

## DATA MINING *CLUSTERING* DATA OBAT-OBATAN MENGGUNAKAN ALGORITMA *K-MEANS* PADA RSU AN NI'MAH WANGON

Mukrodin<sup>1</sup>, Rohmat Taufiq<sup>2</sup>, Dwi Sana Rizkiyanti Ermi<sup>3</sup>

<sup>1,3</sup> Sistem Informasi, Fakultas Sains dan Teknologi, Universitas Peradaban, Jl Raya Pagojengan KM 03, Brebes

<sup>2</sup> Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Tangerang, Jl Perintis Kemerdekaan I/33, Tangerang  
Co Responden Email: mukrodins@gmail.com

### Article history

Received 26 Dec 2022

Revised 09 Jan 2023

Accepted 24 Mar 2023

Available online 15 May 2023

### Keywords

Clustering,

Data Mining,

K-Means Algorithm,

Medicine

### Riwayat

Diterima 26 Des 2022

Revisi 09 Jan 2023

Disetujui 24 Mar 2023

Terbit online 15 Mei 2023

### Kata Kunci

Clustering,

Data Mining,

Algoritma K-Means,

Obat-obatan

### Abstract

*The need for the right medicines can help control the supply of medicines effectively and efficiently, so that the availability of medicines of the right type and quantity is in accordance with the needs and can be obtained when needed. In data mining clustering can be used to analyze drug use, control, and drug planning in hospitals. The method used for drug data clustering is the K-Means algorithm which is a non-hierarchical data clustering method that partitions data into clusters so that data with different characteristics are grouped into other groups. The purpose of this research is to classify drug data at An Ni'mah Wangon General Hospital which can be used as a reference in decision making or planning and controlling of medical supplies or drug data at the hospital. Where, the results for making this decision will be obtained on how the pharmacy manages the drug stock that has been grouped.*

### Abstrak

Kebutuhan obat-obatan yang sesuai bisa membantu proses pengendalian pemasokan obat menjadi efektif dan efisien, sehingga ketersediaan berbagai jenis obat dan jumlah ketersediaan obat sesuai kebutuhan dan bisa diperoleh disaat dibutuhkan. Pada *data mining, clustering* bisa dipergunakan untuk melakukan analisa penggunaan obat-obatan, pengendalian, serta perencanaan ketersediaan berbagai macam obat di rumah sakit. Penelitian ini menggunakan metode *clustering* data obat-obatan yaitu algoritma *K-Means* yang merupakan metode data *clustering non hirarki* untuk mempartisi data ke dalam *cluster* hingga data yang memiliki karakteristik tidak sama dikelompokkan pada kelompok yang lain. Tujuan penelitian ini yaitu membentuk kelompok-kelompok data obat-obatan di RSU An Ni'mah Wangon yang dapat dipakai sebagai acuan pada proses pengambilan keputusan ataupun perencanaan dan pengendalian pasokan medis/data obat-obatan di rumah sakit. Dimana, hasil untuk pengambilan keputusan ini akan diperoleh bagaimana untuk apotek mengelola stok obat yang telah dikelompokkan tersebut.

## PENDAHULUAN

Rumah Sakit Umum An Ni'mah Wangon merupakan salah satu rumah sakit yang berada di wilayah kabupaten banyumas tepatnya di Jl. Raya Timur Klapagading Kulon, Wangon Banyumas. Rumah sakit An Ni'mah Wangon berdiri pada tanggal 5 Juni 1998, digawangi oleh dr. Teguh Ariyanto MPH dan istri tercintanya yang juga seorang dokter yaitu dr. Ratna Widarastuti MMR, Rumah Sakit tersebut perkembangannya semakin pesat baik fasilitas, sarana juga prasarana dalam pelayanannya hingga sekarang seperti laboratorium, fisio terapi,

USG Kandungan, HCU, perinatologi dan lainnya. Layanan tersebut tentu tidak terpisahkan dari pelayanan obat yang disediakan oleh farmasi atau apotek rumah sakit, dalam pelayanan apotek ini haruslah cepat dan tepat serta sesuai resep dokter.

Salah satu aspek penting dalam perencanaan kebutuhan obat-obatan penting untuk penentuan dalam pengelolaan obat-obatan, maka dari itu pengadaan obat-obatan haruslah tepat sesuai dengan kebutuhannya dan ketersediaan obat-obatan yang cukup dengan jenis dan jumlah yang tepat. Kegiatan itu dapat memberikan pengaruh dalam hal

pengadaan, distribusi obat serta obat-obatan yang dipakai serta untuk proses pimpinan dalam mengambil keputusan dalam merencanakan serta pengendalian manajemen rantai pasok obat pada Apotik yang ada di Rumah Sakit An Ni'mah Wangon.

RSU An Ni'mah memiliki apotek yang melayani transaksi penjualan obat setiap harinya. Seiring dengan transaksi penjualan obat yang dilakukan tersebut, maka apotek akan menghasilkan data transaksi yang besar sehingga data tersebut semakin lama data obat akan diperbarui yang nantinya akan menyebabkan kebaruan informasi mengenai data obat-obatan. Berdasarkan hasil observasi yang penulis lakukan pada rumah sakit tersebut, data-data transaksi tersebut berfungsi sebagai arsip dan laporan saja. Belum adanya pemanfaatan dari informasi atau data-data tersebut seperti pengelompokan atau *clustering* guna mendapatkan sebuah informasi baru.

Dari penumpukan data obat tersebut dimana belum adanya pemanfaatan untuk dihasilkan sebuah informasi, hal ini berakibat pada persediaan obat yang tidak terkendali sehingga obat yang dibutuhkan pasien seringkali tidak ada atau bahkan habis. Untuk menyelesaikan persoalan tersebut, dibutuhkan perencanaan yang matang kebutuhan obat-obatan dan pemanfaatan data obat terhadap pemakaian rendah, sedang dan tinggi. Sehingga, perlu dilakukan analisis data terhadap data obat salah satunya adalah pengelompokan atau *clustering* data-data obat tersebut serta bisa dijadikan sebagai acuan obat-obatan yang dibutuhkan.

Data mining merupakan sebuah cara pencarian *trend/pola* yang diinginkan dalam sebuah basis data besar (*data warehouse*) dalam rangka membantu proses pengambilan keputusan pada waktu mendatang (Syahdan & Sindar, 2018) (Nurofik et al., 2021). Selanjutnya, Mochamad Wahyudi (2020) menambahkan informasi tentang *data mining* yang merupakan suatu proses iteratif serta interaktif dalam rangka mendapatkan beberapa pola atau model-model baru yang terbaik (sempurna), dan bermanfaat serta dapat dipahami pada sebuah database yang memiliki data sangat besar (*massive databases*).

Definisi *clustering* ialah sebuah metode yang dipakai pada proses *data mining* dimana memiliki cara kerja dengan mencari, pengelompokan data yang memiliki kemiripan karakteristik diantara data satu dan data yang lain (Ong, 2013). Pada metode *clustering* bagian dari objek yang memiliki jenis karakteristik serupa akan dikelompokkan pada sebuah kelompok dan objek yang mempunyai karakteristik tidak sama akan dikelompokkan di kelompok lainnya (Herdiansah et al., 2022). Contoh metode *clustering* yang terkenal serta merupakan algoritma *clustering* ialah *K-means*, metode ini sering digunakan karena sederhana dan efisien.

Algoritma *K-means* merupakan salahsatu algoritma dari metodologi *Clustering*, yang berupa pengelompokan data, dimana data dibagi menjadi dua kelompok atau lebih. selain kesederhanaan dan efisiensinya, *K-means* ini memiliki sifat efisien dan cepat.

Tujuan dari algoritma *K-means* berfungsi mendesain *cluster* objek yang didasarkan pada atribut yang berubah menjadi partisi *k*, cara kerjanya diawali dengan menentukan *cluster*, selanjutnya pada setiap *cluster* di elemen pertama yang dipilih dijadikan sebagai titik pusat *cluster (centroid)*, setelah itu pengulangan setiap step sampai tidak dapat ditemukan objek yang dapat dipindahkan lagi.

## METODE PENELITIAN

### Data Mining

Data mining didefinisikan seperangkat teknik yang bisa digunakan dengan cara otomatis berfungsi mengeksplorasi data dilakukan untuk mengidentifikasi relasi yang kompleks dalam set data besar yang biasanya diorganisir dalam format tabel, seperti yang digunakan dalam teknologi manajemen basis data relasional. (Amril Mutoi Siregar, 2020)

Data mining merujuk pada proses penggalian pengetahuan dalam database. Sebuah data mining merupakan sebuah metode untuk mengekstrak informasi atau menemukan pola dalam kumpulan data besar (Herdiansah et al., 2022).

Selanjutnya (Gorunescu, 2011) memaparkan terdapat 3 langkah sistematis yang terdapat data mining:

1. Pemrosesan awal pada data ‘langkah-langkah yang dilakukan dalam pemrosesan ini adalah Untuk mempersiapkan data sebelum diproses, berbagai teknik digunakan, termasuk pembersihan data, normalisasi data, transformasi data, penanganan data yang tidak valid, reduksi dimensi, pemilihan subset fitur, dan sebagainya.
2. Membangun model dan melakukan validasi berarti melakukan analisis terhadap berbagai model untuk memilih model yang memiliki kinerja prediksi terbaik. Pada tahap ini, berbagai metode digunakan, seperti klasifikasi, regresi, analisis cluster, deteksi anomali, analisis asosiasi, analisis pola sekuensial, dan lain-lain.
3. Tahap ketiga dari proses tersebut melibatkan pengaplikasian model ke dalam data baru dengan tujuan menghasilkan prediksi terhadap masalah yang sedang diinvestigasi.

## Clustering

Analisis *cluster* adalah salah satu Teknik *datamining* yang memiliki tujuan untuk melakukan identifikasi sekelompok objek yang memiliki kemiripan/keserupaan karakteristik tertentu yang bisa dijadikan acuan memisahkan objek tersebut dengan kelompok objek yang lain, hingga pada akhirnya objek yang berada di kelompok yang memiliki kesamaan relatif lebih homogeni dari objek yang terdapat di kelompok yang berbeda (Najmuddin & Herdiansah, 2021). Besarnya kelompok yang bisa diidentifikasi bergantung pada banyak serta variasi objek data. Sedangkan tujuan dari pengelompokan kumpulan objek data menjadi beberapa kelompok yang memiliki karakter tertentu serta bisa dibedakan satu dengan yang lain adalah untuk analisis dan interpretasi lebih lanjut (Nurchayyo, 2015).

Teknik pengelompokan ini dibagi menjadi tiga kategori, yaitu algoritma parsial, hierarkis, dan lokal. Ada sekumpulan objek dan pengelompokan atau kriteria

pengelompokan, dan pengelompokan parsial membagi objek ke dalam kluster sedemikian rupa sehingga objek dalam satu kluster lebih mirip dengan objek dalam satu kluster daripada objek dalam kluster lain. (Varghese et al., 2011).

Analisis *Clustering* atau analisis kelompok adalah sebuah Teknik yang bisa digunakan untuk mengenali berbagai kelompok yang diperoleh dari pengelompokan beberapa unsur yang lebih kecil megacu pada adanya kemiripan satu objek dengan objek lainnya.

## Algoritma K-Means

*K-Means* adalah sebuah algoritma *Clustering* yang merupakan kelompok/kumpulan *Unsupervised learning*, biasanya dipergunakan untuk membagi data jadi beberapa bagian/kelompok menggunakan sistem partisi, algoritma ini dapat memperoleh *input*-an berupa data label kelas (Wanto, 2020).

Algoritma *kmeans* dikenal dan banyak digunakan dalam metode partisi, yaitu cara memisahkan himpunan objek data ke sub himpunan *cluster* yang tak saling tumpang-tindih, hingga setiap objek data tepat berada di dalam sebuah *cluster*.

Menurut (Gustientiedina et al., 2019), algoritma yang umum digunakan dalam metode K-Means adalah sebagai berikut:

1. tetapkan total jumlah cluster k yang terbentuk. Jumlah cluster k dapat ditentukan berdasarkan pertimbangan teoritis dan konseptual yang diajukan.
2. Pilih k titik acak sebagai pusat cluster. Titik pusat pertama/awal diambil dengan acak dari objek tersedia untuk membentuk k klaster. Selanjutnya, titik pusat klaster ke-i dihitung mempergunakan rumus sebagai berikut:

$$v = \frac{\sum_{i=1}^n x_i}{n} ; i = 1, 2, 3, \dots, n$$

dimana:

v : *centroid* pada *Cluster*

x<sub>i</sub> : objek ke-i

n : banyaknya objek/jumlah objek yang menjadi anggota *Cluster*

3. Hitung jarak setiap objek ke masing-masing *centroid* dari masing-masing *Cluster*. Untuk menghitung jarak antara objek dengan *centroid* penulis

$$d(x,y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} ; i = 1,2,3, \dots, n$$

menggunakan Euclidean Distance. dimana:

$x_i$  : obyek x ke-i  $y_i$  : daya y ke-i  $n$  : banyaknya objek

4. Dalam proses alokasi setiap objek ke centroid terdekat saat iterasi, terdapat dua cara umum yang dapat dilakukan, yaitu menggunakan hard k-means atau soft k-means. Pada hard k-means, setiap objek secara tegas dinyatakan sebagai anggota cluster dengan mengukur jarak kedekatannya dengan titik pusat cluster. cara lain dapat dilakukan dengan *fuzzy C means*.
5. Lakukan Iterasi, setelah melakukan iterasi langkah berikutnya menentukan posisi *centroid* baru dengan menggunakan persamaan 1.
6. Ulangi langkah 3 jika posisi *centroid* baru tidak sama.

Pada metode K-Means, data yang dapat diolah dalam perhitungan dibagi menjadi data numerik berupa angka. Setiap objek kemudian dihitung kedekatannya berdasarkan karakter yang dimilikinya dengan pusat klaster yang telah ditentukan, jarak terkecil antar objek dan setiap klaster merupakan anggota klaster terdekat. Setelah menentukan jumlah cluster, tiga objek dipilih secara acak untuk mewakili pusat awal cluster. Pusat cluster tersebut akan dihitung kedekatannya dengan seluruh objek yang ada.

## HASIL DAN PEMBAHASAN

Tahap pendahuluan adalah serangkaian kegiatan untuk memastikan penelitian akan baik. Setelah itu, proses penelitian akan berlangsung secara sistematis dan menghasilkan hasil yang berkualitas. Langkah yang dilakukan sebagai berikut:

1. Penentuan Tujuan

Langkah ini memiliki fungsi untuk memperjelas kerangka tentang apa saja yang menjadi tujuan penelitian. Sedangkan penelitian ini memiliki tujuan pengelompokkan pemakaian data obat-obatan menggunakan algoritma K-means clustering.

### 2. Merumuskan Masalah

Untuk menentukan topik penelitian, dilakukan observasi dan pengumpulan masalah yang ada. Kemudian, rumusan masalah yang sesuai dengan permasalahan dapat ditentukan. Oleh karena itu, rumusan masalah dalam penelitian ini adalah bagaimana cara mengelompokkan data penggunaan obat-obatan di RSUD An Ni'mah Wangon.

### Preprocessing Data

Sebelum pengolahan data dalam data mining, tujuannya adalah untuk membersihkan dan meningkatkan kualitas hasil pengelompokan data. Ini dilakukan dengan melakukan hal-hal seperti menghapus data yang tidak relevan dan mengurangi jumlah rekaman duplikat. Beberapa penjelasan tahapan *preprocessing data* sebagai berikut:

#### 1. Seleksi Atribut/Data

Pada penelitian ini, tidak semua data atau atribut pada *dataset* dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis data atau atribut mana yang akan diambil dari *dataset* tersebut sesuai dengan tujuan dan kebutuhan pada penelitian ini. Adapun atribut yang digunakan antara lain ialah nama obat, persediaan obat yang diambil dari laporan penjualan obat sejak bulan mei hingga agustus. Untuk mengetahui informasi data atau atribut data dapat dilihat pada gambar berikut ini:

Tabel 1. Informasi Data

No	Column	Non-Null	Count	Dtype
1	Jns.Transaksi	18324	non-null	object
2	Nama Obat/alkes	18324	non-null	object
3	Depo Farmasi	18324	non-null	object
4	Kredit	18324	non-null	float64
5	Saldo	18324	non-null	float64

Dari gambar diatas menjelaskan bahwa informasi data memuat 5 kolom, yakni terdiri dari 18323 data dan 5 variabel, yaitu 3 variabel (*JNS.TRANSAKSI, NAMA OBAT/ALKES, DEPO FARMASI*) yang memiliki tipe data objek, dan 2 variabel yakni (*KREDIT dan SALDO RS*) yang memiliki tipe integer.

2. Pembersihan Data (*Data Cleaning*)

Pada langkah awal ini, dilakukan proses penting dalam penelitian yaitu melakukan pembersihan data agar sesuai untuk di mining. Proses pembersihan ini bertujuan untuk menghapus data yang tidak digunakan atau tidak layak untuk diolah dalam proses data mining seperti data noise atau missing value pada data yang tidak konsisten. Selain itu, dalam penelitian ini juga dilakukan penghapusan data yang tidak relevan yang diisi oleh responden dari luar studi kasus penelitian ini.

Selama analisis data, kami membersihkan data agar dapat digunakan untuk penambangan. Ini termasuk menghapus data yang tidak relevan atau tidak cocok dengan proses penambangan data. Kami juga menghapus data yang secara tidak sengaja disertakan dalam penelitian, seperti noise dan missing value.

Secara umum, data yang dimaksud dalam pembersihan data adalah data yang tak lengkap misalkan data hilang ataupun data yang tidak valid. Selain dari hal tersebut, atribut data dimana data tersebut tidak relevan akan harus di-*delete*. Proses pembersihan data yang dilakukan dapat membuat pengaruh kinerja dari teknik proses *data mining* yang digunakan.

Tabel 2. Data Cleaning

No	NAMA OBAT	Kredit	Saldo_RS
0	Acyclovir Salep	5	24
1	Acyclovir TAB 400 MG	55	23
2	Alpara Tab	93	22
3	Ambroxol Syr	24	21
4	Ambroxol Tab	1082	20
5	Aminophyllin Inj	26	182
6	Amoxicilin 500mg	44	157
7	Ampilicilin Inj	43	137
8	Anadium Kaplet	19	1088
9	Analsik tab	645	1087

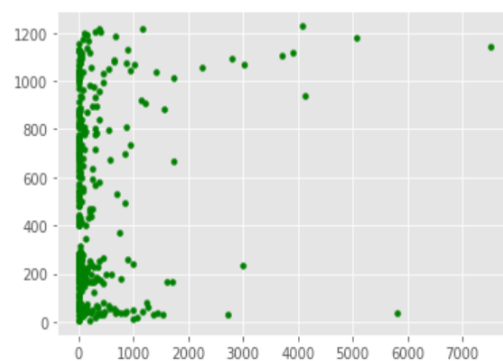
Hasil Penelitian  
Memuat data

Melakukan pemrosesan data/*loading* data di file data Set yang telah dibersihkan/*cleaning* untuk kemudian data di-*upload* ke *google colab*

Tabel 3. Memuat Data

No	NAMA OBAT	Kredit	Saldo_RS
0	Acyclovir Salep	1.0	24.0
1	Acyclovir Salep	1.0	23.0
2	Acyclovir Salep	1.0	22.0
3	Acyclovir Salep	1.0	21.0
4	Acyclovir Salep	1.0	20.0
5	Acyclovir Tab 400 mg	10.0	182.0
6	Acyclovir Tab 400 mg	25.0	157.0
7	Acyclovir Tab 400 mg	20.0	137.0
8	Alpara Tab	10.0	1088.0
9	Alpara Tab	1.0	1087.0
10	Alpara Tab	10.0	1077.0
11	Alpara Tab	10.0	1067.0

Terdapat 2 kolom yang akan kita gunakan, maka data dari kedua kolom di pisahkan terlebih dahulu dan ditampung ke dalam dua variabel yang berbeda. Kedua variabel tersebut, akan di plot ke dalam sebuah *scatter* dengan semua warna data berwarna hijau yang artinya data tersebut masih merupakan seluruh data dibagian yang sama.



Gambar 1. Visualisasi Data Set

Clustering Data

a. Fungsi Euclidean Distance

Buat Fungsi Euclidean Distance untuk menghitung jarak dari data ke *centroid* dan

- juga berguna untuk menghitung jarak antara *centroid*.
- b. Nilai *Cluster* Dan *Centroid*
    - i. Tentukan terlebih dahulu nilai *k* sebagai banyaknya *cluster* yang akan dihitung.
    - ii. Tentukan nilai *centroid* awal pada sumbu *x* dan sumbu *y*.
  - c. Pengelompokan Data Pada Tiap *Cluster*
    - i. Buat variable penampung yang akan menampung data pada tiap *cluster* yang berbeda. Variable tersebut sejumlah dengan jumlah *cluster* yang digunakan.
    - ii. Lakukan pengelompokan selama panjang data.
    - iii. Tampilkan indeks dari tiap data yang ada pada suatu *cluster* dengan banyaknya data pada *cluster* tersebut.

```
Cluster 1: [0, 1, 2, 3, 4, 5, 6, 7, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 81, 82, 83, 84, 85, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 132, 133, 135, 136, 137, 160, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 191, 192, 193, 194, 195, 196, 199, 200, 202, 204, 205, 206, 207, 209, 287, 288, 289, 290, 292, 293, 294, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 385, 386, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435] Jumlah data: 272
```

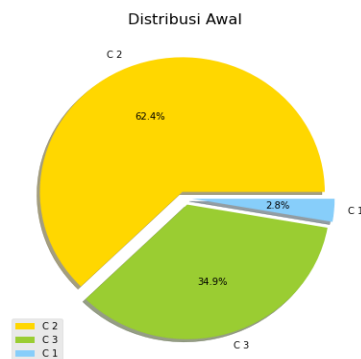
```
Cluster 2: [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 86, 88, 98, 103, 104, 108, 109, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 197, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 210, 211, 212, 213, 214, 215, 216, 217, 218, 219, 220, 221, 222, 223, 224, 225, 226, 227, 228, 229, 230, 231, 232, 233, 234, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 289, 294, 384] Jumlah data: 152
```

```
Cluster 3: [197, 210, 218, 222, 226, 227, 231, 235, 236, 248, 291, 307] Jumlah data: 12
```

Gambar 2. Cluster Awal

### Distribusi Cluster

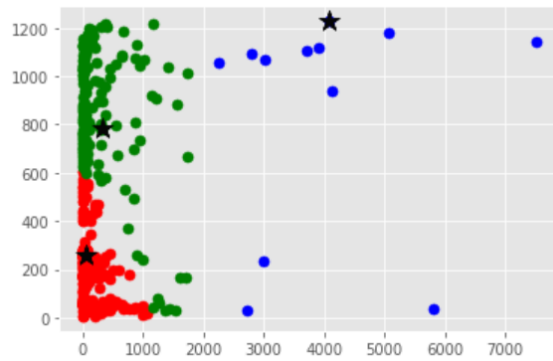
Distribusi *cluster* digunakan untuk menunjukkan kontribusi relatif dari tiap *cluster* terhadap total keseluruhan data. Distribusi *cluster* dapat dilihat pada diagram berikut:



Gambar 3. Distribusi Awal Cluster

Berdasarkan diagram diatas distribusi cluster terbanyak didapati pada C2 dimana terdapat distribusi cluster sebanyak 62,4%, sedangkan C1 dan C3 mendapati distribusi cluster yang lebih kecil yaitu 2,8% dan 34,9%

### Visualisasi Data



Gambar 4. Visualisasi Cluster Awal

### Update Centroid

*Update centroid* digunakan untuk menetapkan *centroid* terbaik. *Update centroid* dapat dilakukan sebagai berikut:

- a. Buat variabel untuk menampung *centroid* lama, yaitu *centroid* yang belum di *update*. *Centroid* lama tersebut akan digunakan untuk menghitung jarak antara *centroid* lama dengan *centroid* baru (yang sudah di-*update*).
- b. Buat variabel untuk menampung jenis *cluster* dari tiap data, sehingga data dengan indeks ke *i* setelah proses perhitungan jarak akan memiliki nilai *cluster*-nya.
- c. Buat variabel untuk menampung jarak antara *centroid* lama dengan *centroid* baru (yang sudah di-*update*).
- d. Lakukan perulangan selama jarak antara *centroid* lama dengan *centroid* baru (yang sudah di-*update*) sudah bernilai 0. Artinya letak *centroid* lama dan *centroid* baru sudah tidak berpindah pindah lagi.
- e. Tampilkan nilai *centroid* baru yang sudah tidak berpindah pindah lagi.

### Pengelompokan Menggunakan Centroid yang sudah di update

Setelah mendapatkan *centroid* yang sudah di-*update* kemudian lakukan pengelompokan kembali untuk mendapatkan hasil cluster terbaik. Pengelompokan dilakukan sebagai berikut:

- a. Buat variable penampung yang akan menampung data pada tiap *cluster* yang berbeda. Variable tersebut sejumlah dengan jumlah *cluster* yang digunakan.
- b. Lakukan pengelompokan selama panjang data.

- c. Tampilkan indeks dari tiap data yang ada pada suatu *cluster* dengan banyaknya data pada *cluster* tersebut.

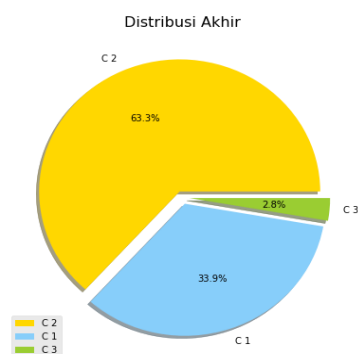
```
Cluster 1: [0, 1, 2, 3, 4, 5, 6, 7, 12, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36, 37, 38, 39, 40, 41, 42, 43, 44, 45, 83, 84, 85, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 99, 100, 101, 102, 103, 104, 105, 106, 107, 108, 109, 110, 111, 112, 113, 114, 115, 116, 117, 118, 119, 120, 121, 122, 123, 124, 125, 126, 127, 128, 129, 130, 132, 133, 135, 136, 137, 161, 162, 163, 164, 165, 166, 167, 168, 169, 170, 171, 172, 173, 174, 175, 176, 177, 178, 179, 180, 181, 182, 183, 184, 185, 186, 187, 188, 189, 190, 191, 192, 193, 194, 195, 196, 198, 199, 200, 201, 202, 203, 204, 205, 206, 207, 208, 209, 288, 289, 290, 292, 293, 294, 295, 296, 297, 298, 299, 300, 301, 302, 303, 304, 305, 306, 307, 308, 309, 310, 311, 312, 313, 314, 315, 316, 317, 318, 319, 320, 321, 322, 323, 324, 325, 326, 327, 328, 329, 330, 331, 332, 333, 334, 335, 336, 337, 338, 339, 340, 341, 342, 343, 344, 345, 346, 347, 348, 349, 350, 351, 352, 353, 354, 355, 356, 357, 358, 359, 360, 361, 362, 363, 364, 365, 366, 367, 368, 369, 370, 371, 372, 373, 374, 375, 376, 377, 378, 379, 380, 381, 382, 383, 384, 385, 386, 388, 389, 390, 391, 392, 393, 394, 395, 396, 397, 398, 399, 400, 401, 402, 403, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415, 416, 417, 418, 419, 420, 421, 422, 423, 424, 425, 426, 427, 428, 429, 430, 431, 432, 433, 434, 435] Jumlah data: 276

Cluster 2: [8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 46, 47, 48, 49, 50, 51, 52, 53, 54, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 72, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 86, 88, 90, 131, 134, 138, 139, 140, 141, 142, 143, 144, 145, 146, 147, 148, 149, 150, 151, 152, 153, 154, 155, 156, 157, 158, 159, 160, 211, 212, 213, 214, 215, 216, 217, 219, 220, 221, 223, 224, 225, 228, 229, 230, 231, 233, 234, 237, 238, 239, 240, 241, 242, 243, 244, 245, 246, 247, 249, 250, 251, 252, 253, 254, 255, 256, 257, 258, 259, 260, 261, 262, 263, 264, 265, 266, 267, 268, 269, 270, 271, 272, 273, 274, 275, 276, 277, 278, 279, 280, 281, 282, 283, 284, 285, 286, 287] Jumlah data: 148

Cluster 3: [197, 218, 218, 222, 226, 227, 231, 235, 236, 248, 291, 387] Jumlah data: 11
```

Gambar 5. Hasil Cluster Akhir

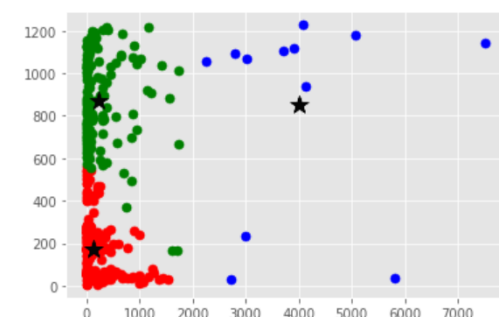
### Distribusi Akhir



Gambar 6. Distribusi Cluster Akhir

Berdasarkan diagram diatas distribusi *cluster* terbanyak dimiliki pada C2 dimana terdapat distribusi *cluster* sebanyak 63,3%, sedangkan C1 dan C3 mendapati distribusi *cluster* yang lebih kecil yaitu 33,9% dan 2,8%.

### Visualisasikan Kembali



Gambar 7. Visualisasi Cluster Akhir

### KESIMPULAN

Dari hasil *clusterisasi* pada data obat-obatan dapat ditarik kesimpulan bahwa kelompok obat yang termasuk pemakaian sedikit rata-rata permintaan obat pada bulan Mei untuk pemakaian kurang dari 31877

buah, dan obat yang termasuk pemakaian sedang rata-rata permintaan obat pada bulan Mei sekitar 42364 buah, sedangkan obat yang termasuk ke dalam obat pemakaian tinggi rata-rata permintaan obat pada bulan Mei diatas 48008 buah. Informasi yang didapat dari hasil klusterisasi obat tersebut setelah dikonfirmasi pada pihak rumah sakit, kemudian dapat diketahui pemakai obat tinggi rendahnya obat. Melalui data mining tersebut kemudian apotek menggunakannya sebagai referensi kebijakan pengambilan keputusan untuk mengendalikan stok obat dirumah sakit, dimana untuk obat dengan pemakaian rendah, rumah sakit akan menyetoknya dibawah 3000 buah, obat dengan pemakaian sedang akan di stok dari 3000-5000 buah dan obat dengan pemakaian tertinggi, rumah sakit akan menyetoknya diatas 5000 buah.

### SARAN

Penelitian ini sangat membantu bidang kesehatan terutama pada bidang apotik, diharapkan dengan adanya penelitian ini dapat mengurangi ada kesalahan pemilihan obat, serta penelitian ini bisa dikembangkan lebih luas terutama untuk implementasi aplikasi obat-obatan.

### REFERENSI

Prasetyo, Eko. (2014). Data Mining: Mengolah Data Menjadi Informasi Menggunakan Matlab. Bojonegoro: Universitas Negeri Malang.

Amril Mutoi Siregar, S. K. M. K. D. A. N. A. P. S. K. M. K. (2020). *DATA MINING: Pengolahan Data Menjadi Informasi dengan RapidMiner*. CV Kekata Group. <https://books.google.co.id/books?id=rTlmDwAAQBAJ>

Bogadenta, A. (2013). *Manajemen Pengelolaan Apotek*. D-Medika.

Creative, S. C. (2017). *Jago Microsoft Excel 2016*. Elex Media Komputindo. <https://books.google.co.id/books?id=r0RGDwAAQBAJ>

Darmawan, U., Destriana, R., & Tisno, W. (2022). Analisis Penerapan Website e-Government Pemerintah Kabupaten Tangerang Menggunakan

- GTMetrix. *Jurnal Telematika*, 17(1), 24-31.
- Departemen Kesehatan RI. (2017). Sterkwerkende *Geneesmiddelen Ordonnantie*, Staatsblad 1949:419); *Peraturan Menteri Kesehatan Republik Indonesia Nomor 9 Tahun 2017 Tentang Apotik*, 1–36.
- Destriana, R., Rusdianto, H., Prabowo, D. C., Erri, D., & Putra, J. L. (2023). Interactive Application Development Using Augmented Reality Concept for Catalog Tire in Planet Ban Store Pasar Kemis. *Jurnal Teknik Komputer AMIK BSI*, 9(1), 41-45.
- Firmansyah, Y., & Pitriani. (2017). Penerapan Metode SDLC Waterfall Dalam Pembuatan Aplikasi Pelayanan Anggota Pada Cu Duta Usaha Bersama Pontianak. *Jurnal Bianglala Informatika*, 5(2), 53–61. <https://ejournal.bsi.ac.id/ejurnal/index.php/Bianglala/article/view/2703/1813>
- Herdiansah, A., Borman, R. I., Nurnaningsih, D., Sinlae, A. A. J., & Al Hakim, R. R. (2022). Klasifikasi Citra Daun Herbal Dengan Menggunakan Backpropagation Neural Networks Berdasarkan Ekstraksi Ciri Bentuk. *JURIKOM (Jurnal Riset Komputer)*, 9(2), 388–395. <https://doi.org/10.30865/jurikom.v9i2.4066>
- Herdiansah, A., Nurnaningsih, D., & Rusdianto, H. (2022). Pemanfaatan Flutter Pada Pengembangan Aplikasi Mobile EBisnis Penyediaan Bahan Baku Bisnis Katering. *Jurnal Teknoinfo*, 16(2), 291–303. <https://doi.org/10.33365/jti.v16i2.1937>
- Gorunescu, F. (2011). Data mining: Concepts, models and techniques. *Intelligent Systems Reference Library*. <https://doi.org/10.1007/978-3-642-19721-5>
- Gustientiedina, G., Adiya, M., & Desnelita, Y. (2019). Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5, 17–24. <https://doi.org/10.25077/TEKNOSI.v5i1.2019.17-24>
- Jollyta, D., Ramdhan, W., & Zarlis, M. (2020). *Konsep Data Mining Dan Penerapan*. Deepublish. <https://books.google.co.id/books?id=pimJEAAAQBAJ>
- Najmuddin, N., & Herdiansah, A. (2021). Decision Support System Fuzzy Analytic Hierarchy Process Method Studi Kasus Pemilihan Vendor Kemasan Terbaik. *Jurnal Teknik Informatika (JIKA)*, 5(2), 124–133
- Nurchahyo, D. N. G. W. (2015). *Algoritma Data Mining Dan Pengujian*. Deepublish. <https://books.google.co.id/books?id=F n-QDwAAQBAJ>
- Nurofik, Agus, Elsy Rahajeng, Novi Yona Sidratul Munti, Sutisna, Hamdan Firmansyah, Amar Sani, Decky Hendarsyah, Sukri Adrianto, Wawang Adi Darma, Arief Herdiansah, Deny Ariestiandy, Desi Nurnaningsih, Iwan Setiawan, Adrianto Sugiarto Wiyono, dan Zaharah. 2021. *Pengantar Teknologi Informasi*. Ed.1. disunting oleh I. Kusumawati dan M. Sari. Cirebon: Insania.
- Sibero, A. F. K. (2013). *Web Programming Power Pack*. Mediakom.
- Wanto, A. (2020). *Data Mining : Algoritma dan Implementasi*. Yayasan Kita Menulis. <https://books.google.co.id/books?id=LSPfDwAAQBAJ>