

Implementasi Data Mining dengan K-Means Clustering untuk Memprediksi Pengadaan Obat

Putri Pratiwi Pane^{*}, Yusuf Ramadhan Nasution, Mhd. Furqan

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

Email: ^{1,*}putriipane2001@gmail.com, ²ramadhannst@uinsu.ac.id, ³mfurqan@uinsu.ac.id

Email Penulis Korespondensi: putriipane2001@gmail.com

Submitted: 06/02/2024; Accepted: 18/02/2024; Published: 20/02/2024

Abstrak—Pusat Kesehatan Masyarakat (Puskesmas) merupakan salah satu lembaga yang menyediakan pelayanan kesehatan. Untuk memberikan jaminan layanan kesehatan yang berkualitas, pihak puskesmas harus dapat melakukan pengelolaan manajemen stok persediaan obat agar menghindari risiko terjadinya kekurangan atau kelebihan stok obat yang ada. Oleh sebab itu tujuan dilakukannya penelitian ini adalah untuk melihat dan melakukan pengelompokan permintaan obat yang ada di Puskesmas Mandala dengan teknik clustering. Data yang digunakan adalah data permintaan obat yang ada pada puskesmas tersebut dimulai dari bulan Januari – Desember 2023. Dalam penerapannya, digunakan aplikasi atau software RapidMiner untuk melakukan pengelompokan dengan algoritma K-Means Clustering. Data obat yang ada akan dikelompokkan ke dalam 3 cluster, yaitu cluster 0 untuk permintaan obat tinggi, cluster 1 untuk permintaan obat menengah dan cluster 2 untuk permintaan obat rendah. Dari 28 data uji yang digunakan, didapatkan hasil cluster pertama yang terdiri 24 item, cluster kedua yang terdiri dari 3 item serta cluster ketiga terdiri dari 1 item dengan nilai Davies Bouldin Index yang didapatkan sebesar 0,276. Dari hasil penelitian ini, maka pihak Puskesmas dapat terus melakukan pengadaan obat pada jenis obat yang tergolong pada cluster permintaan tinggi agar kebutuhan obat dapat selalu terpenuhi.

Kata Kunci: K-Means; Data Mining; Clustering; Obat-obatan; RapidMiner

Abstract—Community Health Center is one of the institutions that provides healthcare services. To ensure the provision of quality healthcare services, the Community Health Center management must be able to effectively manage medicine inventory to avoid the risks of shortages or excess stock. Therefore, the purpose of this research is to observe and perform clustering of medicine demands at Puskesmas Mandala using the K-Means Clustering technique. The data used includes medicine demand data from January to December 2023 at the health center. In its implementation, the RapidMiner application or software is utilized to perform clustering using the K-Means Clustering algorithm. The available medicine data will be grouped into 3 clusters: cluster 0 for high medicine demands, cluster 1 for moderate medicine demands, and cluster 2 for low medicine demands. Out of the 28 test data used, the results show the first cluster consisting of 24 items, the second cluster consisting of 3 items, and the third cluster consisting of 1 item with a Davies Bouldin Index value of 0.276. From this research, the Puskesmas can continue to procure medicine for the types classified under high-demand clusters to ensure that the medicine needs are consistently met.

Keywords: K-Means; Data Mining; Clustering; Medicine; RapidMiner

1. PENDAHULUAN

Pusat Kesehatan Masyarakat (Puskesmas) merupakan salah satu lembaga pelayanan kesehatan yang memiliki peranan penting dalam menyediakan layanan kesehatan bagi masyarakat [1]. Puskesmas harus dapat memberikan pelayanan kesehatan dan pengobatan pasien dengan efektif dan efisien agar kesehatan masyarakat dapat selalu terjamin. Salah satu upaya yang dapat dilakukan adalah dengan mengelola ketersediaan obat yang dibutuhkan. Pengelolaan ketersediaan obat dilakukan dengan tujuan menghindari kekurangan atau kelebihan stok yang berpotensi mengurangi kualitas pelayanan kesehatan [2].

Perencanaan ketersediaan obat merupakan salah satu aspek penting dalam pengelolaan obat, karena perencanaan kebutuhan obat akan mempengaruhi pengadaan, pendistribusian dan pemakaian obat di unit pelayanan kesehatan [3]. Dengan adanya perencanaan kebutuhan obat yang tepat, maka dapat memberikan efek pengadaan obat menjadi lebih efektif dan efisien sehingga tersedianya obat dengan jenis dan jumlah yang cukup sesuai dengan kebutuhan yang diperlukan. Kekurangan obat memiliki dampak negatif pada perawatan pasien dan berimplikasi pada pembiayaan yang mahal. Kelebihan persediaan obat juga akan menimbulkan masalah bagi rumah sakit. Obat yang menumpuk akan menjadi rusak dan kedaluwarsa, hal ini akan menimbulkan kerugian dimasa mendatang [4].

Untuk menghadapi dinamika permintaan obat yang fluktuatif dan beragam, diperlukan adanya suatu sistem yang dapat menganalisa permintaan obat yang ada pada puskesmas. Permasalahan ini dapat diselesaikan dengan konsep keilmuan *data mining*. Pada dasarnya, *data mining* merupakan proses penggalian suatu data dengan tujuan mendapatkan informasi terbaru dengan menentukan pola atau aturan tertentu sehingga berguna dalam pengambilan keputusan [5]. Sebenarnya jika dilakukan pengolahan lebih dalam terhadap data, maka akan didapatkan suatu informasi yang sangat berguna dan dapat dijadikan sebagai suatu pendukung keputusan [6]. Klastering merupakan salah satu teknik yang ada pada *data mining* yang dapat mengelompokkan data ke dalam kelompok tertentu [7]. Teknik ini dapat memberikan hasil analisis berupa pengelompokan permintaan obat-obatan berdasarkan data historis yang ada.. Algoritma K-Means merupakan algoritma yang paling populer dan banyak digunakan dalam melakukan klastering. Algoritma ini akan digunakan dikarenakan penggunaannya yang

cukup sederhana dan cepat dalam menangani permasalahan klustering dengan penggunaan data numerik dalam ukuran data yang cukup besar dan tersebar [8].

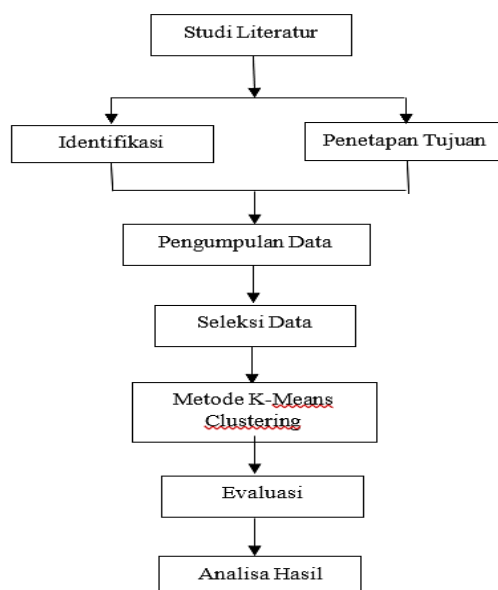
Penelitian dengan menerapkan K-Means Clustering telah banyak dilakukan. Seperti pada penelitian implementasi metode *improved* k-means untuk mengelompokkan film. Hasil analisis yang didapatkan adalah pengelompokan film ke dalam 6 cluster dengan menggunakan atribut atau parameter seperti durasi, jumlah votes dan sutradara. Nilai akurasi yang didapatkan pada penelitian ini adalah sebesar 87.2% yang menunjukkan bahwa metode ini dapat melakukan *clustering* dengan cukup baik [9]. Selanjutnya penelitian yang menerapkan metode K-Means *Clustering* dalam menentukan jenis udang terlaris. Pada penelitian ini digunakan 82 data yang dikelompokkan ke dalam tiga kluster yaitu sebanyak 19 jenis udang kurang laris, 42 jenis udang laris dan 21 jenis udang yang terlaris [10]. Pada penelitian lain menerapkan algoritma K-Means Clustering untuk mengelompokkan data penjualan toko *fashion* hijab. Dari 100 data yang digunakan, didapatkan hasil akhir pengelompokan ke dalam 3 cluster yaitu cluster sangat laris ada 11 item, cluster laris ada 55 item dan 34 item untuk cluster kurang laris [11]. Lalu pada penelitian lain dilakukan penerapan algoritma k-means untuk melakukan pengelompokan penyebaran covid-a9 di provinsi Jawa Barat. Hasil penelitian didapatkan dengan mengelompokkan data ke dalam 3 kluster, yaitu kluster 1 untuk penyebaran paling tinggi sebanyak 2 kota, kluster 2 untuk penyebaran menengah diperoleh 5 kota, dan kluster 3 untuk penyebaran lebih sedikit sebanyak 20 kota [12]. Dan pada penelitian lainnya juga menerapkan algoritma k-means clustering untuk mengelompokkan wilayah produksi tanaman obat ke dalam 3 kluster. Hasil yang didapatkan menunjukkan sebanyak 24 wilayah di cluster 0, 1 wilayah di cluster 1, dan 2 wilayah di cluster 2 [13]. Dari beberapa penelitian yang ada sebelumnya, penelitian ini akan mencoba menerapkan algoritma k-means clustering untuk mengelompokkan data permintaan obat. Berbeda dengan penelitian yang lain, penelitian sebelumnya menerapkan algoritma K-Means Clustering untuk melakukan pengelompokan beberapa permasalahan yang berbeda, seperti jenis udang, jenis film, daerah penyebaran covid19, penjualan pakaian, dan wilayah produksi tanaman obat. Pada penelitian ini akan digunakan untuk mengelompokkan data permintaan obat yang ada pada puskesmas Mandala ke dalam 3 kluster yaitu rendah, sedang dan tinggi.

Manfaat dari klustering permintaan obat ini akan dapat memberikan efisiensi manajemen persediaan, pengoptimalan ruang penyimpanan, dan pengurangan risiko kekurangan atau kelebihan stok. Dengan mengetahui kelompok permintaan obat-obatan yang serupa, lembaga kesehatan dapat mengembangkan strategi pengadaan dan penataan persediaan yang lebih terarah [14]. Dengan demikian, penelitian ini memiliki potensi untuk memberikan kontribusi signifikan dalam konteks manajemen logistik di sektor kesehatan. Dengan adanya penelitian ini, harapannya pihak puskesmas mandala dapat meningkatkan efisiensi operasional, meminimalkan biaya, dan memberikan pelayanan kesehatan yang lebih baik kepada masyarakat.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Penelitian ini dilakukan dengan melakukan beberapa tahapan. Tahapan tersebut dimulai dari studi literatur, identifikasi masalah, penetapan tujuan, pengumpulan data, seleksi data, penerapan algoritma, evaluasi hingga proses analisis hasil.



Gambar 1. Tahapan Penelitian

Gambar 1 merupakan gambaran mengenai tahapan penelitian yang akan dilakukan. Proses awal dalam melakukan penelitian ini adalah melakukan studi literatur yang dilakukan dengan tujuan untuk mengumpulkan beberapa referensi mengenai algoritma yang akan digunakan sehingga peneliti dapat memahami dengan baik algoritma yang akan digunakan. Selanjutnya adalah melakukan pengidentifikasian permasalahan. Dalam penelitian ini ditemukan permasalahan kurangnya strategi manajemen pengelolaan obat yang ada di puskesmas Mandala. Oleh karena itu, peneliti akan mencoba menerapkan algoritma K-Means Clustering untuk melakukan prediksi pengadaan obat di masa yang akan datang dengan melakukan pengelompokan klaster berdasarkan data permintaan obat sebelumnya.

Langkah selanjutnya adalah melakukan pengumpulan data yang dibutuhkan pada penelitian. Pada penelitian ini digunakan data historis permintaan obat yang ada pada puskesmas Mandala selama 1 tahun. Selanjutnya akan dilakukan pengolahan serta penerapan algoritma k-means clustering pada data yang ada untuk mendapatkan hasil klaster permintaan obat. Hasil klaster tersebut kemudian akan dianalisis, dan dapat digunakan oleh pihak puskesmas Mandala untuk melakukan strategi pengadaan obat di masa yang akan datang.

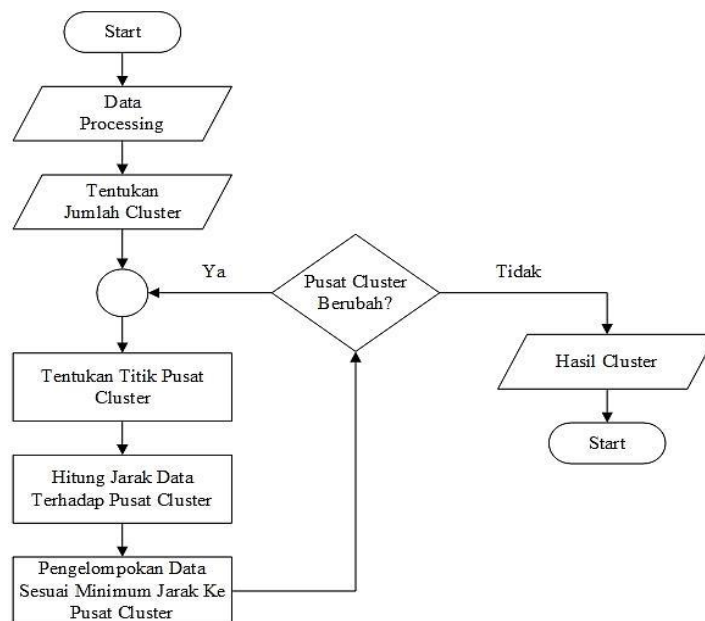
2.2 Data Mining

Data *mining* merupakan suatu konsep yang diterapkan dalam mencari nilai lebih yang tersimpan di dalam database yang besar. Data mining ialah sebuah tahapan semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan machine learning untuk mengekstraksi dan mengidentifikasi informasi pengetahuan yang bermanfaat yang tersimpan didalam database yang besar. Data mining juga mempunyai peran yang sangat penting dalam beberapa bidang dalam kehidupan sehari-hari antara lain ialah bidang industri, bidang cuaca, keuangan dan bidang ilmu dan teknologi[6].

Salah satu teknik data *mining* yang ada adalah teknik klastering. Analisis cluster yaitu menemukan kumpulan objek hingga objek-objek dalam satu kelompok sama (punya hubungan) dengan yang lain dan berbeda (tidak berhubungan) dengan objek-objek dalam kelompok lain. Tujuan dari analisis cluster adalah meminimalkan jarak di dalam cluster dan memaksimalkan jarak antar cluster. Clustering merupakan sebuah proses dalam pembentukan kelompok dari sebuah data yang besar yang sebelumnya tidak diketahui kelompoknya berdasarkan kesamaan karakteristiknya. Berbeda dengan model prediksi atau klasifikasi, dataset yang digunakan untuk melakukan pengklasteran tidak memiliki variabel data target [15].

2.3 K-Means Clustering

K-Means merupakan salah satu metode pengelompokan atau klastering yang berusaha membagi data yang ada ke dalam bentuk dua atau lebih kelompok. *Clustering* merupakan proses membagi objek ke dalam kumpulan objek sejenis yang memiliki kesamaan satu sama lain dan berbeda dengan objek di cluster lain [16]. Dataset yang digunakan untuk melakukan pengklasteran tidak memiliki variabel data target [15]. Algoritma K-Means Clustering merupakan salah satu algoritma yang dapat melakukan *clustering* dengan secara sederhana yang akan melakukan pengelompokan dengan melalui langkah iteratif.



Gambar 2. Diagram Alir Algoritma K-Means Clustering

Proses penerapan algoritma K-Means Clustering diawali dengan proses data *processing*. Tahapan ini dilakukan dengan tujuan untuk mempersiapkan data penelitian yang ada agar dapat diterapkan untuk dilakukan pembelajaran mesin. Proses selanjutnya adalah menentukan nilai cluster atau kelompok. Penentuan nilai kluster

dilakukan dengan tujuan agar *output* yang dihasilkan sesuai dengan banyaknya kluster yang diinginkan. Selanjutnya dilakukan penentuan nilai titik pusat *cluster* dan jarak data terhadap kluster serta pengelompokan data sesuai dengan nilai minimum jarak ke pusat *cluster*. Iterasi algoritma ini akan terus berjalan sampai keadaan dimana pusat *cluster* tidak berubah. Secara umum, cara kerja dari algoritma K-Means clustering adalah sebagai berikut [17]

- a. Menentukan k sebagai jumlah kluster yang diinginkan.
- b. Menentukan nilai random untuk pusat *centroid* (*cluster* awal) sebanyak k.

$$v = \frac{\sum_k^n 1(x_i)}{N}, i = 1, 2, 3, n \tag{1}$$

Nilai v diartikan sebagai nilai *centroid* pada kelompok. Variabel x_i merupakan objek ke-I, di mana nilai objek berubah sesuai dengan iterasi. Dan variabel n adalah banyaknya objek/data/

- c. Menghitung jarak setiap data *input* terhadap masing-masing *centroid* menggunakan rumus jarak *Euclidean Distance*. Berikut adalah persamaan pada rumus *Euclidean Distance* :

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

Nilai d diartikan sebagai nilai jarak data ke pusat *cluster*. Variabel x_i merupakan objek x ke-i dan variabel y_i merupakan data y ke-i. Variabel n menjelaskan banyaknya objek/data

- d. Mengklasifikasikan data berdasarkan kedekatannya dengan *centroid* (jarak terkecil).
- e. Memperbaharui nilai *centroid*. Nilai *centroid* baru diperoleh dari rata-rata *cluster* yang bersangkutan dengan menggunakan rumus 1.
- f. Melakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

3. HASIL DAN PEMBAHASAN

3.1 Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data yang dikumpulkan di puskesmas Mandala. Data yang diambil merupakan data penjualan obat berdasarkan bulan Januari-Desember 2023. Kemudian data tersebut diseleksi dan akan digunakan untuk diolah untuk menentukan prediksi persediaan obat farmasi. Analisis ini dilakukan dengan tujuan melakukan prediksi atau peramalan mengenai hasil – hasil yang telah ditargetkan dengan suatu metode tertentu [18]. Berikut merupakan dataset yang telah dikumpulkan:

Tabel 1. Dataset Penelitian

Nama Obat	Jan	Feb	Mare t	Apr	Mei	Jun	Jul	Ags	Sep	Okt	Nov	Des
Acetylcysteine 200 mg	2305	1855	380	2300	2266	2317	2185	2205	2218	2280	2300	1305
Alcohol Swab	3	1	0	6	7	11	4	8	2	10	5	9
Amitriptyline 25 mg	40	1	0	39	23	52	54	89	78	15	35	30
Antasida tablet	26808	1046	1188	16808	26769	26820	29981	30001	30014	16788	16808	25808
Asam ascorbat (Vit C) 50 mg	33048	2087	60	31048	33009	33060	36457	36477	36490	32980	33000	31048
...
...
...
Sput 3ml / Nipro	271	270	350	268	261	267	261	266	265	264	262	263
Sput 5ml / Nipro	430	359	356	-27	391	342	310	330	343	310	331	230
Surgical Gloves Steril 7,5	230	148	247	1029	1271	1222	1190	1210	1223	1010	1030	1210

Tabel 1 merupakan kumpulan dataset jenis obat yang didapatkan. Pada tabel tersebut dijelaskan terdapat 142 jenis obat dan banyaknya permintaan obat pada setiap bulannya selama satu tahun di Puskesmas Mandala. Sebelum digunakan, data penelitian di atas akan diolah terlebih dahulu agar dapat menghasilkan data penelitian yang baik.

Tabel 2. Data *Integration*

Nama Obat	Rerata Permintaan
Acetylcysteine 200 mg	1993
Alcohol Swab /Serenity	5,5
Amitriptyline 25 mg	38
Antasida tablet	20736,58
Asam ascorbat (Vit C) 50 mg	28230,33
...	...
...	...
...	...
Sput 3ml / Nipro	272,3333
Sput 5ml / Nipro	308,75
Surgical Gloves Steril 7,5 (Ambumax)	918,3333

Data pada tabel 2 merupakan hasil pengolahan data yang melalui tahapan data *integration*. Data *integration* dilakukan dengan menggabungkan data – data permintaan obat pada setiap bulannya ke dalam rerata pemakaian dalam satu tahun. Penggabungan data dilakukan dengan menjumlahkan seluruh permintaan dimulai dari bulan Januari – Desember 2023 kemudian dilakukan perhitungan rerata pemakaian. Sehingga, didapatkan atribut/parameter rerata permintaan. Atribut ini nantinya akan digunakan untuk menerapkan klastering dengan algoritma K-Means.

3.2 Penerapan K-Means Clustering

Langkah awal dalam penerapan algoritma K-Means Clustering adalah menentukan nilai k atau jumlah *cluster*. Dalam penelitian ini, data akan dikelompokkan ke dalam 3 *cluster*, yaitu *cluster* 0 untuk penjualan kategori tinggi, *cluster* 1 untuk penjualan obat menengah, dan *cluster* 2 untuk penjualan kategori rendah. Selanjutnya adalah menentukan nilai *centroid* awal yang ditentukan secara acak. Berikut merupakan hasil penentuan nilai *centroid* awal :

Tabel 3. Data Pusat *Cluster* Iterasi 1

No.	<i>Cluster</i> Pusat	<i>Cluster</i>
1	356,75	C0
2	20102	C1
3	6487,667	C2

Penentuan titik *cluster* pusat pada iterasi 1 dilakukan secara acak yang dipilih dari dataset yang ada. Dapat dilihat pada tabel 3, telah ditentukan nilai *cluster* pusat untuk setiap *cluster*. Diketahui nilai *cluster* pusat C0 adalah 356,75; *cluster* pusat C1 adalah 20102; dan *cluster* pusat C2 adalah 6487,667. Setelah didapatkan titik pusat awal *cluster*, kemudian dilakukan perhitungan jarak *Euclidean* dan mengelompokkan berdasarkan jarak terkecil sehingga akan didapatkan nilai *centroid* baru untuk acuan perhitungan berikutnya sampai nilai *centroid* sebelum dan sesudah bernilai sama.

Hitung *Euclidean Distance* dari semua data ke tiap titik pusat pertama :

$$C0(1) = \sqrt{(356,75 - 1993)^2} = 1636,25$$

$$C0(2) = \sqrt{(356,75 - 5,5)^2} = 351,25$$

$$C0(3) = \sqrt{(356,75 - 38)^2} = 318,75$$

$$C0(4) = \sqrt{(356,75 - 20736,58)^2} = 2039,83$$

$$C0(5) = \sqrt{(356,75 - 28230,33)^2} = 27873,58$$

$$C0(6) = \sqrt{(356,75 - 471,16)^2} = 114,41$$

$$C0(7) = \sqrt{(356,75 - 16930,33)^2} = 16573,58$$

$$C0(8) = \sqrt{(356,75 - 4560,5)^2} = 4203,75$$

$$C0(9) = \sqrt{(356,75 - 5,83)^2} = 350,91$$

$$C0(10) = \sqrt{(356,75 - 4,91)^2} = 351,83$$

.....

$$C0(142) = \sqrt{(356,75 - 918,33)^2} = 561,58$$

Hitung *Euclidean Distance* dari semua data ke tiap titik pusat kedua :

$$C1(1) = \text{sqrt}((20102 - 918,33)^2) = 18109$$

$$C1(2) = \text{sqrt}((20102 - 5,5)^2) = 20096,5$$

$$C1(3) = \text{sqrt}((20102 - 38)^2) = 20064$$

$$C1(4) = \text{sqrt}((20102 - 20736,58)^2) = 634,58$$

$$C1(5) = \text{sqrt}((20102 - 28230,33)^2) = 8128,33$$

$$C1(6) = \text{sqrt}((20102 - 471,16)^2) = 19630,83$$

$$C1(7) = \text{sqrt}((20102 - 16930,33)^2) = 3171,66$$

$$C1(8) = \text{sqrt}((20102 - 4560,5)^2) = 15541,5$$

$$C1(9) = \text{sqrt}((20102 - 5,83)^2) = 20097,08$$

$$C1(10) = \text{sqrt}((20102 - 4,91)^2) = 20090,67$$

.....

$$C1(142) = \text{sqrt}((20102 - 918,33)^2) = 19136,67$$

Hitung *Euclidean Distance* dari semua data ke tiap titik pusat ketiga :

$$C2(1) = \text{sqrt}((6847,66 - 1993)^2) = 4494,66$$

$$C2(2) = \text{sqrt}((6847,66 - 5,5)^2) = 6842,16$$

$$C2(3) = \text{sqrt}((6847,66 - 38)^2) = 6449,66$$

$$C2(4) = \text{sqrt}((6847,66 - 20736,58)^2) = 14248,92$$

$$C2(5) = \text{sqrt}((6847,66 - 28230,33)^2) = 21742,67$$

$$C2(6) = \text{sqrt}((6847,66 - 471,66)^2) = 6016,5$$

$$C2(7) = \text{sqrt}((6847,66 - 16930,33)^2) = 10442,67$$

$$C2(8) = \text{sqrt}((6847,66 - 4560,5)^2) = 1927,16$$

$$C2(9) = \text{sqrt}((6847,66 - 5,83)^2) = 6481,83$$

$$C2(10) = \text{sqrt}((6847,66 - 4,92)^2) = 6482,75$$

.....

$$C2(142) = \text{sqrt}((6847,66 - 272,33)^2) = 6215,33$$

Hasil perhitungan *Euclidean Distance* dapat dilihat pada tabel di bawah ini:

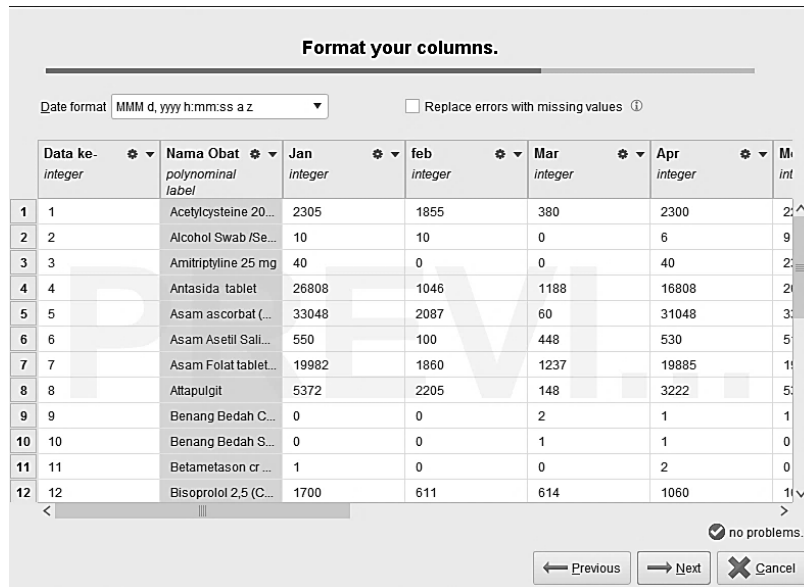
Tabel 4. Perhitungan *Euclidean Distance* Iterasi 1

Nama Obat	C0	C1	C2	Jarak	Cluster
Asam Asetil Salisilat (Asetosal) 80 mg	114,4167	19630,83	6016,5	114,41667	C0
Betametason cr 0,1% (sebagai Valerat)	345,4167	20090,67	6476,334	345,41667	C0
Bisoprolol 5 mg	641,5833	19163,07	5489,334	641,58333	C0
Calcium Lactate 500 mg	1708,667	18036,58	4422,25	1708,6667	C0
Chloramphenicol / Bufacetin salep kulit	257,6667	20002,92	6388,584	257,66667	C0
Codein 10 mg	132,3333	19612,92	5998,584	132,33333	C0
Diphenhydramine Injeksi	337,75	20083	6468,667	337,75	C0
Erythromycine 250 mg	7402,417	12342,83	1271,5	1271,5	C2
Gentamycine /Genoint salep mata 0,3%	272,75	20018	6403,667	272,75	C0
Gentamycine tetes mata 0,3%	297	20042,25	6427,917	297	C0

Tabel 4 menunjukkan hasil perhitungan jarak dengan menerapkan algoritma K-Means Clustering. Nilai *cluster* ditentukan dengan melihat nilai jarak terkecil untuk setiap *cluster*. Sebagai contoh untuk data obat Asam Asetil Salsilat (Asetosal) diketahui nilai *cluster* C0 = 114,4; C1 = 19630,8; dan C2 = 6016,5. Nilai jarak terkecil terletak pada *cluster* 0, sehingga untuk data pada baris pertama nilai *cluster* nya terletak pada C0.

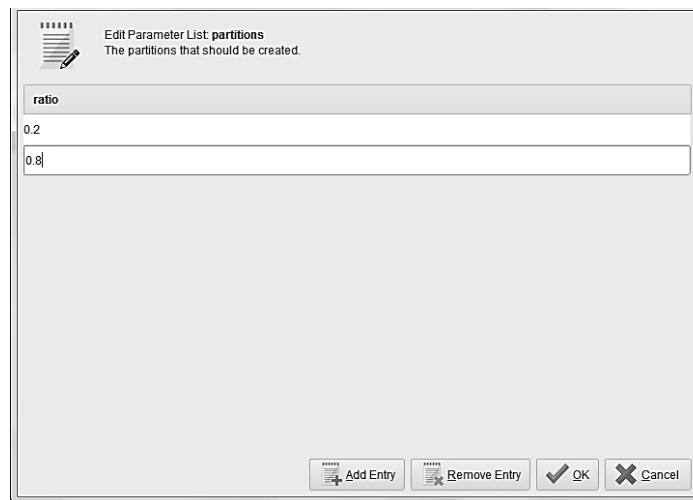
3.3 Proses K-Means Clustering dengan RapidMiner

RapidMiner merupakan salah satu *software data mining* pengolahan data set untuk mencari pola data sesuai dengan tujuan dari pengolahan data tersebut. RapidMiner memungkinkan pengguna untuk dengan mudah mengimpor, mengelola, dan menganalisis data dari berbagai sumber [19]. Langkah awal yang dilakukan untuk menerapkan algoritma K-Means Clustering adalah dengan melakukan *import dataset* penelitian yang digunakan. Berikut merupakan tampilan ketika melakukan *import dataset* pada aplikasi RapidMiner:



Gambar 3. Import Dataset

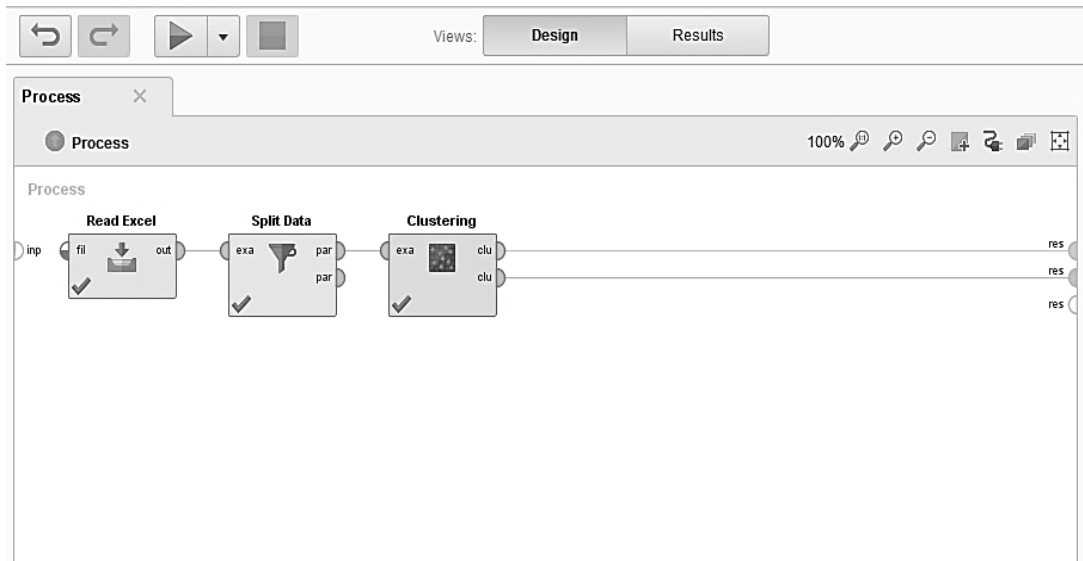
Gambar 3 di atas menunjukkan proses *import* dataset penelitian ke dalam aplikasi RapidMiner. Hal ini dilakukan agar dataset yang telah ada dapat diakses atau digunakan dalam aplikasi sehingga dapat dilakukan proses penerapan algoritma K-Means. Setelah *importing* data, selanjutnya adalah membagi data ke dalam dua bagian yaitu data *training* dan data *testing* dengan rasio perbandingan 80% dan 20%. Sehingga jumlah data *training* yang digunakan sebesar 114 data dan data *testing* sebanyak sebesar 28 data.



Gambar 4. Penggunaan Operator Split Data

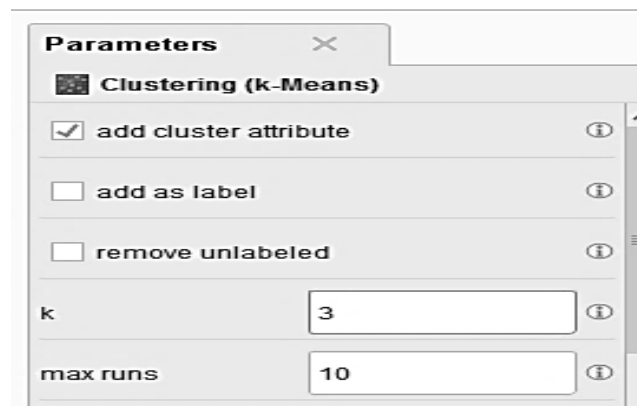
Proses gambar 4 di atas dilakukan untuk melakukan pembagian data pada aplikasi RapidMiner dengan menggunakan operator *split* data. Operator ini digunakan untuk membagi dataset menjadi dua subset data yaitu data *training* dan data *testing*. Nilai rasio merupakan nilai perbandingan untuk melakukan pembagian dataset penelitian. Oleh karena itu nilai rasio diisi nilai 0,2 dan 0,8 agar menghasilkan rasio perbandingan dataset 20:80%.

Kemudian untuk menerapkan algoritma K-Means Clustering digunakan operator Clustering yang telah disediakan di aplikasi RapidMiner. Karena yang digunakan 3 kategori, Selanjutnya melakukan pengaturan algoritma K-Means, yang diatur pada menu Parameter Clustering K Means dengan k=3. Sehingga tampilan model yang dibangun pada aplikasi RapidMiner adalah sebagai berikut:



Gambar 5. Tampilan Rancangan Model Clustering

Gambar 5 di atas merupakan tampilan desain model pada aplikasi RapidMiner untuk melakukan proses clustering. Digunakan tiga operator yaitu *read excel*, *split data*, dan *clustering*. Operator *read excel* digunakan untuk melakukan proses *import* dataset. Operator *split data* dipergunakan untuk membagi data dan operator *clustering* digunakan untuk menerapkan algoritma *clustering* pada data.



Gambar 6. Menentukan Jumlah Cluster

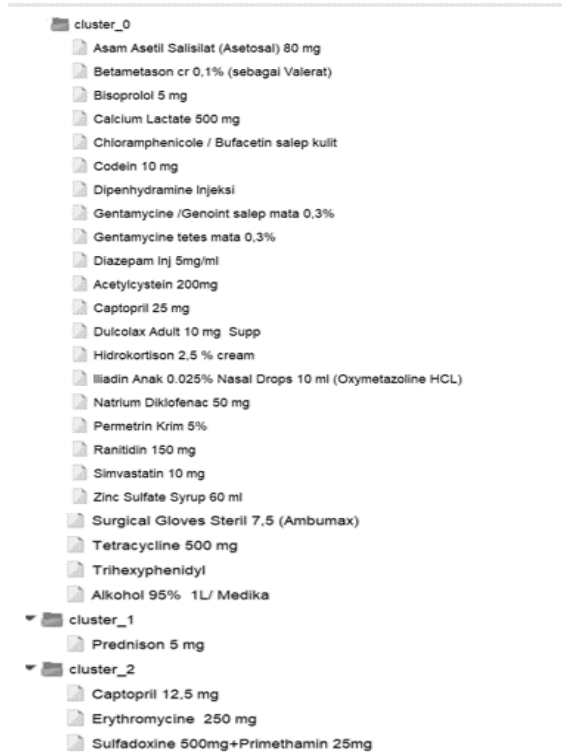
Seperti penjelasan sebelumnya, langkah awal dalam menerapkan proses *clustering* adalah menentukan nilai *cluster*. Gambar 6 menjelaskan proses penentuan nilai parameter *cluster* (*k*) pada aplikasi RapidMiner. Dalam penelitian ini ditentukan nilai *cluster* yaitu tiga. Tampilan output atau hasil yang muncul pada aplikasi RapidMiner dapat dilihat dalam dua bentuk yaitu *example set* dan *text view cluster*. *Example set* merupakan tampilan hasil cluster data secara keseluruhan sesuai dengan data yang telah dimasukkan. Sedangkan *text view cluster* merupakan tampilan ringkasan hasil clustering yang dilakukan.

Cluster Model

```
Cluster 0: 24 items
Cluster 1: 1 items
Cluster 2: 3 items
Total number of items: 28
```

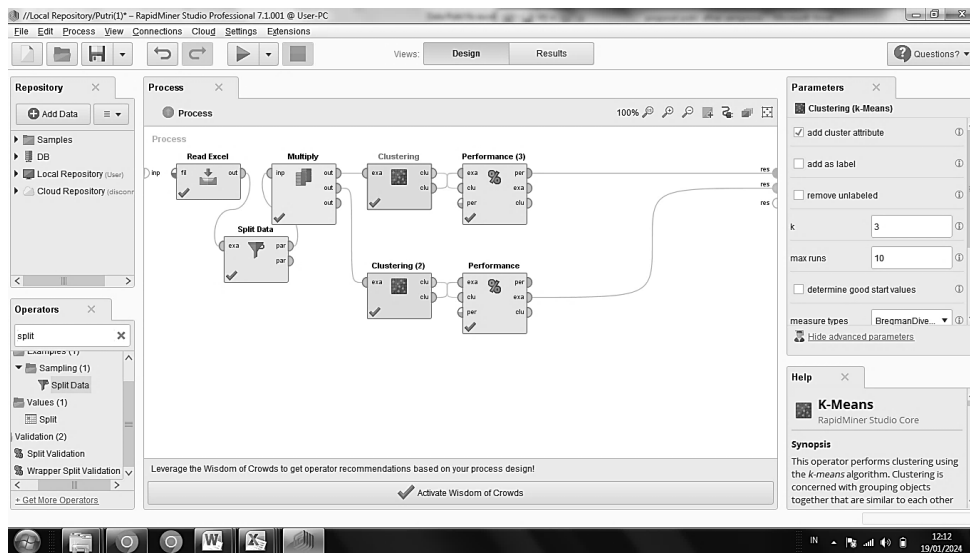
Gambar 7. Tampilan Text View Cluster

Dari gambar 7, diketahui hasil *text view cluster* yang dikelompokkan ke dalam 3 *cluster*. Dari 28 data pengujian, didapatkan sebanyak 24 data jenis obat tergolong dalam *cluster* 0 (C0), 1 data jenis obat tergolong dalam *cluster* 1 (C1) dan 3 data jenis obat tergolong dalam *cluster* 2 (C2)



Gambar 8. Tampilan Hasil Cluster

Gambar 8 menjelaskan mengenai rincian pengklasteran data jenis obat. Dari gambar di atas dapat dilihat jenis obat apa saja yang tergolong ke dalam cluster C0,C1, dan C2. Selanjutnya, untuk memastikan akurasi model klastering yang dihasilkan dilakukan perhitungan nilai evaluasi *Davies Bouldin Index*. Untuk mengukur kinerja model klastering yang ada, maka digunakan operator pembantu yang ada pada aplikasi RapidMiner yaitu *Cluster Distance Performance*. Sehingga desain model akan tampak pada gambar di bawah ini:



Gambar 9. Model Evaluasi *Davies Bouldin Index*

Tampilan gambar di atas merupakan tampilan desain model evaluasi pada aplikasi RapidMiner. Dari gambar 9 dapat dilihat penggunaan berbagai operator yang dipergunakan untuk melakukan proses evaluasi kinerja model klastering yang dihasilkan.

Davies Bouldin

Davies Bouldin: -0.276

Gambar 10. Hasil Evaluasi *Davies Bouldin Index*

Gambar 10 merupakan hasil akhir nilai evaluasi yang dihasilkan pada aplikasi RapidMiner sesuai dengan desain model pada gambar 9. Dapat diketahui bahwa dengan menggunakan parameter *cluster* $k=3$, diperoleh *Davies Bouldin Index* sebesar 0,276. Semakin kecil nilai *davies bouldin index* maka semakin baik *cluster* yang diperoleh dari pengelompokan menggunakan metode *clustering* [20]. Berikut merupakan hasil pengklasteran 28 jenis data obat yang dirangkum kedalam bentuk tabel:

Tabel 5. Hasil Klastering Permintaan Obat

Data ke-	Nama Obat	Kategori Penjualan
1	Asam Asetil Salisilat (Asetosal) 80 mg	Tinggi
2	Betametason cr 0,1% (sebagai Valerat)	Tinggi
3	Bisoprolol 5 mg	Tinggi
4	Calcium Lactate 500 mg	Tinggi
5	Chloramphenicol / Bufacetin salep kulit	Tinggi
6	Codein 10 mg	Tinggi
7	Dipenhydramine Injeksi	Tinggi
8	Erythromycine 250 mg	Rendah
9	Gentamycine /Genoint salep mata 0,3%	Tinggi
10	Gentamycine tetes mata 0,3%	Tinggi
11	Diazepam Inj 5mg/ml	Tinggi
12	Prednison 5 mg	Sedang
13	Acetylcystein 200mg	Tinggi
14	Captopril 25 mg	Tinggi
15	Dulcolax Adult 10 mg Supp	Tinggi
16	Hidrokortison 2,5 % cream	Tinggi
17	Iliadin Anak 0.025% Nasal Drops 10 ml (Oxymetazoline HCL)	Tinggi
18	Natrium Diklofenac 50 mg	Tinggi
19	Permetrin Krim 5%	Tinggi
20	Ranitidin 150 mg	Tinggi
21	Simvastatin 10 mg	Tinggi
22	Zinc Sulfate Syrup 60 ml	Tinggi
23	Sulfadoxine 500mg+Primethamin 25mg	Rendah
24	Surgical Gloves Steril 7,5 (Ambumax)	Tinggi
25	Tetracycline 500 mg	Tinggi
26	Trihexyphenidyl	Tinggi
27	Alkohol 95% 1L/ Medika	Tinggi
28	Captopril 12,5 mg	Rendah

Dari tabel 5 di atas, dapat dilihat hasil pengklasteran 28 data jenis obat. Hasil tersebut menunjukkan bahwa proses pengklasteran dikelompokkan menjadi 3 jenis kategori di mana *cluster* pertama (C0) terdiri dari 24 item, *cluster* kedua (C1) yang terdiri dari 1 item serta *cluster* ketiga (C2) terdiri dari 3 item.

4. KESIMPULAN

Dari hasil penelitian yang telah dilakukan oleh peneliti, maka dapat disimpulkan bahwa metode clustering dengan menggunakan algoritma K-Means dapat digunakan untuk mengelompokkan data penjualan obat farmasi UPT Puskesmas Mandala berdasarkan jumlah penjualannya selama tahun 2023 (Januari-Desember). Klastering dikelompokkan ke dalam 3 kategori yaitu *Cluster* 0 untuk penjualan tinggi, *Cluster* 1 untuk penjualan menengah, dan *Cluster* 2 untuk penjualan rendah. Berdasarkan hasil penerapan dengan software RapidMiner didapatkan hasil pengelompokan dan pengujian yang terdiri dari tiga cluster yaitu cluster pertama (C0) yang terdiri 24 item, cluster kedua (C1) yang terdiri dari 3 item serta cluster ketiga (C2) terdiri dari 1 item dengan nilai Davies Bouldin Index yang didapatkan sebesar 0,276. Hasil penerapan metode *clustering* ini dapat digunakan untuk membantu pihak UPT Puskesmas dalam melihat tingkat penjualan obat mereka.

REFERENCES

- [1] S. Afrioza and I. Baidillah, "Hubungan Tingkat Kepuasan Pasien Terhadap Pelayanan Kesehatan di Puskmas Sepatan," *Journal of Nursing Practice and Education*, vol. 1, no. 2, pp. 169–180, Jun. 2021, doi: 10.34305/jnpe.v1i2.305.
- [2] M. P. Baybo, W. A. Lolo, and M. Jayanti, "Analisis Pengendalian Persediaan Obat di Puskesmas Teling Atas," *Pharmacy Medical Journal*, vol. 5, no. 1, pp. 7–13, 2022.
- [3] S. Rusman, "Analisis Sistem Penyusunan Perencanaan Obat di Dinas Kesehatan Kabupaten Padang Pariaman," *Jurnal Human Care*, vol. 5, no. 1, pp. 411–420, 2020.

- [4] A. L. Y. Pratasik, Fatimawali, and O. J. Sumampouw, "Analisis Perencanaan, Pengadaan, dan Pengendalian Obat di Instalasi Farmasi Unit Pelaksana Teknis Daerah Rumah Sakit Manembo Nembo Tipe C Bitung," *Jurnal Kesehatan Tambusai*, vol. 4, no. 4, pp. 5249–5266, 2023.
- [5] Mustika, Y. Ardilla, and A. Manuhutu, *Data Mining dan Aplikasinya*. Bandung: Widina Bhakti Persada Bandung, 2021.
- [6] A. O. P. Dewi, "Big Data di Perpustakaan dengan Memanfaatkan Data Mining," *ANUVA*, vol. 4, no. 2, pp. 223–230, 2020.
- [7] T. A. Aria, Yuliadi, M. Julkarnain, and F. Hamdani, "Penerapan Algoritma K-Means Clustering Untuk Data Obat," *KLIK: Kajian Ilmiah Informatika dan Komputer*, vol. 4, no. 1, pp. 649–657, 2023, doi: 10.30865/klik.v4i1.1117.
- [8] Syaifuddin, Ramlah, I. Hakim, Y. Berliana, and Nurhayati, "Pemetaan Produksi Tanaman Tomat di Indonesia Berdasarkan Provinsi Menggunakan Algoritma K-Means Clustering," *Journal of Computer System and Informatics (JoSYC)*, vol. 3, no. 4, pp. 222–228, Aug. 2022, doi: 10.47065/josyc.v3i4.2206.
- [9] A. Ayunda, P. Sari, and M. Muhajir, "Implementasi Metode Improved K-Means dengan Algoritma Dbscan untuk Pengelompokan Film," *Unisda Journal of Mathematics and Computer Science*, vol. 6, no. 1, pp. 1–8, 2020.
- [10] A. Fahmi, Risawandi, and Z. Husna, "Implementasi Metode K-Means Clustering dalam Menentukan Jenis Udag Terlaris pada UD. Tegar Rizqi," *TECHSI - Jurnal Teknik Informatika*, vol. 13, no. 1, pp. 92–106, Apr. 2021, doi: 10.29103/techsi.v13i1.3148.
- [11] Normah, S. Nurajizah, and A. Salbinda, "Penerapan Data Mining Metode K-Means Clustering untuk Analisa Penjualan pada Toko Fashion Hijab Banten," *Jurnal Teknik Komputer AMIK BSI*, vol. 7, no. 2, pp. 158–163, 2021, doi: 10.31294/jtk.v4i2.
- [12] N. Mirantika, A. Tsamratul'ain, and F. D. Agnia, "Penerapan Algoritma K-Means Clustering untuk Pengelompokan Penyebaran Covid-19 di Provinsi Jawa Barat," *Jurnal Nuansa Informatika*, vol. 15, no. 2, pp. 92–98, 2021, [Online]. Available: <https://journal.uniku.ac.id/index.php/ilkom>
- [13] M. K. Hidayat and R. Fitriana, "Penerapan Sistem Intelijensia Bisnis dan K-Means Clustering untuk Memantau Produksi Tanaman Obat," *Jurnal Teknologi Industri Pertanian*, vol. 32, no. 2, pp. 204–219, Aug. 2022, doi: 10.24961/j.tek.ind.pert.2022.32.2.204.
- [14] Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 5, no. 1, pp. 17–24, Apr. 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [15] D. Suryanto and I. R. Adevi, "Analisa Penjualan Toko Hijab Kiki Hn Dengan Penerapan Data Mining Metode K-Means Clustering Dan Market Basket Analysis," *Jurnal Mirai Management*, vol. 8, no. 3, pp. 167–176, 2023.
- [16] H. R. Putri, Maimunah, and E. R. Arumi, "Penentuan Prioritas Pendaftaran Tanah Sistematis Lengkap Melalui Clustering Jumlah Sertifikat Hak Atas Tanah Menggunakan K-Means," *Journal of Computer System and Informatics (JoSYC)*, vol. 4, no. 4, pp. 931–941, Aug. 2023, doi: 10.47065/josyc.v4i4.3928.
- [17] I. Sanela, A. Nazir, F. Syafria, E. Haerani, and L. Oktavia, "Penerapan Metode Clustering Dengan K-Means Untuk Memetakan Potensi Tanaman Padi di Sumatera," *Journal of Computer System and Informatics (JoSYC)*, vol. 5, no. 1, pp. 82–92, 2023, doi: 10.47065/josyc.v5i1.4523.
- [18] Z. Setiawan, M. Fajar, and A. M. Priyatno, *Buku Ajar Data Mining*. Jambi: PT. Sonpedia Publishing Indonesia, 2023.
- [19] R. Nofitri and N. Irawati, "Analisis Data Hasil Keuntungan Menggunakan Software RapidMiner," *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, vol. 5, no. 2, pp. 199–204, Jul. 2019, doi: 10.33330/jurteks.v5i2.365.
- [20] B. Kristanto, A. Turmudi, and M. Fatchan, "Analisis Penentuan Karyawan Tetap dengan Algoritma K-Means dan Davies Bouldin Index," *Bulletin of Information Technology (BIT)*, vol. 4, no. 1, pp. 112–120, 2023, doi: 10.47065/bit.v3i1.