

## **Peningkatan Ketersediaan Darah Sesuai Segmentasi Umur Menggunakan K-Means Clustering**

### ***Improvement of Blood Availability According to Age Segmentation Using K-Means Clustering***

**Titania Dwi Andini<sup>1\*</sup>, Lia Farokhah<sup>2</sup>**

Program Studi Teknik Informatika, Institut Teknologi dan Bisnis Asia, Malang

\*E-mail: [titania.andini@asia.ac.id](mailto:titania.andini@asia.ac.id)

#### **Abstrak**

Ketersediaan darah pada bank darah di kota Malang mengalami fluktuasi yang tidak merata disetiap bulan dan pada setiap tahunnya. Ini sangat mempengaruhi permintaan darah yang sangat tinggi disetiap harinya terlebih lagi disaat pandemi sekarang ini. Untuk memenuhi permintaan darah yang tinggi, diperlukan program yang bertujuan dapat menarik warga kota Malang terlebih lagi dari segala segmen usia untuk mendonorkan darahnya sehingga ketersediaan darah setiap bulannya dapat terpenuhi. Penelitian ini bertujuan melihat data stok darah dari BPS (Badan Pusat Statistik) Kota Malang tiap bulan dalam tiga tahun (2018-2020) yang akan di kluster / dikelompokkan menggunakan metode K-Means sesuai dengan segmentasi umur produktif ( 17, 18-24, 25-44, 45-59 tahun) sehingga akan menghasilkan informasi kapan akan diadakan program peningkatan penggalangan donor darah sesuai segmentasi umur dengan memaksimalkan bulan yang menjadi favorit segmentasi umur dan bulan yang sepi peminat donor. Dari penelitian ini dihasilkan kelompok kluster berdasarkan bulan dengan analisis pada jumlah kluster terbaik sehingga evaluasi menghasilkan pengetahuan bulan tertinggi yang memiliki banyak pendonor adalah bulan Oktober, sedangkan bulan terendah yang memiliki sedikit pendonor adalah bulan juni serta Desember. Pemerintah kota malang dan pihak terkait bisa membuat program peningkatan dan memaksimalan program donor darah pada bulan tersebut.

**Kata kunci:** Clustering, Ketersediaan Darah, K-Means.

#### **Abstract**

Availability of blood at blood banks in Malang city fluctuates unevenly every month and every year. This greatly affects the demand for blood which is very high every day, especially during the current pandemic. To meet the high demand for blood, a program is needed that aims to attract Malang city residents, especially the productive age segment to donate blood so that the availability of blood every month can be met. This study aims to look at blood stock data from BPS (Central Statistics Agency) Malang City every month for three years (2018-2020) which will be clustered / grouped using K-Means method according to the productive age segmentation (17, 18-24, 25-44, 45-59 years) so that it will produce information when there will be an increase in blood donor raising program according to age segmentation by maximizing the month that is the favorite month of the age segmentation and the month that is quiet for donors. This research resulted in cluster groups based on months with analysis on the best number of clusters so that the evaluation resulted in knowledge that the highest month with many donors was October, while the lowest month with few donors was June and December. The government of Malang and related parties can make a program to increase and maximize the blood donor program in that month.

**Keywords:** Clustering, Blood Availability, K-Means.

Naskah diterima 18 Agust. 2022; direvisi 10 Okt. 2022; dipublikasikan 13 Okt. 2022.  
JAMIKA is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



## I. PENDAHULUAN

Darah merupakan satu elemen penting dalam kehidupan. Darah merupakan cairan yang berfungsi menyuplai tubuh dengan nutrisi, mengangkut bahan-bahan kimia hasil metabolisme, mengirimkan zat-zat oksigen yang dibutuhkan oleh jaringan tubuh, dan juga sebagai pertahanan tubuh terhadap virus atau bakteri. Luka yang mengakibatkan pendarahan hebat dapat menyebabkan kehilangan darah dalam tubuh seseorang secara signifikan. Kehilangan darah dalam jumlah besar tidak hanya dikarenakan luka, namun bisa disebabkan karena trauma, syok atau organ pembentuk sel darah merah tidak berfungsi. Untuk meminimalisasi korban dikarenakan kekurangan darah, maka dibutuhkan transfusi darah dan ketersediaan darah yang memadai.

Unit Transfusi Darah (UTD) PMI (Palang Merah Indonesia) kota Malang merupakan salah satu instansi penyedia darah terletak di jalan Buring no.10 Malang yang bertanggungjawab menyediakan stok darah di kota

Malang dan sekitarnya. Instansi tersebut memiliki berbagai macam program supaya stok darah tetap tersedia dan terjaga. Salah satunya mendukung program PMI pusat untuk menarik pendonor untuk berpartisipasi dalam mendonorkan darahnya melalui program “ayudonor” dalam skala nasional. Stok darah pada UTD PMI kota Malang mengalami fluktuatif terlebih di bulan Ramadan dan bulan program vaksinasi COVID 19 yang dilakukan secara serentak di kota Malang[1]. Hal ini disebabkan karena jika pendonor melakukan vaksinasi pertama, tidak diperbolehkan sama sekali untuk mendonorkan darahnya, pada vaksinasi kedua membutuhkan jarak waktu minimal 2 minggu untuk bisa mendonorkan darahnya. Sehingga semakin sedikit pendonor yang bisa melakukan donor darah.

Menurut Badan Pusat Statistik (BPS), total pendonor darah pada usia 17 tahun, 18 hingga 24 tahun mengalami penurunan cukup signifikan pada tahun 2018 hingga 2020 sekitar 500 hingga 2.000 orang tingkat penurunannya. Untuk kelompok usia antara 25 hingga 44 tahun penurunannya sekitar 1.000 orang dalam rentang tahun yang sama [2]. Namun yang menarik, untuk segmen umur 45 hingga 59 tahun, mengalami kenaikan 1000 orang tiap tahun untuk rentang tahun yang sama pula yaitu antara 2018 hingga 2020. Untuk penyebarannya tiap bulan masing-masing segmentasi umur juga mengalami fluktuasi yang beragam. Program ketersediaan darah cukup penting karena terkait Kesehatan untuk berbagai macam kebutuhan. Untuk menarik minat pendonor sesuai klasifikasi badan pusat statistik, pemanfaatan data yang sudah tercatat merupakan salah satu bagian penting untuk menggali sebuah informasi yang berharga. Informasi ini bisa dipakai untuk mengatur kebijakan terkait ketersediaan darah yang ada di PMI kota Malang sehingga stok darah bisa terus ditingkatkan.

Dari paparan masalah di atas, dibutuhkan sebuah metode pengelompokan yang dapat memberikan informasi sehingga dapat menghasilkan informasi bulan penting dalam setahun dalam program donor darah sesuai dengan segmentasi umur pendonor. K-Means merupakan salah satu metode pengelompokan / klasifikasi data nonhierarki (sekatan) yang dapat mempartisi data ke dalam dua kelompok atau lebih dan diharapkan bisa menjawab permasalahan di atas [3]. Metode tersebut akan menyekat/memartisi data ke dalam suatu kelompok sesuai karakteristiknya[4]. Tujuan dari pengelompokan, yaitu untuk meminimalisir fungsi objektif yang diset dalam proses pengelompokan pada umumnya. Metode ini akan meminimalkan variasi di dalam suatu kelompok dan memaksimalkan variasi antar kelompok[5].

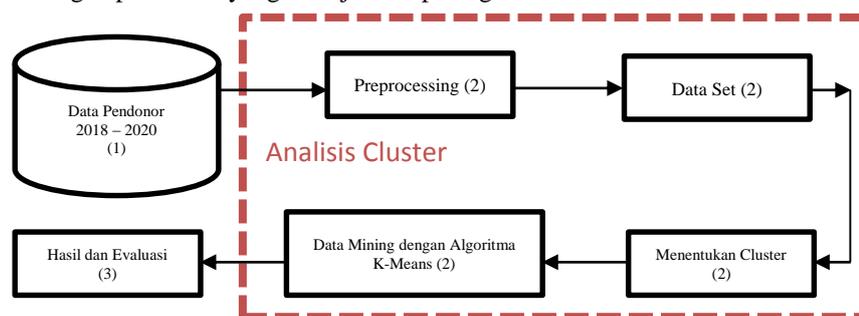
Klasifikasi tentang pendonor darah cukup banyak diangkat oleh penelitian sebelumnya seperti kelayakan calon pendonor darah[6],[7],[8] menggunakan klasifikasi pada data mining lainnya, namun belum ada yang menggunakan metode K-Means. Ketiga penelitian sebelumnya mengangkat kelayakan pendonor darah dalam mengikuti program donor darah (sebelum program). Namun masih sedikit penelitian yang mengolah data pendonor setelah melaksanakan program. Penelitian ini dilakukan bertujuan mengetahui klusterisasi data pendonor sesuai dengan segmentasi umur 17 tahun, 18-24 tahun, 25-44 tahun, 45-59 tahun. Dataset yang ada dari tahun 2018-2020. Keluaran yang dihasilkan adalah kelompok donor darah per bulan sesuai segmentasi umur dan rekomendasi untuk peningkatan program ketersediaan darah.

## II. METODE PENELITIAN

Penelitian ini akan memiliki empat pengelompokan pendonor darah berdasarkan range umur yang telah ditentukan dan bulan sebagai variabel penelitian. Dataset pendonor darah yang diambil selama tiga tahun. Aplikasi pendukung untuk melakukan pengolahan data menggunakan metode K-Means adalah RapidMiner Studio[9],[10] sebagai salah satu aplikasi yang akurat dalam klarifikasi data[11] dan sering digunakan.

### Tahapan Penelitian

Dalam penelitian ini untuk lebih terarah dan terstruktur dalam melakukan olah data, maka tahapan-tahapan yang dilakukan, yaitu: pencarian dataset di BPS, analisis kluster, evaluasi, dan validasi kemudian dibuatlah kerangka penelitian yang ditunjukkan pada gambar 1 di bawah ini :



Gambar 1. Diagram Kerangka Penelitian

Adapun langkah-langkah detail dari gambar 1 di atas adalah :

1. Tahapan pertama adalah pencarian dataset. Dataset di ambil dari situs BPS kota Malang Dengan tahun 2018 sampai tahun 2020. Adapun variabel pada dataset ini adalah tahun dan id data merupakan bulan selama satu tahun. Data ini merupakan data resmi sehingga bisa dipertanggungjawabkan.
2. Pada tahap kedua, dilakukan pengecekan data dan pembersihan data. Pada tahap ini dataset melalui proses *preprocessing* dengan tujuan supaya bersih dan memenuhi bentuk yang diperlukan Ketika mengolah data klustering (sesuai dengan variabel yang dibutuhkan) dan siap dimodelkan sehingga menjadi Data Set. Pada tahap permodelan, k-means *clustering* dipilih sebagai metode pengelompokan. Jumlah kluster yang diujicobakan ada tiga, yaitu K=2, K=3, K=5. Percobaan jumlah kluster dilakukan untuk mendapatkan hasil evaluasi kelompok terbaik.
3. Pada tahap ketiga, dilakukan evaluasi dengan memilih nilai K terbaik melalui metode *davies bouldin index* untuk mengukur apakah jumlah kluster yang terpilih itu terbaik. Setelah tahap clustering kelompok sesuai segmentasi umur selesai maka dilakukan analisis lanjutan terhadap kelompok kluster yang terbentuk. Analisis ini dilanjutkan untuk menentukan bulan apa yang sepi peminat dan ramai pendonor. Analisis lanjutan dilakukan dengan cari mencari rata-rata pertahun antar kluster.

### **Data mining**

*Data mining* adalah pembelajaran dalam mengumpulkan, membersihkan, memproses, menganalisis, dan memperoleh manfaat dari data[12]. Variasi yang luas ada dalam hal domain masalah, aplikasi, formulasi, dan representasi data yang ditemui dalam aplikasi nyata. Karena itu, data mining merupakan istilah umum yang digunakan untuk menggambarkan aspek-aspek yang berbeda dari pengolahan data[13]. *Data mining* berisi pencarian tren atau menginginkan pola dalam database yang besar untuk membantu pengambilan keputusan diwaktu yang akan datang. Pola-pola ini dikenali oleh perangkat tertentu dimana dapat memberikan suatu analisa data yang berguna dan berwawasan lalu kemudian dapat dipelajari dengan lebih teliti, ini bisa saja menggunakan perangkat pendukung keputusan yang lain[14]. *Data mining* mulai menarik minat berbagai stakeholder baik pemerintah atau swasta dalam menganalisis kondisi lembaga mereka atau untuk memajukan bisnis mereka. Operasi Data mining menurut sifatnya dibedakan menjadi dua, yaitu bersifat (1) prediksi (*prediction driven*) untuk menjawab pertanyaan apa dan sesuatu yang bersifat remang-remang atau transparan. Operasi prediksi digunakan untuk validasi hipotesis, *querying* dan pelaporan (misal: *spreadsheet* dan *pivot table*), analisis multidimensi (*dimensional summary*); OLAP (*Online Analytic Processing*) serta analisis statistik. (2) Penemuan (*discovery driven*) bersifat transparan dan menjawab pertanyaan “mengapa?”[15].

*Clustering* merupakan salah satu metode data mining yang bersifat tanpa arahan (*unsupervised*) yang memiliki arti bahwa karakteristik tiap kluster tidak ditentukan sebelumnya melainkan berdasarkan kemiripan atribut-atribut dari suatu kelompok atau kluster. Kluster membagi data menjadi kelompok- kelompok atau kluster berdasarkan suatu kemiripan atribut-atribut diantara sekumpulan data, atribut-atribut yang mirip tersebut dipresentasikan sebagai titik-titik dalam ruang multidimensi. Dalam data mining ada dua jenis metode *Clustering* yang digunakan dalam pengelompokan data, yaitu *Hierarchical Clustering* dan *non-hierarchical Clustering*.

### **Algoritma K-Means**

Metode K-means merupakan algoritma klasterisasi non-hirarki (*non-hierarchical Clustering*) yang paling tua dan banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan implementasinya[16],[17]. Ide dasar algoritma K-Means sangatlah sederhana, yaitu meminimalkan *Sym of Squared Error* (SSE) antara objek-objek data dengan beberapa langkah. Langkah tersebut antara lain:

1. Menentukan k sebagai jumlah kluster yang dibentuk.
2. Bangkitkan k Centroid (titik pusat kluster) awal secara random:

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

dimana :

v : centroid pada kluster

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

n : banyaknya objek/jumlah objek yang menjadi anggota kluster

- Menghitung jarak setiap objek ke masing-masing centroid dari masing-masing kluster

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

dimana :

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

- Mengalokasikan masing-masing objek ke dalam centroid terdekat.
- Melakukan iterasi, kemudian menentukan posisi centroid baru dengan persamaan (1).
- Mengulangi langkah (3) jika posisi centroid baru tidak sama.

Untuk pemilihan metode K-Means, dipilih evaluasi kluster menggunakan *Davies Bouldin Index* (DBI) karena merupakan salah satu metode evaluasi internal.

#### **Evaluasi Kluster Menggunakan Davies Bouldin Index**

Salah satu metode evaluasi internal untuk mengukur evaluasi kluster pada suatu metode pengelompokan yang berdasarkan pada nilai kohesi dan separasi adalah evaluasi kluster *Davies Bouldin Index* atau yang umum disingkat dengan DBI[18]. Untuk mengetahui persamaan yang digunakan dalam matrik kohesi dalam sebuah kluster ke-i adalah *Sum of square within cluster* (SSW) yang dirumuskan sebagai berikut:

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{m_i} d(x_j, c_i) \quad (3)$$

Dari persamaan di atas,  $m_i$  merupakan jumlah data dalam kluster ke-i,  $c_i$  adalah centroid kluster ke-i, dan  $d()$  merupakan jarak setiap data ke centroid yang dihitung menggunakan jarak Euclidean. Sedangkan *Sum of square between cluster* (SSB) merupakan persamaan yang digunakan untuk mengetahui separasi antar kluster dengan persamaan[19]:

$$SSB_{i,j} = d(c_i, c_j) \quad (4)$$

Jika nilai kohesi dan separasi diperoleh, maka dilakukan rasio ( $R_{ij}$ ) untuk mengetahui nilai perbandingan antara kluster ke-i dan kluster ke-j. Kluster yang baik adalah kluster yang memiliki kohesi sekecil mungkin dan separasi yang sebesar mungkin. Persamaan untuk menghitung nilai rasio adalah:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSB_{i,j}} \quad (5)$$

Nilai rasio yang diperoleh digunakan untuk mencari nilai DBI dari persamaan berikut:

$$DBI = \frac{1}{k} \sum_i^k \max_{i \neq j} (R_{i,j}) \quad (6)$$

Dari persamaan di atas,  $k$  merupakan jumlah kluster yang digunakan. Semakin kecil nilai DBI (non-negatif  $\geq 0$ ), maka semakin baik kluster yang diperoleh dari pengelompokan K-Means yang digunakan.

### III. HASIL DAN PEMBAHASAN

Pada penelitian ini dilakukan klustering terhadap data donor darah PMI kota Malang selama 3 tahun. *Input* dari *dataset* adalah data donor darah berdasarkan beberapa segmentasi umur, yaitu umur 17 tahun, 18-24 tahun dan 25-44 tahun terhadap bulan dalam satu tahun. Adapun luaran yang diharapkan adalah hasil analisis peminat donor terhadap bulan dalam satu tahun sesuai dengan segmentasi umur. Hal ini bertujuan untuk meningkatkan jumlah pendonor darah bulan sepi peminat atau mendeteksi bulan yang ramai peminat berdasarkan hasil kluster yang terbentuk. Hal ini digunakan untuk memberikan masukan pada pihak terkait untuk membuat program untuk peningkatan stok darah. Tahapan penelitian dibagikan menjadi tiga, yaitu percobaan beberapa nilai  $k$  dari kluster, evaluasi kelompok kluster terbaik menggunakan *davies-bouldin index* dan validasi serta analisis hasil kluster melalui titik *centroid* yang terbentuk. Adapun kontribusi utama dari

penelitian ini adalah cara menganalisis hasil kelompok cluster yang terbentuk melalui titik *centroid*. Hal ini jarang dilakukan oleh penelitian sebelumnya dimana hanya sampai terbentuknya kelompok cluster atau hasil evaluasi menggunakan *davies-bouldin index*[18],[19],[20] (DBI).

### Segmentasi Umur 17 Tahun

Dari data yang diperoleh, untuk segmentasi umur 17 tahun didapat hasil evaluasi DBI sebagai berikut:

TABEL 1.  
HASIL DBI UNTUK SEGMENTASI UMUR 17 TAHUN

K=2	K=3	K=5
0,255	0,461	0,487

Dari hasil DBI di atas maka pembentukan k=2 adalah yang terbaik. Hal ini dikarenakan semakin kecil nilai DBI maka tingkat SSW- Keterikatan antar anggota dalam cluster/Homogenitas (kecil) dan SSB- Jarak antar cluster/ heterogenitas(besar). Untuk lebih jelasnya akan dipaparkan pada tabel hasil pengelompokan berdasarkan umur 17 tahun setiap bulan menggunakan aplikasi RapidMiner pada gambar 2 di bawah ini:

Row No.	Bulan	cluster	2018	2019	2020
1	Januari	cluster_0	75	15	34
2	Februari	cluster_1	61	14	123
3	Maret	cluster_0	94	4	17
4	April	cluster_0	26	13	22
5	Mei	cluster_0	39	7	19
6	Juni	cluster_0	19	2	10
7	Juli	cluster_0	54	21	10
8	Agustus	cluster_0	51	17	9
9	September	cluster_0	29	13	11
10	Oktober	cluster_0	120	13	8
11	November	cluster_0	50	40	7
12	Desember	cluster_0	35	13	8

Gambar 2. Hasil pengelompokan menggunakan rapidminer untuk segmentasi usia 17 tahun

Hasil anggota dari cluster di atas adalah sebagai berikut : untuk cluster-0 terdapat bulan Januari, Maret, April, Mei, Juni, Juli, Agustus, September, Oktober, November, dan Desember yang menjadi anggotanya, sedangkan untuk cluster-1 hanya bulan Februari yang menjadi anggotanya. Untuk mengetahui titik *centroid* masing-masing cluster dapat dilihat pada tabel 2 berikut:

TABEL 2  
TITIK CENTROID SEGMENTASI UMUR 17 TAHUN

Attribute	Kluster-0	Kluster-1
2018	53.818	61
2019	14.364	14
2020	14.091	123

Dari Analisa 2 kluster di atas, dilakukan analisis lebih lanjut dengan titik *centroid* yang terbentuk antar kelompok kluster maka kluster-1, yaitu bulan Februari diklasifikasikan sebagai bulan yang banyak pendonor dibandingkan kluster-0. Hal ini menjadi rekomendasi untuk PMI kota Malang untuk meningkatkan program yang semakin banyak untuk para remaja umur 17 tahun terkait donor darah pada bulan Februari yang ada di kluster-1 dan bulan yang dipandang sepi pendonor, yaitu kluster-0.

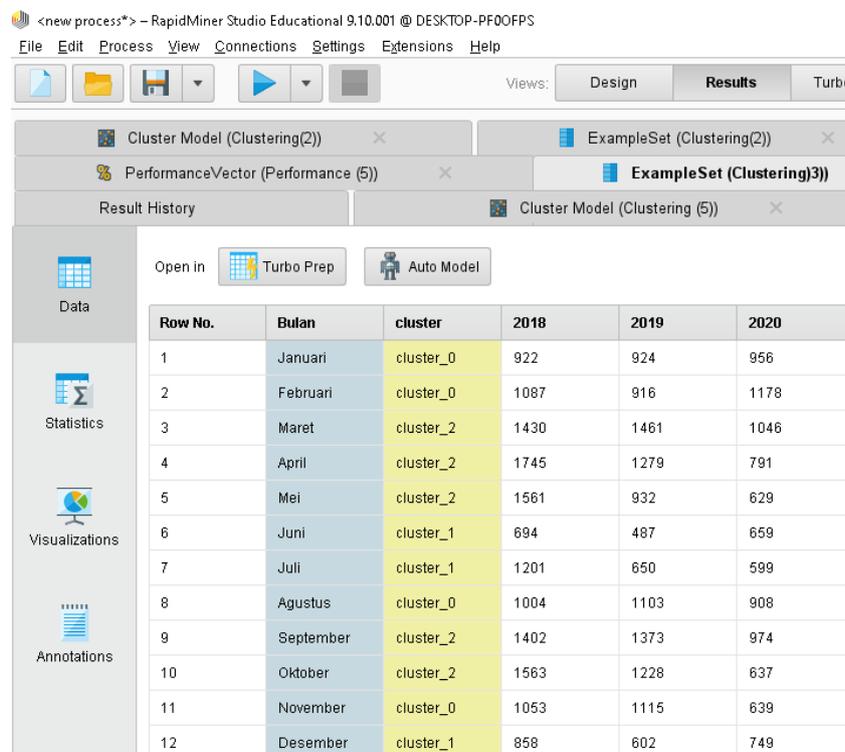
### Segmentasi Umur 18-24 Tahun

Untuk segmentasi umur 18 hingga 24 tahun, untuk evaluasi DBI diperoleh sebagai berikut:

TABEL 3  
HASIL DBI UNTUK SEGMENTASI UMUR 18-24 TAHUN

K=2	K=3	K=5
0,839	0,811	0,848

Dari hasil DBI di atas maka pembentukan k=3 adalah yang terbaik karena memiliki nilai DBI terkecil dibandingkan k lainnya. Selanjutnya akan dikelompokkan kembali berdasarkan umur 18 hingga 24 tahun dan hasil anggota dari kluster di atas menggunakan aplikasi RapidMiner sesuai pada gambar 3 di bawah ini:



Row No.	Bulan	cluster	2018	2019	2020
1	Januari	cluster_0	922	924	956
2	Februari	cluster_0	1087	916	1178
3	Maret	cluster_2	1430	1461	1046
4	April	cluster_2	1745	1279	791
5	Mei	cluster_2	1561	932	629
6	Juni	cluster_1	694	487	659
7	Juli	cluster_1	1201	650	599
8	Agustus	cluster_0	1004	1103	908
9	September	cluster_2	1402	1373	974
10	Oktober	cluster_2	1563	1228	637
11	November	cluster_0	1053	1115	639
12	Desember	cluster_1	858	602	749

Gambar 3. Hasil pengelompokan menggunakan rapidminer untuk segmentasi usia 18-24 tahun

Dari gambar 3 di atas terdapat tiga kluster dimana kluster 0 ada pada bulan Januari, Februari, Agustus, dan November. Untuk kluster 1 ada pada bulan Juni, Juli, dan Desember. Untuk kluster 2 tersebar pada bulan

Maret, April, Mei, September dan Oktober. Titik *centroid* masing-masing kluster dapat dilihat pada tabel 4 di bawah:

TABEL 4  
TITIK CENTROID SEGMENT UMUR 18-24 TAHUN

Attribute	Kluster-0	Kluster-1	Kluster-2
2018	1016.500	917.667	1540.200
2019	1014.500	579.667	1254.600
2020	920.250	669	815.400

Dari tabel 4 *centroid* di atas, maka untuk kluster rendah adalah kluster 1, kluster sedang ada di kluster 0 dan kluster tinggi berada pada kluster 2. Untuk kluster tinggi adalah kluster 2 di bulan Maret, April, Mei, September dan Oktober, sedangkan kluster rendah di bulan Juni, Juli dan Desember, sehingga PMI dapat membuat program di kedua kluster di atas.

#### Segmentasi Umur 25-44 Tahun

Hasil evaluasi DBI yang dilakukan untuk segmentasi umur 25-44 tahun, dapat dilihat dari tabel 5 di bawah ini:

TABEL 5  
HASIL DBI UNTUK SEGMENTASI UMUR 25-44 TAHUN

K=2	K=3	K=5
0,885	0,733	0,626

Dari hasil DBI pada tabel 5, maka untuk segmentasi umur 25-44 tahun pembentukan k=5 adalah yang terbaik karena memiliki nilai paling rendah. Kemudian dilakukan pengelompokan berdasarkan umur 25 hingga 44 tahun menggunakan rapidminer seperti pada gambar 4 di bawah ini:

Row No.	Bulan	cluster	2018	2019	2020
1	Januari	cluster_2	1775	1734	1872
2	Februari	cluster_3	1678	1638	1540
3	Maret	cluster_0	1894	1675	1418
4	April	cluster_1	1694	1722	2470
5	Mei	cluster_2	1887	1779	2000
6	Juni	cluster_3	1747	1443	1688
7	Juli	cluster_4	2179	1959	1403
8	Agustus	cluster_3	1581	1701	1764
9	September	cluster_3	1526	1637	1714
10	Oktober	cluster_4	2035	1985	1726
11	November	cluster_0	1610	1599	1326
12	Desember	cluster_3	1520	1635	1721

Gambar 4. Hasil pengelompokan menggunakan rapidminer untuk segmentasi usia 25-44 tahun

Pada gambar 4 terdapat 5 kluster, yaitu kluster 0 dikisaran bulan Maret dan November, untuk kluster 1 pada bulan April, sedangkan untuk kluster 2 dikisaran bulan Januari dan Mei. Selanjutnya untuk kluster 3 pada bulan Februari, Juni, Agustus, September dan Desember. Untuk kluster 4 pada kisaran bulan Juli dan bulan Oktober. Masing-masing titik *centroid* kluster dapat diketahui pada tabel 6 berikut:

TABEL 6  
TITIK CENTROID SEGMENT UMUR 25-44 TAHUN

Attribute	Kluster-0	Kluster-1	Kluster-2	Kluster-3	Kluster-4
2018	1752	1694	1831	1610.400	2107
2019	1637	1722	1756.500	1610.800	1972
2020	1372	2470	1936	1685.400	1564.500

Dari Analisa 5 kluster dengan memperhatikan tabel 6 di atas, dilakukan analisis lebih lanjut dengan melihat titik tertinggi dan terendah dengan hasil kluster 4 untuk nilai tertinggi dan kluster 3 untuk nilai terendah. Hal ini menjadi rekomendasi untuk PMI kota Malang untuk meningkatkan program untuk segmentasi umur 25-44 tahun di bulan Juli dan Oktober untuk pendonor tertinggi yang terletak pada kluster 4 dan bulan yang sepi pendonor di bulan Februari, Juni, Agustus, September dan Desember yang masuk dalam kluster 3.

#### Segmentasi Umur 45-59 Tahun

Seperti pada segmentasi sebelumnya, dilakukan evaluasi DBI untuk segmentasi umur 45-59 tahun dan dapat dilihat dari tabel 7 berikut:

TABEL 7  
HASIL DBI UNTUK SEGMENTASI UMUR 45-59 TAHUN

K=2	K=3	K=5
0,951	0,852	0,636

Dari hasil SBI di atas maka untuk segmentasi umur 45-59 pembentukan k=5 adalah yang terbaik. Hasil anggota dari kluster di atas diolah menggunakan aplikasi rapidminer dan hasilnya seperti pada gambar 5 di bawah ini:

Row No.	Bulan	cluster	2018	2019	2020
1	Januari	cluster_0	923	1058	1159
2	Februari	cluster_4	807	900	959
3	Maret	cluster_4	925	947	851
4	April	cluster_2	967	1060	1541
5	Mei	cluster_3	931	1173	1266
6	Juni	cluster_0	1003	881	1118
7	Juli	cluster_1	1121	1008	1009
8	Agustus	cluster_0	826	990	1203
9	September	cluster_0	918	917	1086
10	Oktober	cluster_3	1072	1113	1178
11	November	cluster_4	832	997	876
12	Desember	cluster_0	970	1009	1072

Gambar 5 Hasil pengelompokan menggunakan rapid miner untuk segmentasi usia 45-59 tahun

Dari Analisa gambar 5 di atas dapat memperoleh hasil untuk kluster 0 pada bulan Januari, Juni, Agustus, September dan Desember. Untuk kluster 1 tersebar pada bulan Juli, sedangkan untuk kluster 2 pada bulan April, untuk kluster 3 pada bulan Mei dan Oktober. Untuk kluster 4 tersebar pada bulan, Februari, Maret, dan bulan November. Kemudian dihitung titik *centroid* kluster selanjutnya, dapat dilihat pada tabel 8 di bawah ini:

TABEL 8  
 TITIK CENTROID SEGMENT UMUR 45-59 TAHUN

Attribute	Kluster-0	Kluster-1	Kluster-2	Kluster-3	Kluster-4
2018	928	1121	967	1001.500	854.667
2019	971	1008	1060	1143	948
2020	1127.600	1009	1541	1222	895.333

Selama tiga tahun data yang diambil, untuk segmentasi umur 45 hingga 59 tahun didapat hasil kluster tertinggi pada kluster 3 dan terendah pada kluster 4. Dapat direkomendasikan untuk PMI membuat program untuk pendonor darah paling tinggi di bulan Mei dan Oktober yang terletak pada kluster 3, dan untuk bulan yang bisa dikatakan sepi tahu rendah pendonor adalah bulan Februari, Maret dan bulan November yang termasuk dalam kluster 4. Berikut ringkasan hasil donor darah kota malang berdasarkan segmentasi umur setelah dikelompokkan menggunakan menggunakan metode K-Means:

TABEL 9  
 HASIL ANALISIS (DALAM HITUNGAN BULAN) KLUSTER TERTINGGI DAN TERENDAH SETIAP SEGMENTASI UMUR

Segmentasi Umur	Hasil Analisis (bulan)	
	Rendah	Tinggi
17	Januari, Maret, April, Mei, Juni, Juli, Agustus, September, Oktober, November, Desember	Februari
18-24	Juni, Juli, Desember	Maret, April, Mei, September, Oktober
25-44	Februari, Juni, Agustus, September, Desember	Juli, Oktober
45-59	Februari, Maret, November	Mei, Oktober

Dari tabel 9 di atas dapat digunakan untuk pihak PMI dan pihak terkait untuk membuat program sesuai dengan bulan-bulan tertinggi dan terendah pendonor mendonorkan darahnya sesuai segmentasi umur. Bulan oktober merupakan bulan dengan irisan segmentasi umur paling banyak, yaitu umur 18-24, umur 23-44 dan umur 45-59. Sedangkan untuk segmentasi remaja yaitu 17 tahun bulan februari sebagai bulan terbaik dan terbanyak untuk donor darah. Bulan dengan pendonor tertinggi ini bisa dimanfaatkan untuk memaksimalkan program donor. Disisi lain, untuk bulan terendah dari semua segmentasi umur cenderung berbeda beda. Namun irisan paling banyak ada pada bulan Juni dan Desember.

#### IV. KESIMPULAN

Pada penelitian ini, dataset donor dari PMI Kota malang dari tahun 2018-2020 diolah menggunakan algoritma *k-means* klustering. Algoritma ini menghasilkan luaran kelompok kluster. Pada penelitian ini, dilakukan eksperimen dengan nilai kluster dan dievaluasi menggunakan DBI. Hasil dari kluster yang terbentuk dengan DBI terbaik, dianalisis kembali untuk mendapatkan karakteristik dari masing-masing kelompok kluster melalui titik centroid yang terbentuk untuk mencari tentang pola minat donor darah. Hal ini jarang dilakukan oleh penelitian yang sudah ada. Rata-rata hanya sampai terbentuknya kelompok kluster. Hasil penelitian ini memiliki pola bulan dengan pendonor yang tinggi adalah bulan Oktober, Juni, dan bulan Desember merupakan bulan yang terendah orang mendonorkan darah. Hal ini memungkinkan untuk meningkatkan program pada bulan dengan peminat sedikit dan memaksimalkan pada bulan yang ramai pendonor sehingga ketersediaan darah di kota malang selalu aman dan terpenuhi. Pada penelitian selanjutnya, bisa diperluas dengan

perbandingan berbagai metode klastering seperti DBSCAN, Agglomerative, K-medoids untuk mempertajam terbentuknya kluster terbaik.

DAFTAR PUSTAKA

- [1] H. Oktavia, "Stok Darah di PMI Kota Malang Rawan Menipis," *rri.co.id*, 2021.
- [2] B. Malang, "Jumlah Donor Darah Menurut Kelompok Umur dan Bulanan di Kota Malang (Jiwa), 2018-2020," *Badan Pusat Statistik*, 2022.
- [3] G. Gustientiedina, M. H. Adiya, and Y. Desnelita, "Penerapan Algoritma K-Means Untuk Clustering Data Obat-Obatan," *J. Nas. Teknol. dan Sist. Inf.*, vol. 5, no. 1, pp. 17–24, 2019, doi: 10.25077/teknosi.v5i1.2019.17-24.
- [4] S. Aulia, "Klasterisasi Pola Penjualan Pestisida Menggunakan Metode K-Means Clustering (Studi Kasus Di Toko Juanda Tani Kecamatan Hutabayu Raja)," *Djtechno J. Teknol. Inf.*, vol. 1, no. 1, pp. 1–5, 2021, doi: 10.46576/djtechno.v1i1.964.
- [5] L. R. Ananda, "Clustering Untuk Menentukan Calon Mahasiswa Berprestasi," *Jiti*, vol. 1, no. 2, pp. 16–19, 2018.
- [6] M. R. Firdaus, A. Latif, and W. Gata, "Klasifikasi Kelayakan Calon Pendorong Darah Menggunakan Neura L Network," *Sistemasi*, vol. 9, no. 2, p. 362, 2020, doi: 10.32520/stmsi.v9i2.840.
- [7] M. Lestandy, L. Syafa'ah, and A. Faruq, "Classification of potential blood donors using machine learning algorithms approach," *J. Teknol. dan Sist. Komput.*, vol. 8, no. 3, pp. 217–221, 2020, doi: 10.14710/jtsiskom.2020.13619.
- [8] E. B. Nugroho, M. T. Furqon, and N. Hidayat, "Klasifikasi Pendorong Darah Menggunakan Metode Support Vector Machine pada Dataset RFMTC," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. 10, pp. 3860–3864, 2018, [Online]. Available: <https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/2751>
- [9] R. W. Sari, A. Wanto, and A. P. Windarto, "IMPLEMENTASI RAPIDMINER DENGAN METODE K-MEANS (STUDY KASUS: IMUNISASI CAMPAK PADA BALITA BERDASARKAN PROVINSI)," vol. 2, pp. 224–230, 2018.
- [10] J. Nasional, S. Informasi, V. Riandaru, H. Lazuardi, A. Adhi, and C. Lauw, "Penerapan Aplikasi RapidMiner Untuk Prediksi Nilai Tukar Rupiah Terhadap US Dollar Dengan Metode Regresi Linier," vol. 01, pp. 8–17, 2021.
- [11] P. Seminar and N. Matematika, "Akurasi Algoritma Klasifikasi pada Software Rapidminer dan Weka," vol. 4, pp. 493–499, 2021.
- [12] U. Kokate, A. Deshpande, P. Mahalle, and P. Patil, "Data stream clustering techniques, applications, and models: Comparative analysis and discussion," *Big Data Cogn. Comput.*, vol. 2, no. 4, pp. 1–30, 2018, doi: 10.3390/bdcc2040032.
- [13] E. D. Sikumbang, "Penerapan Data Mining Penjualan Sepatu Menggunakan Metode Algoritma Apriori," vol. 4, no. 1, pp. 156–161, 2018.
- [14] T. Jurnal, S. Dan, R. Rayuwati, and K. Koko, "Implementasi data mining untuk menentukan strategi penjualan buku bekas dengan pola pembelian konsumen menggunakan metode Apriori (studi kasus : Kota Medan)," vol. 16, no. 1, pp. 69–82, 2020.
- [15] A. Anas and R. H. Delima, "Penerapan Data Mining Untuk Menganalisis Data Mahasiswa Baru STIE-GK Muara Bulian," vol. 15, no. 2, pp. 119–128, 2021.
- [16] P. Subekti, T. D. Andini, and M. Islamiyah, "Sistem Penentuan Konsentrasi Jurusan Bagi Mahasiswa Informatika Menggunakan Metode K-Means Di Institut Asia Malang Determination System for Department Concentration for Informatics Students Using the K-Means Method at the Institute of Asia Malang," vol. 12, no. April, pp. 25–39, 2022, doi: 10.34010/jamika.v12i1.6452.
- [17] B. Harahap, "Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung)," pp. 394–403, 2019, [Online]. Available: <https://ptki.ac.id/jurnal/index.php/readystar/article/view/82>
- [18] D. . Jumadi, Bernard, "PENINGKATAN HASIL EVALUASI CLUSTERING DAVIES-BOULDIN INDEX DENGAN PENENTUAN TITIK PUSAT CLUSTER AWAL ALGORITMA K-MEANS," UNIVERSITAS SUMATERA UTARA, 2018.
- [19] Y. Arie Wijaya, D. Acmad Kurniady, E. Setyanto, W. Sanur Tarihoran, D. Rusmana, and R. Rahim, "Davies Bouldin Index Algorithm for Optimizing Clustering Case Studies Mapping School Facilities," *TEM J.*, vol. 10, no. 3, pp. 1099–1103, 2021, [Online]. Available: [https://www.temjournal.com/content/103/TEMJournalAugust2021\\_1099\\_1103.pdf](https://www.temjournal.com/content/103/TEMJournalAugust2021_1099_1103.pdf)

- 
- [20] A. Badruttamam and D. A. I. Maruddani, "PENERAPAN ANALISIS KLASTER K-MODES DENGAN VALIDASI DAVIES BOULDIN INDEX DALAM MENENTUKAN KARAKTERISTIK KANAL YOUTUBE DI INDONESIA (Studi Kasus: 250 Kanal YouTube Indonesia Teratas Menurut Socialblade )," vol. 9, pp. 263–272, 2020.