

PENERAPAN ALGORITME K-MEANS DALAM KLASTERISASI DATA PENJUALAN OBAT APOTEK KIDANGRANGGA

Kunaifi Kurnia Abdullah, Iqbal Maulana, Aries Suharso, Garno, Aji Primajaya

Program Studi Informatika S1, Fakultas Ilmu Komputer
Universitas Singaperbangsa Karawang, Karawang Jawa Barat
1910631170199@student.unsika.ac.id

ABSTRAK

Kesehatan adalah aspek yang penting dalam kehidupan manusia. Tanpa kesehatan yang baik, manusia akan mengalami hambatan dalam menjalankan aktivitas sehari-hari secara optimal. Oleh karena itu, isu-isu kesehatan selalu menjadi fokus perhatian di seluruh dunia. Salah satu cara untuk meningkatkan kesehatan adalah dengan meningkatkan kualitas layanan kesehatan. Pelayanan kesehatan menjadi tempat di mana masyarakat berkonsultasi mengenai kesehatan mereka. Salah satu aspek penting dalam meningkatkan layanan kesehatan adalah pengelolaan obat yang efektif. Apotek Kidangrangga, yang terletak di kecamatan Rengasdengklok Kabupaten Karawang, telah menggunakan sistem yang saling terhubung untuk mengelola data, termasuk data obat. Namun, setelah mewawancarai pemilik apotek tersebut, diketahui bahwa sering terjadi kelebihan dan kekurangan obat di Apotek Kidangrangga. Algoritme K-Means dapat diimplementasikan dengan baik untuk melakukan klusterisasi data obat dengan metode KDD dan software excel. Hasil dari menggunakan *K-Means* dibagi menjadi 2 kluster, kluster 1 dengan anggota 81 data obat yang memiliki label penjualan rendah, sedangkan kluster 2 terdiri dari 418 data obat yang memiliki label penjualan tinggi. Hasil klusterisasi dengan menggunakan algoritme dievaluasi dengan menggunakan *davies bouldin index* dan grafik *silhouette coefficient* dan menghasilkan titik kluster optimal di $C=2$ dengan nilai DBI 0.8148198.

Kata kunci: *Klusterisasi, K-Means, Obat*

1. PENDAHULUAN

Kesehatan adalah aspek yang penting dalam kehidupan manusia. Tanpa kesehatan yang baik, manusia akan mengalami hambatan dalam menjalankan aktivitas sehari-hari secara optimal. Oleh karena itu, isu-isu kesehatan selalu menjadi fokus perhatian di seluruh dunia. Salah satu cara untuk meningkatkan kesehatan adalah dengan meningkatkan kualitas layanan kesehatan. Pelayanan kesehatan menjadi tempat di mana masyarakat berkonsultasi mengenai kesehatan mereka. Salah satu aspek penting dalam meningkatkan layanan kesehatan adalah pengelolaan obat yang efektif [1]. Apotek Kidangrangga, yang terletak di kecamatan Rengasdengklok Kabupaten Karawang, telah menggunakan sistem yang saling terhubung untuk mengelola data, termasuk data obat. Namun, setelah mewawancarai pemilik apotek tersebut, diketahui bahwa sering terjadi kelebihan dan kekurangan obat di Apotek Kidangrangga.

Data Mining adalah penerapan otomatis teknik analisis data untuk mengungkapkan hubungan yang sebelumnya tidak terdeteksi antara item-item data. Proses ini digunakan untuk menemukan pola data yang sama dari berbagai bidang seperti keuangan, industri ritel, ilmu pengetahuan, statistik, ilmu kedokteran, kecerdasan buatan, dan pengelompokan data. Tujuannya adalah untuk memahami pola universal yang terdapat dalam data. Selain itu, deteksi anomali transaksi juga penting untuk mengidentifikasi tindak lanjut yang perlu diambil. Semua upaya ini bertujuan

untuk mendukung kegiatan operasional perusahaan dan mencapai tujuan akhir yang diharapkan [2].

Klusterisasi merupakan teknik yang digunakan untuk mengelompokkan rekaman dalam database berdasarkan kondisi tertentu. Konsep dasar dari klusterisasi adalah menggabungkan beberapa objek ke dalam cluster, di mana cluster yang optimal adalah yang memiliki tingkat kemiripan yang tinggi antara objek-objek dalam cluster tersebut dan tingkat perbedaan yang tinggi dengan objek-objek di cluster lainnya [3]. Selain itu klusterisasi merupakan teknik analisis data yang digunakan untuk mengelompokkan data dengan karakteristik atau pola yang serupa ke dalam kelompok atau kluster. Dalam bidang farmasi, klusterisasi dapat digunakan untuk mengelompokkan obat-obatan yang tersedia di apotek berdasarkan kriteria tertentu, seperti efek samping, indikasi, atau jenis obat. Hal ini dapat membantu apotek dalam mengoptimalkan stok obat dan memberikan pelayanan yang lebih baik kepada konsumen.

Algoritma K-Means merupakan algoritma pembelajaran tanpa pengawasan yang dapat mengelompokkan data ke dalam K kluster berdasarkan jarak antara data dan pusat kluster. algoritma ini memiliki keunggulan dalam kecepatan dan efisiensi dalam mengolah data yang besar. Algoritma K-Means adalah algoritma yang digunakan untuk analisis pengelompokan dengan tujuan membagi n objek pengamatan ke dalam k kelompok atau cluster. Pada algoritma K-Means, setiap objek pengamatan ditempatkan dalam kelompok yang memiliki mean

terdekat, dan algoritma melakukan iterasi untuk mencari pusat dari setiap kelompok dalam data [4].

Penelitian ini bertujuan untuk menerapkan algoritma K-Means dalam klasterisasi data obat di apotek Kidangrangga. Dalam penelitian ini, dilakukan analisis terhadap data obat yang tersedia di apotek dan kemudian dilakukan klasterisasi menggunakan algoritma K-Means, dengan adanya penelitian ini diharapkan dapat mengimplementasikan penerapan data mining pada penjualan obat di apotek Kidangrangga guna mendapatkan sumber informasi mengenai ketersediaan stok obat yang optimal.

2. TINJAUAN PUSTAKA

2.1. Data

Data merupakan hasil pencatatan peneliti dalam bentuk kata-kata maupun angka. Pendapat lain menyatakan bahwa data merujuk pada semua hasil pengukuran atau observasi yang telah dicatat untuk tujuan tertentu. Data dapat berarti fakta dari objek yang diamati, baik dalam bentuk angka maupun kata-kata. Dalam konteks statistika, data merupakan fakta-fakta yang digunakan sebagai dasar untuk membuat kesimpulan [5].

2.2. Obat

Obat adalah zat atau campuran zat, termasuk produk biologi, yang digunakan untuk mempengaruhi atau menyelidiki sistem fisiologi atau keadaan patologi dalam rangka menetapkan diagnosis, mencegah, menyembuhkan, memulihkan, meningkatkan kesehatan, dan sebagai kontrasepsi bagi manusia [6].

2.3. Data Mining

Data Mining adalah penerapan otomatis teknik analisis data untuk mengungkapkan hubungan yang sebelumnya tidak terdeteksi antara item-item data. Proses ini digunakan untuk menemukan pola data yang sama dari berbagai bidang seperti keuangan, industri ritel, ilmu pengetahuan, statistik, ilmu kedokteran, kecerdasan buatan, dan pengelompokan data [7].

2.4. Algoritme K-Means

Algoritme K-Means adalah algoritma yang digunakan untuk analisis pengelompokan dengan tujuan membagi n objek pengamatan ke dalam k kelompok atau cluster. Pada algoritma K-Means, setiap objek pengamatan ditempatkan dalam kelompok yang memiliki mean terdekat, dan algoritma melakukan iterasi untuk mencari pusat dari setiap kelompok dalam data [8].

2.5. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah istilah yang sering digunakan untuk menggambarkan proses penggalian informasi yang tersembunyi dalam basis data besar. Proses KDD meliputi beberapa tahapan, yaitu pemilihan data, pembersihan data, transformasi data, data mining, dan

interpretasi/evaluasi. Tahap pemilihan data melibatkan seleksi data dari basis data operasional dan menyimpan data yang dipilih dalam sebuah berkas terpisah. Tahap pembersihan data melibatkan proses membuang duplikasi, memeriksa inkonsistensi dan kesalahan pada data [9].

2.6. Silhouette Coefficient

Silhouette Coefficient merupakan metode validasi cluster internal yang Metode ini merupakan sebuah metode evaluasi cluster yang menggabungkan metode *cohesion* dan *separation*. *Cohession* diukur dengan cara menghitung semua objek yang ada dalam suatu cluster, sedangkan *separation* diukur dengan menghitung jarak rata-rata antara setiap objek dalam sebuah cluster dengan cluster terdekatnya. Jarak antara data dihitung menggunakan rumus *euclidean distance* [10].

2.7. Davies-Bouldin Index

Akurasi diukur menggunakan indeks *Davies-Bouldin* dan faktor siluet. Ini merupakan tanda yang baik dalam pengelompokan karena nilai Indeks *Davies-Bouldin* (DBI) mendekati 0 dan tidak negatif. Nama ini menggunakan kedua nama tersebut, yaitu Indeks *Davies-Bouldin* (DBI), untuk evaluasi pengelompokan. DBI memiliki skema penilaian pengelompokan internal. Skor pengelompokan dianggap baik atau buruk berdasarkan jumlah dan kedekatan skor pengelompokan [11].

2.8. R Studio

R Studio dan bahasa pemrograman R, yang merupakan bahasa pemrograman untuk komputasi statistik dan grafik. *RStudio* adalah sebuah Integrated Development Environment (IDE) yang digunakan untuk pengembangan dan analisis data dengan menggunakan bahasa pemrograman R. IDE ini menyediakan antarmuka pengguna yang intuitif dan berbagai fitur yang memudahkan pengguna dalam menulis, menjalankan, dan menganalisis kode R. *RStudio* merupakan perangkat lunak sumber terbuka (*open-source*) dan tersedia dalam beberapa versi, termasuk *RStudio Desktop*, *RStudio Server*, *RStudio Connect*, dan *RStudio Cloud* [12].

3. METODE PENELITIAN

Metodologi penelitian yang digunakan dalam penelitian ini adalah metodologi knowledge discovery in Database (KDD). Tahapan dari metodologi menggunakan metode KDD adalah *data selection*, *data preprocessing*, *data transformation*, *data mining*, *knowledge interpretation/evaluation*. Adapun tahapan KDD yang digunakan yaitu:

1. Data Selection

Pada tahap awal ini, data-data pada database akan diseleksi kemudian disimpan didalam berkas yang terpisah.

2. Pre-processing/Cleaning

Tahapan berikutnya adalah melakukan pembersihan terhadap data-data yang sudah diperoleh sebelum dilakukan proses data mining.

3. Transformation

Tahapan ini merupakan suatu proses dimana akan dilakukannya transformasi atau perubahan bentuk data terhadap data-data yang sudah di bersihkan sehingga nantinya data tersebut dapat dilanjutkan ke dalam proses data mining.

4. Data Mining

Di dalam tahapan ini akan dilakukan proses pencarian pola atau informasi dengan menggunakan metode atau teknik tertentu terhadap data-data.

5. Interpretation/Evaluation

Tahap ini merupakan tahap terakhir yang dimana akan dilakukannya proses evaluasi terhadap hasil dari proses data mining yang sudah dilakukan.

4. HASIL DAN PEMBAHASAN

Hasil dari penelitian ini yang telah dilakukan adalah klasterisasi data obat di apotek kidangranga dengan menggunakan algoritma *K-Means* dan evaluasi klaster internal yaitu Silhouette Coefficient dan Davies-Bouldin Index.

1. Data Selection

Data Selection merupakan proses pemilihan data. Data yang diambil data list stok apotek kidangranga setiap bulan pada tahun 2022. Pengambilan data obat ini dikarenakan di dalam data tersebut terdapat data obat yang paling terbaru yang bisa diperoleh. Data yang didapat dari sistem masih dalam bentuk excel sehingga perlu dilakukan konversi dan dilakukan seleksi data obat. Contoh data obat yang masih berbentuk excel ditunjukkan oleh gambar 1.

LIST STOCK OF NAME									
KODE	DESKRIPSI	UNIT	HARGA JUAL	KURS	DEPT	GLDANG	GROUPA	QTY	
OKT0001	ACIFAR SALEP	TUBE	Rp. 7.000,00	IDR	Head Quarter	Head Quarter	No Group	155	
OKT2001	ACLOVIR KIRM 5 GR	TUBE	Rp. 12.430,00	IDR	Head Quarter	Head Quarter	No Group	210	
OKT0015	ACLOVIR 5% 5 GR	TUBE	Rp. 4.500,00	IDR	Head Quarter	Head Quarter	No Group	186	
OKT2204	ACLOVIR CREAM 5 GR (FM)	TUBE	Rp. 4.500,00	IDR	Head Quarter	Head Quarter	No Group	161	
OKT2337	ACLOVIR CREAM 5 GR (KF)	TUBE	Rp. 5.000,00	IDR	Head Quarter	Head Quarter	No Group	135	
OKT0001	ALBOTHYL CONC 10 ML	FILES	Rp. 49.200,00	IDR	Head Quarter	Head Quarter	No Group	124	
OKT2961	ALBOTHYL CONC 3ML	FILES	Rp. 29.400,00	IDR	Head Quarter	Head Quarter	No Group	199	
OKT0006	ALOCALIR PLUS GEL	Pcs	Rp. 110.860,00	IDR	Head Quarter	Head Quarter	No Group	141	
OKT0007	ALOCALIR PLUS MOUTHWASH	Pcs	Rp. 120.890,00	IDR	Head Quarter	Head Quarter	No Group	201	
OKT0008	ALOCALIR PLUS ORAL RINSE 60 ML	Pcs	Rp. 124.850,00	IDR	Head Quarter	Head Quarter	No Group	104	
OKT0009	ALOCALIR SPRAY 150ML	Pcs	Rp. 124.830,00	IDR	Head Quarter	Head Quarter	No Group	105	
OKT0006	APOLAR - N CREAM 15 GR	TUBE	Rp. 50.360,00	IDR	Head Quarter	Head Quarter	No Group	34	
OKT0002	APOLAR CREAM 10 GR	TUBE	Rp. 46.200,00	IDR	Head Quarter	Head Quarter	No Group	3	
OKT0122	BACTODERM 10GR	TUBE	Rp. 98.120,00	IDR	Head Quarter	Head Quarter	No Group	107	
OKT0123	BACTODERM 5GR	TUBE	Rp. 60.090,00	IDR	Head Quarter	Head Quarter	No Group	1	
OKT0127	BENOSON - N CREAM 5 GR	TUBE	Rp. 74.690,00	IDR	Head Quarter	Head Quarter	No Group	1	
OKT0128	BENOSON - G CREAM 10 GR	TUBE	Rp. 95.260,00	IDR	Head Quarter	Head Quarter	No Group	244	
OKT0129	BENOSON - M CREAM 5 GR	TUBE	Rp. 88.000,00	IDR	Head Quarter	Head Quarter	No Group	187	
OKT0130	BENOSON - N CREAM 15 GR	TUBE	Rp. 89.180,00	IDR	Head Quarter	Head Quarter	No Group	143	

Gambar 1. Contoh Data Obat

Data yang akan dipilih dari data list stock opname dengan rentang waktu bulan Januari hingga Desember tahun 2022 adalah data obat yang dijual pada apotek Kidangranga. Data tersebut akan dilakukan analisis untuk dapat memperoleh informasi dan mengelompokkan data obat. Total jumlah data obat dari bulan Januari hingga Desember adalah sebanyak 5.988 data yang terbagi ke dalam 499 jenis obat berdasarkan rekapan bulanan. Selanjutnya dilakukan analisis atau pendeskripsian pada data dengan mendeskripsikan format serta keterangan atribut yang

merujuk pada informasi dari data obat. Data dideskripsikan menjadi atribut, tipe data, dan keterangan. Contoh tabel jenis atribut dan keterangannya ditunjukkan pada tabel 1.

Tabel 1. Jenis Atribut dan Keterangannya

Atribut	Tipe Data	Keterangan
Kode	Numerik	Menunjukkan kode obat
Nama Obat	Nominal	Menunjukkan nama obat
Unit	Nominal	Keterangan jenis satuan pada obat
Harga Jual	Numerik	Harga jual perobat
Kurs	Nominal	Mata uang
Dept	Nominal	Supplier obat
Quantity	Numerik	Satuan obat

2. Data Preprocessing/Cleaning

Setelah proses pengumpulan data obat tahapan selanjutnya yaitu data preprocessing pada tahap ini dilakukan proses pembersihan pada data obat yang menjadi fokus penelitian. Terdapat beberapa tahap dalam cleaning yaitu tahap penggabungan data, tahap pemeriksaan missing value, dan tahap penghapusan data-data yang tidak di perlukan. Berikut contoh gambar hasil dari *preprocessing* ditunjukkan pada gambar 2.

NAMA OBAT	JANUARI	FEBRUARI	MARET	APRIL	MAY	JUNI	JULI	AUGUSTUS	SEPTEMBER	OCTOBER	NOVEMBER	DECEMBER
1 ACIFAR SALEP	155	77	109	35	34	225	236	49	90	230	60	46
2 ACLOVIR KIRM 5 GR	210	54	74	126	234	66	241	139	131	150	90	90
3 ACLOVIR 5% 5 GR	186	228	46	125	94	122	41	62	216	179	0	0
4 ACLOVIR CREAM 5 GR (FM)	161	5	1	2	4	3	7	0	0	1	4	6
5 ACLOVIR CREAM 5 GR (KF)	135	113	43	31	125	209	136	95	231	130	120	104
6 ALBOTHYL CONC 10 ML	124	50	36	106	246	66	11	59	6	231	13	200
7 ALBOTHYL CONC 3ML	199	81	85	122	2	216	14	112	182	80	135	3
8 ALOCLAIR PLUS GEL	141	50	153	56	37	244	14	151	21	238	64	36
9 ALOCLAIR PLUS MOUTHWASH	201	93	95	95	39	70	211	17	220	174	26	217
10 ALOCLAIR PLUS ORAL RINSE 60 ML	104	140	126	106	165	56	46	170	215	7	188	249
11 ALOCLAIR SPRAY 150ML	105	47	29	10	23	33	143	69	79	188	156	163
12 APOLAR - N CREAM 15 GR	34	238	150	54	157	250	54	199	173	187	34	16
13 APOLAR CREAM 10 GR	3	2	1	0	0	0	7	0	2	4	2	1
14 BACTODERM 10GR	107	117	57	60	225	236	230	223	199	205	237	181
15 BACTODERM 5GR	1	5	4	2	0	9	3	0	0	3	1	1
16 BENOSON - N CREAM 5 GR	1	3	2	1	0	0	1	0	2	9	10	8
17 BENOSON - G CREAM 10 GR	244	211	13	140	123	145	107	204	230	98	145	4
18 BENOSON - M CREAM 5 GR	187	9	57	72	133	169	240	34	157	8	129	202
19 BENOSON - N CREAM 15 GR	143	152	159	133	136	147	201	226	54	197	95	106
20 BENOSON - N CREAM 5 GR	161	68	125	35	17	75	209	119	3	96	136	17
21 BENOSON CREAM 5GR	119	209	55	123	56	106	133	95	23	52	62	203
22 BENZOJAC 2,5% CREAM 5 GR	3	4	0	5	5	0	2	3	3	2	1	5

Gambar 2. Hasil Preprocessing

Dalam proses gambar 2. diatas, atribut yang tidak relevan dan tidak konsisten juga akan dibuang. Penggabungan data obat ini didasarkan pada pemakaian obat atau jumlah obat yang keluar (Qty keluar) serta nama obat. Serta dilakukan penghapusan beberapa atribut yaitu atribut kode, unit, harga jual, kurs, dan dept. Atribut tersebut dihapus karena dianggap tidak relevan dengan penelitian. Tujuan dari langkah ini adalah agar hasil klasterisasi dapat mencapai hasil yang maksimal dalam mengelompokkan data obat berdasarkan pemakaian obat. Hasil dari penggabungan data obat tersebut terlihat dalam gambar 2, dengan total 5.988 data obat. Data obat yang berhasil digabungkan.

3. Transformation Data

Transformasi data merupakan proses pengubahan data ke dalam suatu format sehingga dapat dilanjutkan ke dalam proses data mining. Data yang sudah di bersihkan kemudian dilakukan transformasi data terlebih dahulu sebelum masuk ke tahap data mining. Adapun proses transformasi data yang akan dilakukan adalah menggunakan fungsi penskalaan (scale) yang mana nilai dari masing-masing jumlah penjualan pada tiap bulan akan dilakukan penskalaan agar tidak terdapat rentang nilai yang terlalu jauh yang dapat mempengaruhi kualitas cluster. Berikut contoh gambar hasil dari *Transformation Data* ditunjukkan pada gambar 3.

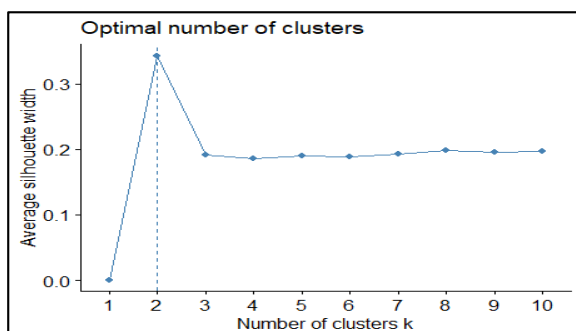
	JANUARI	FEBRUARI	MARET	APRIL	MAY	JUNI	JULI	AUGUSTUS	SEPTEMBER	OCTOBER	NOVEMBER	DESEMBER
1	5.40380302	-0.41155070	0.39767619	-0.64500916	-0.50212750	1.42944040	1.81855249	-0.76367472	-0.22060021	1.65807961	-0.51949720	-0.17470203
2	1.50034202	-0.73602191	-0.16406980	1.80000304	1.63032040	-0.67500703	1.67000000	0.28403045	0.28907081	0.60901780	-0.24400870	-0.18101487
3	0.87400890	1.49147142	-0.58136740	1.54010166	-0.13504940	0.07500708	-0.01040902	-0.33710443	1.34941620	0.88709030	-0.27803300	-0.33626261
4	0.28270388	-0.31214950	-0.33071497	-0.38718417	-0.26819147	-0.08890400	-0.30350004	-0.38950010	-0.33941787	0.88912000	-0.33041013	-0.26207407
5	0.02000012	0.01040507	-0.00161070	-0.01040507	0.00040407	1.78404740	0.38000000	0.01000000	0.01000000	0.01000000	0.01000000	0.01000000
6	0.01040507	-0.01040507	-0.01040507	0.01040507	0.01040507	-0.01040507	-0.01040507	-0.01040507	-0.01040507	-0.01040507	-0.01040507	-0.01040507
7	1.04000000	-0.00000000	0.01000000	1.00000000	0.00000000	1.00000000	0.01000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
8	0.00000000	-0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
9	1.00000000	-0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
10	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
11	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
12	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
13	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
14	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
15	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
16	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
17	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
18	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
19	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
20	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000
21	-0.16200000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000	0.00000000

Gambar 3. Hasil *Transformation Data* obat

Proses transformasi menggunakan fungsi scale dilakukan hanya pada tabel yang bersifat numerik.

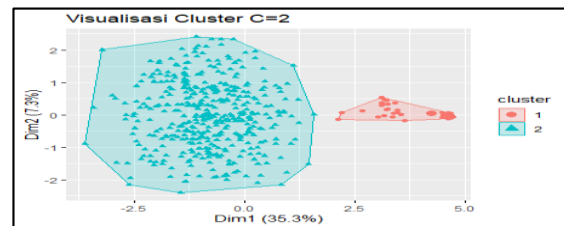
4. Data Mining

Tahap Data Mining akan dilakukan proses pengolahan data dan pencarian pola atau informasi menggunakan teknik clustering dengan menggunakan algoritme K-Means dan Pengolahan data menggunakan software R studio. Sebelum masuk ke dalam proses data mining, dilakukan pencarian jumlah cluster optimal menggunakan grafik *silhouette coefficient*. Berdasarkan grafik dapat dilihat pada gambar 4 bahwa cluster optimal terletak pada titik C=2 yang artinya cluster dinyatakan optimal jika dibagi ke dalam 2 cluster.



Gambar 4. Grafik *Cluster Optimal*

Setelah mengetahui jumlah cluster optimal tahapan selanjutnya adalah proses data mining dengan menggunakan algoritma *k-means*.



Gambar 5. Visualisasi Cluster C1 dan C2

Hasil clustering disajikan dalam bentuk tabel agar lebih mudah dianalisis dan dipahami. Berikut contoh tabel hasil klastering penjualan tinggi dan rendah ditunjukkan pada tabel 2.

Tabel 2. Contoh Hasil Klastering Penjualan Tinggi Dan Rendah

C1	C2
ACYCLOVIR CREAM 5 GR (FM)	ACIFAR SALEP
APOLAR CREAM 10 GR	ACLOVIR KRIM 5 GR
BACTODERM 5GR	ACYCLOVIR 5% 5 GR
BAYCUTEN - N CREAM 5 GR	ACYCLOVIR CREAM 5 GR (KF)
BENZOLAC 2,5 % CREAM 5 GR	ALBOTHYL CONC 10 ML
BERSOL CREAM 15 GR	ALBOTHYL CONC 5ML
BORAGINOL N KRIM 15	ALOCLAIR PLUS GEL
CENTABIO CR 20GR	ALOCLAIR PLUS MOUTHWASH
CLOBETASOL CR 10GR	ALOCLAIR PLUS ORAL RINSE 60 ML
DERMATIX ACNE SPOT CARE	ALOCLAIR SPRAY 15ML
DERMATIX ULTRA 5 GR GEL	APOLAR - N CREAM 10 GR
DESOLEX CREAM 10 GR	BACTODERM 10GR
DIPROSONE- OV CREAM 5 GR	BENOSON - G CREAM 10 GR
FLAMAR GEL	BENOSON - M CREAM 5 GR
FLUCORT N CREAM 5GR	BENOSON - N CREAM 15 GR
FUTADERM OINT	BENOSON - N CREAM 5 GR
GARAMYCIN CREAM 5 GR	BENOSON CREAM 5GR
GENALTEN SK	BENZOLAC 5 % CREAM 5 GR
GENTASOLON CREAM 5 GR	BENZOLAC CL CREAM
IFLACORT CR 5GR	BERSOL CREAM 5 GR

Setelah data dikelompokkan ke dalam tabel dengan atribut C1 dan C2, langkah selanjutnya adalah

pelabelan yang didasarkan pada karakteristik cluster. Berikut gambar karakteristik dari cluster yang telah terbentuk ditunjukkan pada gambar 6.

Cluster	JANUARI	FEBRUARI	MARET	APRIL	MAY	JUNI	JULI	AGUSTUS	SEPTEMBER	OKTOBER	NOVEMBER	DESEMBER
1	10.9	11.6	13.7	12.4	12.2	10.7	10.5	12.7	15.2	11.7	12.9	12.0
2	137	131	97.9	74.6	122	136	125	127	126	128	128	122

Gambar 6. Hasil penentuan Cluster

Berdasarkan data diatas dapat diambil kesimpulan bahwa cluster 1 merupakan cluster dengan tingkat penjualan rendah karena memiliki rata-rata yang rendah. Sedangkan cluster 2 merupakan cluster dengan tingkat penjualan tinggi karena memiliki rata-rata yang tinggi.

5. Knowledge Interpretation/Evaluation

Pada tahap terakhir ini hasil dari proses mining akan divisualisasikan kedalam bentuk grafik. Adapun evaluasi dilakukan dengan menggunakan validasi cluster menggunakan Davies Bouldin Index (DBI). Pada tahap ini akan dilakukan evaluasi untuk mengukur hasil atau kualitas klasterisasi yang telah dilakukan. Berikut hasil tabel nilai evaluasi cluster DBI ditunjukkan pada tabel 3.

Tabel 3. Hasil Nilai Ealuasi Cluster DBI

Jumlah Cluster	Nilai DBI
2	0.8148198
3	2.923235
4	2.804235
5	2.74096
6	2.713477
7	2.670065
8	2.599761
9	2.472374
10	2.45956

Berdasarkan nilai DBI pada masing-masing cluster dapat dilihat bahwa nilai DBI yang mendekati paling rendah dan mendekati 0 terletak pada C=2. Maka dari itu cluster C=2 berdasarkan *silhouette coefficient* dan pengujian nilai DBI merupakan jumlah cluster yang paling optimal.

5. KESIMPULAN DAN SARAN.

Dari hasil penelitian yang telah dilakukan yaitu klasterisasi data obat menggunakan data dari apotek kidangrangga maka terdapat beberapa hal yang dapat disimpulkan yaitu Algoritme K-Means dapat diimplemetasikan dengan baik untuk melakukan klasterisasi data obat dengan metode KDD dan software excel, R Studio serta menggunakan bahasa pemrograman R. Hasil dari klasterisasi dengan menggunakan K-Means mengelompokkan obat menjadi 2 klaster, klaster 1 dengan anggota 81 data obat yang memiliki label penjualan rendah, sedangkan klaster 2 terdiri dari 418 data obat yang memiliki label penjualan tinggi. Hasil klasterisasi dengan

menggunakan algoritme K-Means dievaluasi dengan menggunakan *davies bouldin index* dan grafik *silhouette coefficient* dan menghasilkan titik klaster optimal di C=2 dengan nilai DBI 0.8148198. Saran untuk penelitian ini, Untuk mencapai hasil klasterisasi yang lebih optimal, diperlukan penggunaan algoritma-algoritma terbaik yang tersedia. Selain itu, peningkatan kualitas klasterisasi dapat dicapai dengan memperluas jumlah data yang digunakan dalam proses klasterisasi. Dengan menambah data, model akan lebih akurat dalam mengidentifikasi pola dan hubungan antara entitas.

DAFTAR PUSTAKA

- [1] M. R. Nugroho, I. E. Hendrawan, and P. P. Purwantoro, "Penerapan Algoritma K-Means Untuk Klasterisasi Data Obat Pada Rumah Sakit ASRI," *Nuansa Inform.*, vol. 16, no. 1, pp. 125–133, 2022, doi: 10.25134/nuansa.v16i1.5294.
- [2] W. Purba, W. Siawin, and . H., "Implementasi Data Mining Untuk Pengelompokkan Dan Prediksi Karyawan Yang Berpotensi Phk Dengan Algoritma K-Means Clustering," *J. Sist. Inf. dan Ilmu Komput. Prima(JUSIKOM PRIMA)*, vol. 2, no. 2, pp. 85–90, 2019, doi: 10.34012/jusikom.v2i2.429.
- [3] M. Herviany, S. Putri Delima, T. Nurhidayah, and K. Kasini, "Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokkan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 1, no. 1, pp. 34–40, 2021, doi: 10.57152/malcom.v1i1.60.
- [4] N. Nurahman, A. Purwanto, and S. Mulyanto, "Klasterisasi Sekolah Menggunakan Algoritma K-Means berdasarkan Fasilitas, Pendidik, dan Tenaga Pendidik," *MATRIK J. Manajemen, Tek. Inform. dan Rekayasa Komput.*, vol. 21, no. 2, pp. 337–350, 2022, doi: 10.30812/matrik.v21i2.1411.
- [5] Makbul, "Metode Pengumpulan Data Dan Instrumen Penelitian," *Skripsi*, no. 2, p. 6, 2021.
- [6] H. Hairunnisa, "Sulitnya Menemukan Obat Baru di Indonesia," *Farmasetika.com (Online)*, vol. 4, no. 1, p. 16, 2019, doi: 10.24198/farmasetika.v4i1.22517.
- [7] J. Han, J. Pei, and M. Kamber, *Data mining: concepts and techniques*. Elsevier, 2011.
- [8] A. Nur Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput.*, pp. 1–12, 2016, [Online]. Available: <https://ilmukomputer.org/category/datamining/>
- [9] R. H. I. Risa Helilintar, Intan Nur Farida, "Penerapan metode k-means clustering pada data penerimaan mahasiswa baru.," *Senat. Semin. Nas. Teknol. Inf. dan Komun. Progr. Stud. Tek. Inform. - Univ. pgri madiun . H. I.*, vol. 14, pp. 14–20, 2018, [Online]. Available: [https://doc-pak.undip.ac.id/8488/1/Impression Management Pustakawan Dinas Kearsipan dan Perpustakaan](https://doc-pak.undip.ac.id/8488/1/Impression%20Management%20Pustakawan%20Dinas%20Kearsipan%20dan%20Perpustakaan)

- Provinsi Jawa Tengah.pdf
- [10] S. Paembonan and H. Abduh, "Penerapan Metode Silhouette Coefficient untuk Evaluasi Clustering Obat," *PENA Tek. J. Ilm. Ilmu-Ilmu Tek.*, vol. 6, no. 2, p. 48, 2021, doi: 10.51557/pt_jiit.v6i2.659.
- [11] Abdussalam Amrullah, Intam Purnamasari, Betha Nurina Sari, Garno, and Apriade Voutama, "Analisis Cluster Faktor Penunjang Pendidikan Menggunakan Algoritma K-Means (Studi Kasus: Kabupaten Karawang)," *J. Inform. dan Rekayasa Elektron.*, vol. 5, no. 2, pp. 244–252, 2022, doi: 10.36595/jire.v5i2.701.
- [12] R. P. Primanda, A. Alwi, and D. Mustikasari, "DATA MINING SELEKSI SISWA BERPRESTASI UNTUK MENENTUKAN KELAS UNGGULAN MENGGUNAKAN METODE K-MEANS CLUSTERING (Studi Kasus di MTS Darul Fikri)," *Komputek*, vol. 5, no. 1, p. 88, 2021, doi: 10.24269/jkt.v5i1.686.