

CLUSTERING DATA BLOOD TRANSFUSION METHOD FUZZY C-MEANS

Nikolaus Rio Saputra

Program Studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara
Jl. Letjen S.Parman No.1, Jakarta Barat, DKI Jakarta, Indonesia 11410
Email: nikoblu12@gmail.com

ABSTRAK

Penelitian ini menggunakan metode clustering untuk menggambarkan perilaku donor darah. Tujuan dari clustering ini adalah memberikan kemampuan untuk menentukan segmentasi donor darah untuk meningkatkan akurasi prediksi manajer. Kami menggunakan algoritma Fuzzy k- means untuk jumlah cluster 2, 3, 4, dan 5 kemudian kami menggunakan metode silhouette untuk mengetahui kualitas suatu cluster. Hasil silhouette menunjukkan bahwa cluster optimal adalah 3 yang dikonfirmasi melalui tahap pengujian. Hasil clustering donor darah menggambarkan bahwa laki-laki memiliki kecenderungan untuk mendonorkan darahnya.

Kata Kunci: *Clustering, Fuzzy K-Means, Silhouette, donor darah*

ABSTRACT

This study uses the clustering method to describe the behavior of blood donors. The purpose of this clustering is to provide the ability to determine the segmentation of blood donors to improve the prediction accuracy of managers. We use the Fuzzy k-means algorithm for the number of clusters 2, 3, 4, and 5 then we use the silhouette method to determine the quality of a cluster. The silhouette results show that the optimal cluster is 3 which is confirmed through the testing phase. The results of blood donor clustering illustrate that men have a tendency to donate blood.

Keywords: *Clustering, Fuzzy K-Means, Silhouette, Blood Donation*

1. PENDAHULUAN

Salah satu masalah utama yang dihadapi dalam sistem pelayanan kesehatan di negara berkembang adalah kurangnya persediaan darah aman yang memadai untuk keperluan transfusi darah bagi yang membutuhkannya. Transfusi darah adalah proses pemindahan darah dari donor ke sakit/membutuhkan (resipien). Transfusi darah merupakan hal yang penting dalam pelayanan kesehatan. Sebab dengan penerapan yang benar, transfusi darah dapat menyelamatkan nyawa pasien ataupun meningkatkan derajat kesehatan pasien. Namun dalam beberapa kasus, ketersedian akan darah yang tidak pasti dan sulit di kontrol. Artinya pasokan darah bisa saja habis saat permintaan tinggi dan bisa pula kadaluwarsa akibat terlalu lama dalam penyimpanan. Sehingga, 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[5] dan untuk memenuhi target Teresian darah

Tujuan ini di lakukan untuk mengolah data lebih lanjut untuk menghasilkan informasi baru. Dengan mengelompokan sejumlah data yang di clustering untuk mendapatkan pengelompokan objek data ke dalam kelompok berdasarkan kemiripannya. Dan salah satu algoritme yang di gunakan dalam clustering adalah algoritma Fuzzy-Kmeans. Fuzzy C-Means menggunakan model pengelompokan Fuzzy sehingga data dapat menjadi anggota dari semua kelas atau cluster yang terbentuk dengan derajat atau tingkat keanggotaan yang berbeda. Dalam konsep fuzzy, Adapun kelebihan dari fuzzy k-means adalah mampu menempatkan suatu data yang terletak diantara dua atau lebih gerombol yang lain pada suatu gerombol dan [1]. Adapula Metode FCM memiliki

beberapa kelemahan, antara lain membutuhkan banyaknya kelompok dan matriks keanggotaan kelompok yang ditetapkan sebelumnya[2]. Sehingga Untuk mengatasi keterbatasan-keterbatasan yang ada pada FCM, banyak data scientist yang mengusulkan dan menerapkan algoritme-algoritme optimasi[3]. Atau Untuk menghubungkan masalah kekurangan dan kelebihan kedua metode tersebut diusulkan sebuah metode baru yang merupakan penggabungan dari keduanya disebut *Subtractive Fuzzy C-Means* (SFCM). Metode ini digunakan Liu, Xiao, Wang, Shi, & Fang, dalam penelitiannya dan menyimpulkan bahwa SFCM memberikan solusi yang lebih baik dibandingkan dengan FCM serta memberikan kecepatan yang lebih tinggi dalam konvergensi fungsi objektif[4].

Fuzzy C-Means

Konsep dasar fuzzy k-means pertama kali adalah menentukan pusat cluster pada kondisi awal, pusat cluster ini masih belum akurat dan tiap objek memiliki derajat keanggotaan untuk tiap-tiap cluster dengan cara memperbaiki pusat cluster dan nilai keanggotaan tiap objek secara berulang maka akan dapat dilihat bahwa pusat cluster akan bergerak menuju lokasi yang tepat.

Ketika gerombol-gerombol menjadi overlapping atau setiap objek memungkinkan termasuk ke beberapa gerombol, maka μ_{ik} dapat diinterpretasikan sebagai fungsi keanggotaan yaitu $\mu_{ik} \in [0,1]$. Maka fungsi objektif J yang dirumuskan sebagai fungsi dari U dan V sebagai berikut:

$$J(U, V) = \sum_{k=1}^n \sum_{l=1}^c \mu_{ik} d^2(x_k, v_l)$$

dengan:

U : matriks keanggotaan objek ke masing-masing gerombol V : matriks centroid / rata masing-masing gerombol μ_{ik} : fungsi keanggotaan objek ke- k ke gerombol ke- i x_k : objek ke- k v_i : nilai centroid gerombol ke- i d : ukuran jarak

2. HASIL DAN PEMBAHASAN

2.1 Data

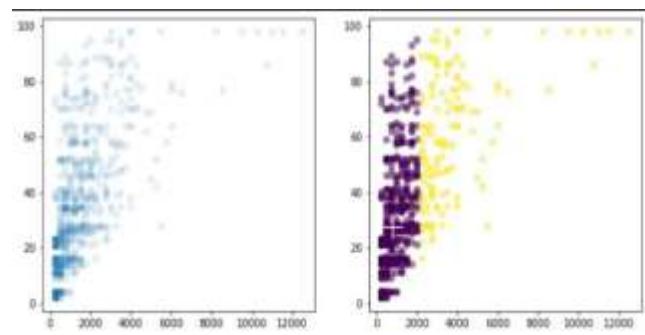
Menggunakan 748 data donor darah yang di dapatkan dari Kaggle yang bertujuan medapatkan hasil dari pengumpulan pendonor selama tiga bulan dengan menggunakan sample yang di pakai :

	Recency (months)	Frequency (times)	Monetary (c.c. blood)	Time (months)	whether he/she donated blood in March 2007
0	2	80	1200	90	1
1	0	19	3250	28	1
2	1	10	4000	35	1
3	2	20	5000	45	1
4	1	24	6000	77	0

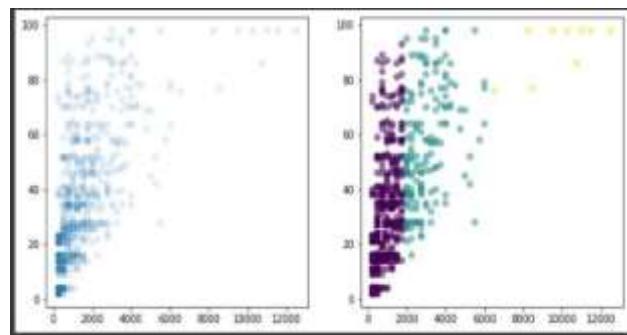
- Recency (months),
- Frequency (times),
- Monetary (c.c. blood),
- Time (months),
- whether he/she donated blood in March 2007

2.2 Pra-Pemrosesan Data

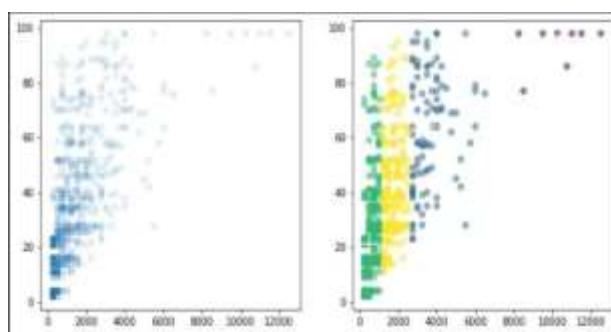
Dalam proses data menggunakan metode fuzzy C-Kmeans yang digunakan untuk mendapatkan hasil silhouette yang di lakukan dengan pengujian clustering 2, 3, 4 dan 5 untuk mendapatkan nilai tertinggi.



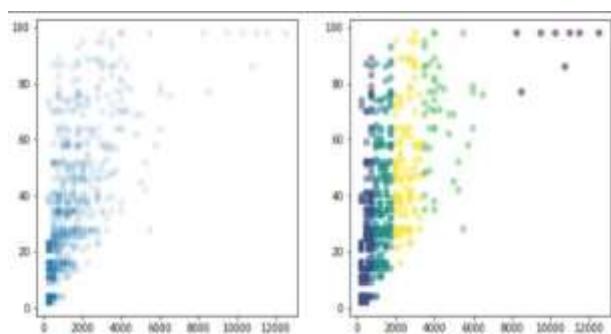
Gambar 1. Cluster n2



Gambar 2. Cluster n3



Gambar 3. Cluster n4



Gambar 4. Cluster n5

2.3 Eksperimen

Data set yang di gunakan untuk pengujian metode Fuzzy C-Kmeans adalah recency dan frequency data di pakai sebagai tahap uji untuk mendapatkan berapa aktif pendonor yang mendonor selama tiga bulan yang di kumpulkan melalui tahap clustering dengan metode Fuzzy C-Kmeans yang dapat di lihat dari gambar 5.

```
[1] features = iris.drop(['Recency', 'Frequency'], axis = 1)
species = iris.drop(['Recency', 'Frequency', 'Monetary', 'Time', 'whether he/she donated blood in March 2007'], axis = 1)
```

Gambar 5. Data Set

Dari data set yang di pakai telah dilakukan beberapa tahap clustering dan di dapatkan nilai tertinggi untuk hasil shilhoutte terpadat pada clustering n2 dengan nilai = 0.6867772288409422 yang menunjukan hasil tertinggi dengan melakukan tahap pengujian clustering dari 2,3,4 dan 5 dari hasil yang dapat di lihat dari gambar 6 di bawah ini.

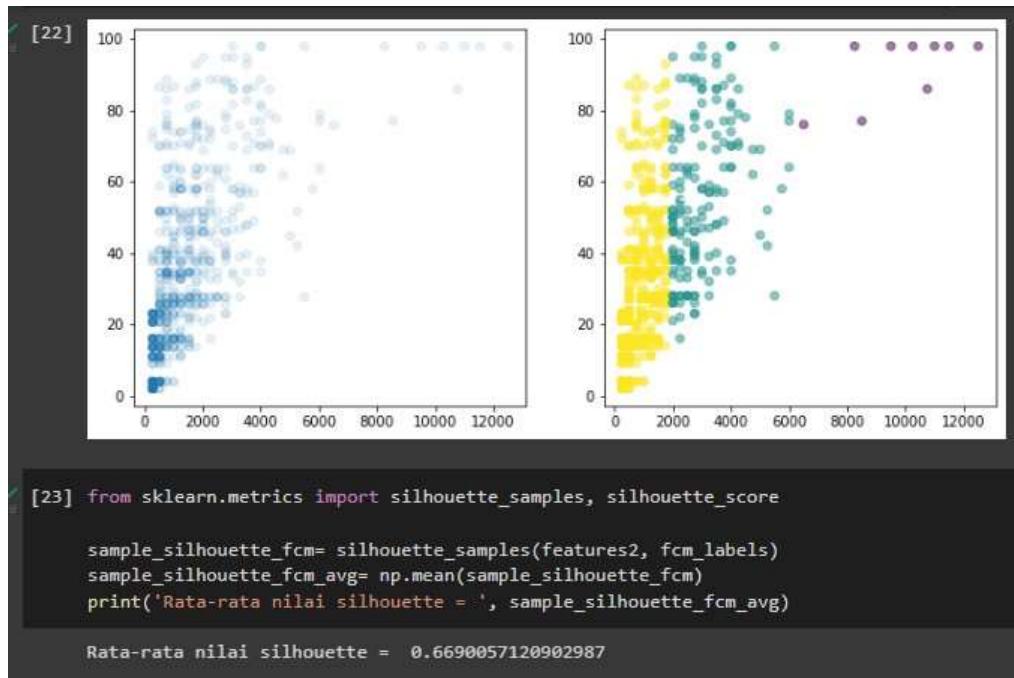


Gambar 6. Hasil Silhouette N2

Dan di lakukan lagi tahap perbandingan pada clustering n3 untuk melihat apakah hasil dari clustering n2 memang terbukti memiliki hasil shilhoutte tertinggi dengan pengujian terhadap clustering n3 yang dapat di lihat pada gambar 7. Melihat hasil shilhouette pada n3 memiliki nilai yang lebih rendah pada gambar 8, sehingga dapat disimpulkan bahwa benar n2 memiliki data tertinggi yang dapat di artikan bahwa pendonor lebih sering di lakukan pada pria.

```
[17] from fcmeans import FCM
fcm = FCM(n_clusters=3)
```

Gambar 7. Clustering n3



Gambar 8. Silhouette N3

2.4 Evaluasi

Proses pengujian pada data yang sudah tersedia dan yang dipakai dalam mendapatkan hasil akurasi dengan menggunakan metode fuzzy c-means dan sudah menampilkan hasil yang baik sesuai dengan hasil yang di butuhkan untuk melihat perbandingan silhouette tertinggi.

2.5 Analisis

Dari hasil penelitian data tersebut bahwa di ketahui bahwa (1 berarti mendonor darah; 0 berarti tidak mendonor darah) dari data tersebut untuk mendapatkan kecendurungan pendonor maka dilakukannya tahap clustering dengan metode fuzzy c-means yang memiliki tingkat akurasi dan hasil yang cukup baik sehingga bisa di dapatkan hasil perbandingan yang dapat menentukan hasil tertinggi dari pengujian tersebut dengan tujuan untuk mendapatkan informasi pendonor yang aktif mendonor selama tiga bulan.

KESIMPULAN

Dalam pelayanan Kesehatan, transfusi darah menjadi hal yang penting. Hal tersebut menjadi penting untuk meningkatkan derajat kesehatan dan menyelamatkan nyawa pasien. Namun, ketersediaan darah tidak pasti dan sulit dikontrol. Untuk mengatasinya, Kota Hsin- Chu di Taiwan mengumpulkan darah dengan adanya 748 pendonor untuk melihat kemungkinan pendonor mendonor setiap tiga bulan. Tujuannya adalah untuk mengolah lanjut data untuk menghasilkan informasi.

Penelitian ini menggunakan metode clustering untuk menentukan segmentasi donor darah untuk meningkatkan akurasi prediksi. Adapun data set yang di gunakan untuk pengujian metode Fuzzy C-Kmeans adalah recency dan frequency data di pakai sebagai tahap uji untuk mendapatkan jumlah aktif pendonor yang mendonor selama tiga bulan yang di kumpulkan melalui tahap clustering dengan metode Fuzzy C-Kmeans. Hasil clustering donor darah menggambarkan bahwa laki-laki memiliki kecenderungan untuk mendonorkan darahnya.

Saran

Pada proses penelitian ini masih banyak sumber data yang cukup sulit untuk di dapatkan maka menurut penulis, proses menyediakan data sangat penting sehingga untuk tujuan clustering yang dapat memberikan hasil akurasi yang lebih baik lagi maka dibutukan data yang lebih. Dan penggunaan metode fuzzy c-means masih memiliki banyak kekurangan sehingga menurut penulis untuk hasil yang lebih baik dapat dilakukan nya beberapa tahap metode lain untuk dipakai sebagai perbandingan dalam proses clustering

DAFTAR PUSTAKA

- [1] Lathifaturrahmah, "Perbandingan Hasil Penggerombolan K-Means, Fuzzy Kmeans, Dan Two Step Clustering," pp. 39-62, 1 Juli – Desember 2014.
- [2] R. K. Baiq Nurul Haqiqi1, "Analisis Perbandingan Metode Fuzzy," vol. Vol. 8 No. 2, pp. 59-67, Desember 2015.
- [3] R. K. Edhi Prabowo1, "Optimasi Algoritma Fuzzy Clustering dengan," vol. Volume 4 No.1, Pp. 1-6, Januari 2019.
- [4] M. N. H. G. Widya Suermi1, "Penerapan Metode Subtractive Fuzzy C-Means," Vol. Volume 2 Nomor 2, pp. 63-74, Desember 2020.
- [5] L. F. Titania Dwi Andini1, "Peningkatan Ketersediaan Darah Sesuai Segmentasi Umur," vol. Volume 12 Nomor 2, pp. 1-18, Oktober 2022.