kritis dalam analisis data, di mana peneliti atau analis data harus mengidentifikasi dan memilih subset data yang relevan dan sesuai dengan tujuan analisis yang ingin dicapai. Seleksi data harus selaras dengan tujuan analisis yang telah ditetapkan. Pilih variabel atau fitur yang memiliki hubungan langsung dengan tujuan analisis. Setelah tahapan seleksi data selesai, maka siap untuk melanjutkan ke tahap berikutnya dalam analisis data. Proses ini menentukan apa saja yang digunakan dalam proses *clustering*. Hasil atribut setelah dilakukan seleksi ditunjukkan pada tabel 7.

Table 8 Hasil Seleksi Dataset

Umur	Jenis Narkoba	Lama Penggunaan (Bulan)
26	1	4
27	1	2
26	1	2
22	3	3
22	2	5
32	1	12
27	1	5
29	1	1
44	1	7
...
18	1	4
34	1	8
27	1	5
50	1	2
33	1	5

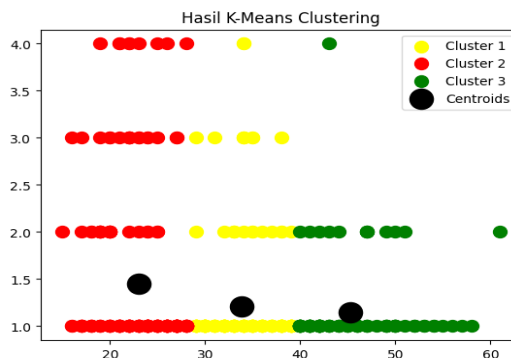
d. Proses Clustering

Clustering dilakukan di *Jupyter Notebook* menggunakan bahasa pemrograman *Python 3*. Proses *clustering* dimulai dengan mencari K optimal menggunakan Metode *Elbow* dengan *random state* : 42.



Gambar 3. Grafik K Optimal

Jika grafik percobaan dengan Metode Siku (*Elbow*) menunjukkan Penyusutan dan lengkungan yang signifikan, K maksimum ditentukan pada titik lokasi *cluster* yang ideal adalah tiga *cluster*. Grafik percobaan terlihat pada Gambar 3. Setelah memeriksa hasil diagram dari Metode *Elbow*, pengelompokan menggunakan *K-Means* dan dibagi menjadi tiga *cluster*. 3 *cluster* tersebut berwarna kuning, merah dan hijau. Gambar 4 menggambarkan temuan *cluster*.



Gambar 4 Hasil Clustering

Hasil *Clustering* kemudian sudah diketahui *centroid* atau titik pusatnya yang dapat dilihat pada gambar 5 dan jumlah anggota setiap *cluster* dapat dilihat pada gambar 6 .

```
[[33.86746988  1.21084337  4.41566265]
 [23.01675978  1.45251397  4.06145251]
 [45.27        1.15        4.66        ]]
```

Gambar 5. Centroid Cluster

Cluster 0: 166 data
Cluster 1: 179 data
Cluster 2: 100 data

Gambar 6. Jumlah Anggota Setiap Cluster

Setelah melakukan perhitungan dengan menggunakan *Jupyter Notebook* maka didapat hasil dari perhitungan yaitu:

1. *Cluster 0* dengan *centroid* 33.86746988, 1.21084337, 4.41566265 (terdapat 166 data). Dapat diketahui bahwa *cluster 0* pada rata-rata umur Dapat diketahui bahwa *cluster 1* pada rata-rata umur 26-45 tahun yang menggunakan narkotika selama 1 - >5 bulan dengan jenis narkoba yang digunakan yaitu sabu, ganja, inex dan *extacy*.
2. *Cluster 1* dengan *centroid* 23.01675978, 1.45251397, 4.06145251 (terdapat 179 data). (Dapat diketahui bahwa *cluster 1* pada rata-rata umur <16 - 25 tahun yang menggunakan narkotika selama 1 - >5 bulan dengan jenis narkoba yang digunakan yaitu sabu, ganja, inex dan *extacy*.
3. *Cluster 2* dengan *centroid* 45.27, 1.15, 4.66 (terdapat 100 data). Dapat diketahui bahwa *cluster 2* pada rata-rata umur 40 - >55 tahun yang menggunakan narkotika selama 3 - >5 bulan dengan jenis narkoba yang digunakan yaitu sabu, ganja dan *extacy*.

4. KESIMPULAN

Dari penelitian ini, dapat disimpulkan bahwa narkoba merupakan zat yang jika dikonsumsi secara tidak wajar dapat menyebabkan ketergantungan dan berbagai kondisi medis yang berbahaya. Penggunaan narkoba kronis terus berkembang, menghadirkan tantangan serius bagi masyarakat dan pemerintah dalam mengatasi penyalahgunaan narkoba. Untuk mengatasi masalah ini, rehabilitasi menjadi salah satu upaya penting yang dilakukan oleh para ahli. Kesimpulan dari penelitian ini adalah penggunaan algoritma *K-Means clustering* dalam pengelompokan data pasien rehabilitasi narkoba berhasil menghasilkan tiga kelompok (*cluster*) berdasarkan karakteristik mereka. *Cluster* pertama merupakan kelompok umur 26-45 tahun dengan lama penggunaan 1->5 bulan dan dengan jenis narkoba yang digunakan adalah sabu, ganja, inex, dan *extacy*. *Cluster* kedua adalah kelompok umur <16-25 tahun dengan pola penggunaan narkoba yang serupa dengan *cluster* pertama. *Cluster* ketiga merupakan kelompok umur 40->55 tahun dengan lama penggunaan 3->5 bulan dan dengan jenis narkoba yang digunakan adalah sabu, ganja, dan *extacy*. Penelitian ini memberikan informasi penting bagi Badan Narkotika Nasional Provinsi Sumatera Utara untuk menyusun program rehabilitasi yang lebih sesuai dengan kebutuhan masing-masing kelompok pasien. Berdasarkan hasil *cluster* yang telah didapat, bahwa program *parenting* dapat diterapkan pada kelompok *cluster* dua. Selain itu, penelitian ini juga menunjukkan bahwa penggunaan algoritma *K-Means clustering* dengan bahasa pemrograman *Python* dan *tools Jupyter Notebook* mampu memberikan hasil *clustering* yang optimal. Hasil penelitian ini dapat digunakan sebagai dasar pengembangan aplikasi dan penelitian lebih lanjut, serta perbandingan dengan algoritma *clustering* lain dalam pengelompokan data pasien rehabilitasi narkoba.

REFERENCES

- [1] B. L. Hasibuan, Sofiah, and E. Yolanda, "Pengklasifikasian Data Pasien Tes Urine Dengan Metode Clustering Pada Kantor Badan Narkotika Nasional Provinsi Sumut (BNNP SUMUT)," *JUKI J. Komput. dan Inform.*, vol. 4, no. 2, pp. 176–187, 2022.
- [2] A. Winarta and W. J. Kurniawan, "Optimasi cluster k-means menggunakan metode elbow pada data pengguna narkoba dengan pemrograman python," *J. Tek. Inform. Kaputama*, vol. 5, no. 1, pp. 113–119, 2021.
- [3] S. Hidayatun and Y. Widowaty, "Konsep Rehabilitasi Bagi Pengguna Narkotika yang Berkeadilan," *J. PENEKAKAN Huk. DAN KEADILAN*, vol. 1, no. 2, pp. 166–181, 2020, doi: 10.18196/jphk.1209.
- [4] P. U. F. Aulia and S. Saepudin, "PENERAPAN DATA MINING K-MEANS CLUSTERING UNTUK," *SISMATIK(Seminar NasionalSistemInformasidan ManajemenInformatika)*, pp. 209–2017, 2021.
- [5] S. Handoko, Fauziah, and Endah Tri Esti Handayani, "Implementasi Data Mining Untuk Menentukan Tingkat Penjualan Paket Data Telkomsel Menggunakan Metode K-Means Clustering.," *J. Ilm. Teknol. dan Rekayasa*, vol. 25, no. 1, pp. 76–88, 2020.
- [6] N. Mirantika, A. Tsamratul'ain, and Futry Diviana Agnia, "Penerapan Algoritma K-Means Clustering Untuk Mengelompokan Penyebaran Covid-19 Di Provinsi Jawa Barat.," *J. NUANSA Inform.*, vol. 15, no. 2, pp. 92–98, 2021.
- [7] B. A. Pangestu, N. A. Kristiawan, and N. Sulistiyowati, "Clustering Obat Untuk Menentukan Pola Pemasaran Efektif di Apotek Amar Sehat," *J. Ilm. Wahana Pendidik.*, vol. 8, no. 16, pp. 115–126, 2022.
- [8] A. Rohmah, F. Sembiring, and A. Erfina, "IMPLEMENTASI ALGORITMA K-MEANS CLUSTERING ANALYSIS UNTUK MENENTUKAN HAMBATAN PEMBELAJARAN DARING (STUDI KASUS : SMK YASPIM GEGERBITUNG)," *SISMATIK(Seminar Nas. Sist. Inf. dan Manaj. Inform.*, pp. 290–298, 2021.

- [9] A. Sulistiyawati and E. Supriyanto, "Implementasi Algoritma K-means Clustering dalam Penentuan Siswa Kelas Unggulan," *J. TEKNO KOMPAK*, vol. 15, no. 2, pp. 25–36, 2020.
- [10] A. Asroni, H. Fitri, and E. Prasetyo, "Penerapan Metode Clustering dengan Algoritma K-Means pada Pengelompokan Data Calon Mahasiswa Baru di Universitas Muhammadiyah Yogyakarta (Studi Kasus: Fakultas Kedokteran dan Ilmu Kesehatan, dan Fakultas Ilmu Sosial dan Ilmu Politik)," *Semesta Tek.*, vol. 21, no. 1, pp. 60–64, 2018, doi: 10.18196/st.211211.
- [11] N. W. Utami and A. A. I. I. Paramitha, "PENERAPAN DATA MINING UNTUK MENGETAHUI POLA PEMILIHAN PROGRAM STUDI DI STMIK PRIMAKARA MENGGUNAKAN ALGORITMA K-MEANS," *J. Teknol. Inf. dan Komput.*, vol. 7, no. 4, pp. 456–463, 2021.
- [12] D. A. Manalu and G. Gunadi, "IMPLEMENTASI METODE DATA MINING K-MEANS CLUSTERING TERHADAP DATA PEMBAYARAN TRANSAKSI MENGGUNAKAN BAHASA PEMROGRAMAN PYTHON PADA CV DIGITAL DIMENSI," *INFOTECH J. Technol. Inf.*, vol. 8, no. 1, pp. 45–54, 2022.
- [13] N. A. Pramono, M. H. Nurdiansyah, and D. Z. Hidayatullifa, "Rancang Bangun Pembuatan Sistem Pengiriman Sensor Secara Real Time Menggunakan Python dan Raspberry Pi," *Risal. Fis.*, vol. 3, no. 2, pp. 43–46, 2019, doi: 10.35895/rf.v3i2.154.
- [14] D. F. Sengkey, F. D. Kambey, S. P. Lengkong, S. R. Joshua, H. Valentino, and F. Kainde, "Pemanfaatan Platform Pemrograman Daring dalam Pembelajaran Probabilitas dan Statistika di Masa Pandemi CoVID-19," *J. Inform.*, vol. 15, no. 4, pp. 217–224, 2020, [Online]. Available: <https://ejournal.unsrat.ac.id/index.php/informatika/article/view/31685>
- [15] S. R. Pratama and A. H. Mirza, "PENERAPAN DATA MINING UNTUK MEMPREDIKSI TINGKAT INFLASI MENGGUNAKAN METODE REGRESI LINIER BERGANDA PADA BPS," *Bina Darma Conf. Comput. Sci.*, pp. 245–255, 2021.
- [16] N. Purba, Poningsih, and H. S. Tambunan, "Penerapan Algoritma K-Means Clustering Pada Penyebaran Penyakit Infeksi Saluran Pernapasan Akut (ISPA) di Provinsi Riau," *J. Inf. Syst. Res.*, vol. 2, no. 3, pp. 220–226, 2021.
- [17] R. T. Wahyuni, A. M. H. Pardede, and T. Pasaribu, "DATA MINING PENGELOMPOKAN PECANDU NARKOBA DI MASYARAKAT BERDASARKAN FAKTOR PENYEBAB PENGGUNAANYA MENGGUNAKAN METODE K-MEANS CLUSTERING," *J. Inform. Kaputama*, vol. 6, no. 3, pp. 836–844, 2022.
- [18] I. D. M. Suyoto, T. Rachmadi, and L. T. Parulian, "Menentukan Cluster Yang Tepat Dengan K-Means Dalam Rangka Mengukur Efektivitas Pelaksanaan Anggaran Pada Kementrerian Agraria Dan Tata Ruang/Badan Pertahanan Nasional," *INFOTECH J. Technol. Inf.*, vol. 8, no. 1, pp. 13–22, 2022, doi: <https://doi.org/10.37365/jti.v8i1.126>.
- [19] F. Sembiring, Octaviana, and S. Saepudin, "Implementasi Metode K-Means Dalam Pengklasteran Daerah Pungutan Liar di Kabupaten Sukabumi (Studi Kasus : Dinas Kependudukan Dan Pencatatan Sipil)," *J. tekno insentif*, vol. 14, no. 1, pp. 40–47, 2020.
- [20] F. N. Dhewayani, D. Amelia, Alifah, B. N. Sari, and M. Jajuli, "Implementasi K-Means Clustering untuk Pengelompokan Daerah Rawan Bencana Kebakaran Menggunakan Model CRISP-DM," *J. Teknol. dan Inf.*, vol. 12, no. 1, pp. 64–77, Mar. 2022, doi: 10.34010/jati.v12i1.6674.