

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Klaster

Analisis klaster adalah pengelompokan data untuk pembentukan kategori baru dengan mengidentifikasi kelompok yang memiliki karakteristik umum tertentu. Objek-objek pada kelompok (klaster) yang akan dibentuk mirip satu sama lain dan berbeda dengan objek kelompok lain. Analisis klaster juga disebut analisis klasifikasi atau taksonomi numerik (*numerical taxonomi*), karena berkenaan dengan prosedur pengklasretan dimana setiap objek hanya masuk ke dalam satu klaster saja, tidak terjadi tumpang tindih (Supranto, 2004). Kebanyakan metode pengklasteran adalah teknik yang relatif sederhana yang tidak didukung oleh argument yang luas. Kebanyakan metode pengklasteran berdasarkan pada algoritma (*algorithm*). Jadi, analisis klaster berbeda sekali jika dibandingkan dengan analisis varian, regresi berganda, analisis diskriminan dan analisis faktor berdasarkan oleh penalaran statistik yang cukup luas.

Walaupun ada banyak permetodean pengelompokan yang memiliki sifat statistik yang terpenting, kesederhanaan metode ini harus dimengerti. Solusi klaster analisis bersifat berbeda atau tidak biasa, anggota klaster untuk setiap pengerjaan/solusi itu tergantung pada beberapa bagian dari prosedur dan beberapa penyelesaian yang berbeda pendapat dengan mengganti satu atau lebih elemen. Solusi pengelompokan secara keseluruhan tergantung pada variabel-variabel yang digunakan sebagai dasar untuk skor kesamaan. Pengurangan atau penambahan variabel-variabel yang relevansi dapat membuat perbedaan kekonsistenan hasil

analisis kluster. Klustering hampir sama dengan klasifikasi, letak perbedaan terdapat pada kategori yang ada. Pada proses klustering, data tidak dimasukkan ke dalam penggolongan suatu kelompok dengan variabel target tertentu (Event, 2021).

2.2 Metode Pengklasteran

Terdapat banyak metode dalam analisis kluster untuk mengelompokkan observasi di dalam kluster. Secara umum metode pengelompokkan analisis kluster terdiri dari dua metode, yaitu metode hirarki dan metode non-horarki. Metode hirarki merupakan pengelompokkan data yang belum diketahui jumlah kelompok sebelumnya. Sedangkan non-hirarki merupakan metode pengelompokkan data yang sudah diketahui jumlah kelompok yang diinginkan (nilai k telah ditentukan sebelumnya). Salah satu algoritma yang termasuk dalam non-hirarki adalah algoritma k -means. Secara umum proses metode non hirarki adalah sebagai berikut :

1. Pilih *centroid* kelompok awal atau *seed*.
2. Tentukan setiap objek ke dalam kelompok jarak terpendek.
3. Tampilkan lagi setiap objek ke bagian kelompok sesuai peraturan penghentian yang sudah ditetapkan.
4. Jika tidak ada objek yang berubah lagi proses akan dihentikan, apabila belum ulangi langkah kedua.

Untuk memperoleh *seed* awal, menggunakan beberapa aturan sebagai berikut:

1. Pilih objek awal yang tidak akan ada lagi data yang hilang sebagai *centroid*.

2. Pilih objek awal tidak akan ada lagi data yang hilang sebagai *seed* kelompok pertama, setelah itu *seed* kelompok kedua dipilih dari objek yang memiliki jarak paling jauh dari sebelumnya dan selanjutnya. Pilih dengan acak objek dengan tidak ada data yang hilang sebagai pusat klaster.
3. Memperbaiki *seed* yang dipilih dengan memakai peraturan tertentu sampai jarak *seed* tersebut semakin jauh.
4. Gunakan pengetahuan tentang identifikasi pusat klaster sehingga jarak pusat klaster semakin jauh.
5. Gunakan *seed* yang disajikan oleh peneliti.

2.2.1 Metode K-Means

K-means merupakan salah satu metode data klastering non hierarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih klaster atau kelompok sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu klaster yang sama dan data yang mempunyaikarakteristik yang berbeda dikelompokkan ke dalam kelompok yang lainnya (Metisen et al, 2015)

2.2.2 Komponen K-Means

Metode k-means adalah metode klastering berbasis jarak yang membagi ke dalam sejumlah klaster dan algoritma ini hanya bekerja pada atribut numerik. Pada algoritma k-means diperlukan 3 komponen yaitu :

1. Jumlah klaster k

Jumlah k dapat ditentukan melalui pendekatan metode hirarki. Namun perlu diperhatikan bahwa tidak terdapat aturan khusus dalam menentukan jumlah

kluster k , terkadang jumlah kluster yang diinginkan tergantung pada subjektif seseorang.

2. Kluster Awal

Kluster awal yang dipilih berkaitan dengan penentuan pusat kluster awal.

3. Ukuran jarak

Ukuran jarak digunakan untuk menempatkan observasi ke dalam kluster berdasarkan *centroid* terdekat. Ukuran jarak yang digunakan disebut jarak *Euclid*.

2.2.3 Algoritma K-Means

Algoritma k-means merupakan salah satu algoritma yang multivariat dengan tujuan untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Data-data dengan karakteristik yang sama akan dikelompokkan menjadi satu kelompok dan data dengan karakteristik yang berbeda dikelompokkan dalam kelompok lainnya, sehingga data dalam satu kelompok mempunyai tingkat variasi yang kecil. Algoritma k-means menggunakan *descriptive model* untuk menjelaskan algoritma pengelompokan yang dilakukannya (Putri et al., 2020). K-means adalah satu dari beberapa metode klustering *non-hierarchical* yang secara umum ditunjukkan dengan algoritma sebagai berikut:

1. Tentukan jumlah kluster (k) yang digunakan sesuai dengan masalah yang diteliti.
2. Tentukan titik pusat (*centroid*) kluster awal secara acak.
3. Hitung jarak *centroid* ke titik pusat kluster setiap data. Ukuran jarak yang dipakai adalah *Euclidean Distance* menggunakan rumus berikut :

$$d_{ij} = \sqrt{\sum_{k=1}^n (x_{ik} - x_{jk})^2} \quad (2.1)$$

Dengan :

d_{ij} = jarak *euclidean* antar objek ke-i dengan objek ke-j

x_{ik} = nilai dari objek ke-i pada peubah ke-k

x_{jk} = nilai dari objek ke-j pada peubah ke-k

4. Lakukan pengelompokkan objek ke titik pusat berdasarkan jarak terdekat.
5. Lalu kerjakan iterasi selanjutnya dengan mengulangi langkah 2 dan 4.

Untuk mengulang langkah kedua harus ditentukan titik pusat atau *centroid* kluster yang baru. *Centroid* kluster adalah rata-rata dari semua data dalam suatu kluster, dicari dengan persamaan berikut:

$$V = \frac{1}{N} \sum_{k=0}^N X_k \quad (2.2)$$

6. Ulangi perhitungan jarak *centroid* dengan setiap data (ulangi pada langkah ketiga) apabila masih terdapat data yang berpindah kluster.
7. Jika titik pusat baru sama dengan titik pusat lama maka tidak perlu lagi dilakukan iterasi selanjutnya.

2.3 *Silhouette Coeficent*

Untuk mengatasi kesulitan algoritma k-means dalam menentukan jumlah kluster yang tepat berdasarkan data yang digunakan. Maka digunakan indeks validasi yang merupakan metode yang mengevaluasi hasil algoritma klustering dengan tujuan mendapatkan kluster terbaik. Pada penelitian ini akan digunakan *Silhouette Coeficent* sebagai penentuan kluster terbaik.

Untuk melihat kualitas hasil pengelompokan masing-masing perhitungannya jarak, maka perlu dilakukan uji homogenitas. Pengujian dilakukan menggunakan persamaan *Silhouette Coefficient*. Langkah dalam perhitungan *Silhouette Coefficient* yaitu di mulai dengan mencari jarak rata-rata data ke- i dengan semua data di klaster yang sama ($a(i)$). Dan menghitung nilai $b(i)$ jarak rata-rata data ke- i dengan semua data di klaster yang berbeda ($b(i)$) (A.Struyf et al., 1997). Jarak antar data dihitung dengan menggunakan rumus *euclidean distance*. Setelah mencari nilai $a(i)$ dan $b(i)$, selanjutnya menghitung nilai *Silhouette Coefficient* ($s(i)$) dengan rumus sebagai berikut :

$$s(i) = \frac{b(i)-a(i)}{\max(a(i),b(i))} \quad (2.3)$$

2.4 Software Pendukung

Dalam penyelesaian metode k-means pada penelitian ini menggunakan software pendukung yaitu R Studio. R Studio dapat digunakan untuk pengelolaan metode k-means klustering dan *Silhouette Coefficient*. Penggunaan R Studio dapat memberi pengguna kemudahan untuk melakukan komunikasi dengan R Studio sehingga dapat menjalankan fungsi statistika dan data science. R Studio sendiri merupakan salah satu program yang sering digunakan dalam mengolah dan menganalisis data. Hal tersebut tentunya karena R Studio memiliki kelebihan, antara lain :

1. Antar muka atau interface R Studio dapat diatur sedemikian rupa sehingga pengguna dapat melihat tabel data, kode R, grafik, dan semua output dengan jelas dalam waktu bersamaan.

2. Fitur Import – Wizard yang dapat digunakan untuk mengimpor file Excel, CSV, SAS (.sas7bdat), Stata (*.dta), dan SPSS (*.sav) ke dalam R.
3. Memiliki Addins yang dapat difungsikan dalam menjalankan proses secara interaktif (*colourpicker*, *bookdown*, dll).

2.5 Puskesmas

Puskesmas merupakan sarana pelayanan kesehatan terkemuka dengan misi sebagai pusat pengembangan pelayanan kesehatan, dengan tugas melaksanakan pembinaan, pelayanan kesehatan secara menyeluruh dan terpadu kepada masyarakat di daerah tertentu. Pelayanan kesehatan yang dilakukan secara menyeluruh, meliputi aspek-aspek; *promotif* (upaya peningkatan), *preventif* (upaya pencegahan), *kuratif* (upaya penyembuhan), dan *rehabilitative* (upaya pemulihan). Keempat aspek ini harus berjalan secara bersamaan dan tidak boleh ada yang terabaikan (Ramadani, 2021).

Di era globalisasi, perkembangan sektor jasa semakin berkembang dalam usaha meningkatkan kesejahteraan dan kesehatan masyarakat. Puskesmas sebagai pusat terdepan, yang bertanggung jawab secara luas dalam upaya pelayanan kesehatan setiap orang. Upaya kesehatan masyarakat memerlukan campur tangan petugas puskesmas dengan kualitas tinggi (Lestari, 2016). Tugas Puskesmas adalah melaksanakan kebijakan kesehatan untuk mencapai tujuan pembangunan kesehatan di wilayah kerjanya dalam rangka mendukung terwujudnya kecamatan sehat (Permenkes No.75 Tahun 2014 tentang Puskesmas).

2.6 Sumber Daya Manusia (SDM) Kesehatan

Menurut Kebijakan Kesehatan Indonesia tahun 2014 sumber daya manusia kesehatan adalah bagian dari sistem kesehatan nasional yang dianggap sebagai komponen utama sebagai penggerak pembangunan kesehatan, dengan tujuan untuk meningkatkan kewaspadaan, kesadaran, kemauan serta kesempatan hidup sehat bagi semua orang supaya terwujud derajat kesehatan yang setinggi-tingginya. Indonesia termasuk dalam 57 negara yang menghadapi krisis tenaga kesehatan. Padahal 80% keberhasilan pembangunan kesehatan ditentukan oleh sumber daya manusia kesehatan. Distribusi dari tenaga kesehatan masih menjadi isu sistem kesehatan di berbagai Negara di dunia termasuk di Indonesia dimana masih adanya ketidakmerataan dari distribusi SDM Kesehatan. Secara geografis, Indonesia merupakan negara yang memiliki daerah yang sulit dijangkau sehingga hal tersebut dapat mempengaruhi minat tenaga kesehatan di Indonesia (Ramadhanti, 2021).

Upaya meningkatkan kualitas dan kuantitas SDM kesehatan menjadi salah satu indikator keberhasilan pembangunan dalam rangka memperoleh sasaran pembangunan kesehatan. Hingga sampai sekarang ini masih banyak kendala yang dihadapi puskesmas di Indonesia terhadap pengelolaan tenaga kesehatan salah satunya tentang distribusi SDM yang tidak merata. Kurangnya jumlah SDM sangat berhubungan dengan mutu/kualitas pelayanan di puskesmas. Sebagai gambaran pada data yang diperoleh ditemukan beberapa permasalahan terkait SDM kesehatan puskesmas di Indonesia yang tidak merata dan terdapat beberapa indikator-indikator sebagai berikut :

1. Jumlah tenaga medis puskesmas di Indonesia tahun 2021.

2. Jumlah tenaga kesehatan masyarakat, kesehatan lingkungan, dan gizi puskesmas di Indonesia tahun 2021.
3. Jumlah tenaga teknik biomedika, keteknisan medis puskesmas di Indonesia tahun 2021.
4. Jumlah tenaga kefarmasian puskesmas di Indonesia tahun 2021.
5. Jumlah tenaga penunjang/pendukung kesehatan puskesmas di Indonesia tahun 2021