

BAB II

TINJAUAN PUSTAKA

2.1 Analisis Cluster

Analisis *cluster* adalah salah satu dari metode dalam analisis multivariat yang memiliki tujuan utama untuk mengelompokkan objek-objek berdasarkan karakteristik yang dimilikinya. Analisis *cluster* mengelompokkan individu atau objek penelitian, sehingga setiap objek yang paling dekat kesamaannya dengan objek lain berada dalam *cluster* yang sama. *Cluster-cluster* yang terbentuk dalam satu *cluster* mempunyai ciri yang relatif sama (homogen), sedangkan antar *cluster* mempunyai ciri yang berbeda (heterogen). Pengelompokan ini dilakukan berdasarkan variabel-variabel yang diamati (Usman dan Sobari, 2013).

Untuk mendapatkan kelompok yang sehomogen mungkin, maka yang digunakan dasar untuk mengelompokkan adalah kesamaan skor nilai yang dianalisis. Semakin kecil besaran jarak suatu individu terhadap individu lain, maka semakin besar kemiripan individu tersebut. Data mengenai ukuran kesamaan tersebut kemudian dilakukan pengelompokan sehingga dapat ditentukan individu mana yang masuk kelompok mana (Gudono, 2014).

Ciri-ciri *cluster* yang baik yaitu mempunyai:

1. Homogenitas (*within-cluster*), yaitu kesamaan yang tinggi antar anggota dalam satu *cluster*.

2. Heterogenitas (*between-cluster*), yaitu perbedaan tinggi antar kluster yang satu dengan *cluster* yang lain.

2.1.1 Uji Multikolinieritas

Uji multikolinieritas antar variabel adalah salah satu pelanggaran asumsi dalam analisis *cluster*. Multikolinieritas adalah suatu keadaan dimana ada hubungan linier sempurna atau hampir sempurna antara beberapa variabel atau semua variabel dalam analisis *cluster*.

Salah satu cara yang digunakan dalam uji multikolinieritas adalah dengan menghitung *Variance Inflation Factor (VIF)*. Dikatakan bahwa variabel tersebut terjadi multikolinieritas jika nilai $VIF \geq 10$ (Gudono, 2014).

Hipotesis:

H_0 : Tidak ada hubungan linier antar variabel independen

H_1 : Ada hubungan linier antar variabel independen

Statistik uji:

$$VIF = \frac{1}{1-R_j^2} \quad (2.1)$$

dengan:

R_j = koefisien determinasi

Daerah penolakan:

Tolak H_0 jika $VIF \geq 10$

Sedangkan untuk rumus koefisien detereminasi sebagai berikut:

$$R^2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})^2 - (Y_i - \hat{Y})^2}{\sum_{i=1}^n (Y_i - \bar{Y})^2} \quad (2.2)$$

dengan:

R_j = koefisien determinasi

Y_i = nilai data ke-i

\hat{Y} = nilai Y dari persamaan regresi

\bar{Y} = nilai rata-rata variabel Y

i = 1, 2, 3, ... , n

2.1.2 Ukuran Jarak

Dalam penelitian ini ukuran jarak digunakan untuk mengetahui seberapa mirip atau berbeda antara objek-objek yang diteliti menggunakan jarak *euclidean*. Semakin kecil/pendek jarak antar objek, maka semakin mirip objek tersebut dengan objek lainnya. Jarak *euclidean* merupakan besarnya jarak suatu garis lurus yang menghubungkan antar objek yang diteliti. Jarak *euclidean* mengukur jumlah kuadrat perbedaan nilai antar masing-masing variabel. Prasetyo (2014) menyatakan bahwa jarak *euclidean* merupakan ukuran jarak yang dapat memberikan jarak terdekat atau terpendek antar dua data.

$$d_{ij} = \sqrt{\sum_{k=1}^p (x_{ik} - x_{jk})^2} \quad (2.3)$$

dengan:

d_{ij} = jarak antara objek ke-i dan objek ke-j

p = jumlah variabel *cluster*

x_{ik} = data dari subjek ke-i pada variabel ke-k

x_{jk} = data dari subjek ke-j pada variabel ke-k

$$k = 1, 2, 3, \dots, p$$

$$i = 1, 2, 3, \dots, n$$

$$j = 1, 2, 3, \dots, n$$

Dalam pengelompokan data terdapat 2 metode yang dapat digunakan, yaitu:

i) Metode Hierarki

Metode hierarki adalah suatu metode analisis *cluster* yang membentuk tingkatan tertentu seperti pada struktur pohon karena proses pengclusteran dilakukan secara bertingkat atau bertahap. Metode ini biasa digunakan untuk individu yang tidak terlalu banyak, dan jumlah kelompok yang akan dibentuk belum diketahui. Hasil dari pengclusteran dengan metode hierarki adalah dengan bentuk dendogram. Dalam metode hierarki terdapat dua prosedur yang digunakan untuk pengclusteran yaitu metode agglomeratif (*agglomerative method*) dan metode divisif (*devisive method*). Hasil pengelompokan biasanya ditampilkan dalam bentuk diagram dendogram. Dendogram digunakan untuk memperjelas pengelompokan dalam metode hierarki (Jhonson and Winchern, 2007)

ii) Metode Non-Hierarki

Perbedaan dasar dalam metode non-hierarki dan metode hierarki ini adalah ditentukan terlebih dahulu jumlah kelompok sebelum pengelompokan dilakukan. Metode mengelompokkan individu atau objek dalam kelompok sedemikian rupa sehingga jarak antar individu ke pusat kelompok minimum. Metode non-hierarki yang paling banyak digunakan adalah metode *K-means*.

Metode non-hierarki ini dimulai dengan memilih sejumlah nilai *cluster* awal sesuai yang diinginkan. Selanjutnya objek-objek penelitian digabungkan dalam *cluster-cluster* tersebut (Usman dan Sobari, 2013).

2.2 *K-Means Cluster*

K-means cluster adalah salah satu metode yang membutuhkan jumlah *cluster* yang ditentukan terlebih dahulu sebagai k , dan kemudian membagi n objek kedalam k *cluster*. Kemiripan antar anggota dalam satu *cluster* sangat tinggi sedangkan kemiripan antar anggota dengan *cluster* yang lain sangat rendah. Kemiripan anggota pada *cluster* diukur dengan nilai kedekatan objek terhadap mean pada *cluster* atau *centroid cluster* (Muzakir, 2014).

K-means merupakan metode non-hierarki yang paling banyak digunakan untuk pengelompokan karena kemudahan dan kemampuannya untuk mengelompokan dalam jumlah data yang besar dengan waktu komputasi yang cepat dan efisien. *K-means* melihat jarak antar objek terhadap nilai *centroid cluster*. *K-Means* mengelompokan secara tegas data hanya pada satu *cluster*, dengan menggunakan ukuran jarak terkecil terhadap *centroid*.

Langkah-langkah melakukan pengelompokan dengan menggunakan metode *K-means* sebagai berikut:

1. Tentukan jumlah *cluster* k
2. Menentukan nilai *centroid cluster*. Nilai *centroid* merupakan rata-rata objek dalam *cluster* tersebut.

$$C_{ij} = \frac{1}{n_i} \sum_{k=1}^{n_i} x_{jk} \quad (2.4)$$

dengan:

C_{ij} = pusat *cluster* ke- i pada variabel ke- j

x_{jk} = nilai pengamatan pada objek ke- k pada variabel ke- j

n_i = jumlah peubah yang menjadi anggota *cluster* ke- i

3. Menentukan jarak terdekat setiap objek terhadap tiap *centroid cluster* dengan menggunakan ukuran jarak *euclidean* (persamaan 2.3)
4. Hitung kembali *centroid* dari anggota *cluster* yang baru terbentuk
5. Jika *centroid cluster* tidak berubah maka langkah terhenti, namun jika nilai *centroid* berubah maka lakukan langkah ke 3 dan 4 sampai nilai *centroid* tidak berubah lagi.

2.3 Logika Fuzzy

Pertama kali diperkenalkan oleh Prof. Lotfi A. Zadeh pada tahun 1965. Logika *fuzzy* digunakan untuk memetakan permasalahan dari *input* ke *output* yang diharapkan. Dalam logika *fuzzy* terdapat *fuzzy cluster* yang merupakan metode untuk menentukan *cluster* optimal dalam suatu vektor dengan bentuk normal *euclidean* untuk jarak antar vektor (Ramdhana dan Efendi, 2017). Dalam logika *fuzzy* metode pengelompokan yang sering digunakan adalah metode *Fuzzy C-Means*.

2.4 Fuzzy C-Means Cluster (FCM)

Metode ini pertama kali ditemukan oleh Dunn pada tahun 1973 dan selanjutnya pada tahun 1981 dikembangkan oleh Bezdek. *Fuzzy c-means* merupakan pengembangan dari metode *cluster* non-hierarki yaitu *K-means* dengan menggunakan logika *fuzzy*. Metode ini memungkinkan keanggotaan suatu objek tidak hanya pada satu *cluster* tapi lebih dari satu *cluster*. Keanggotaan suatu *cluster* didasarkan pada nilai derajat keanggotaanya. Dimana nilai derajat keanggotaanya dihitung dengan menggunakan rumus berikut:

$$u_{ik} = \frac{(d_{ik})^{-2/(m-1)}}{\sum_{j=1}^c (d_{jk})^{-2/(m-1)}} \quad (2.5)$$

dengan:

u_{ik} = derajat keanggotaan objek ke- k cluster ke- i

d_{ik} = jarak *euclidean* antara objek ke- k dengan pusat *cluster* ke- i

m = *weighted exponent* yang nilainya dari 1 sampai tak hingga

c = jumlah *cluster* yang akan dibentuk

Selanjutnya menghitung *centroid cluster* dengan menggunakan rumus sebagai berikut:

$$v_i = \frac{\sum_{k=1}^n (u_{ik})^m \times x_k}{\sum_{k=1}^n (u_{ik})^m} \quad (2.6)$$

dengan:

v_i = pusat *cluster* ke- i

u_{ik} = derajat keanggotaan objek ke- k ke pusat *cluster* ke- i

m = *weighted exponent* yang nilainya dari 1 sampai tak hingga

n = jumlah objek

Metode FCM untuk *centroid* awal masih belum akurat karena setiap objek memiliki derajat keanggotaan untuk tiap *cluster*. Sehingga dilakukan perulangan untuk memperbaiki *centroid* dan derajat keanggotaan, maka *centroid cluster* akan bergerak menuju tempat yang tepat. Perulangan didasarkan untuk meminimalkan fungsi objektif yang digunakan untuk menggambarkan jarak antara objek terhadap *centroid*. Fungsi objektif yang digunakan yaitu:

$$J_{FCM}(U,V) = \sum_{k=1}^n \sum_{i=1}^c d_{ik}^2 (u_{ik})^m \quad (2.7)$$

dengan:

d_{ik} = jarak *euclidean* antara objek ke- k dengan *centroid cluster* ke- i

$$(d_{ik} = d_{ki})$$

u_{ik} = derajat keanggotaan objek ke- k *cluster* ke- i

$$(u_{ik} \in [0,1], \sum_{i=1}^c u_{ik} = 1)$$

n = jumlah objek

c = jumlah *centroid cluster* yang akan terbentuk

m = *weighted exponent* yang nilainya dari 1 sampai tak hingga

Langkah-langkah pengelompokan dengan menggunakan metode FCM sebagai berikut:

1. Menentukan jumlah *cluster*
2. Membangkitkan bilangan random u_{ik} dengan i merupakan banyaknya data dan k merupakan banyaknya kelompok sebagai elemen-elemen awal matriks keanggotaan awal U

3. Menghitung *centroid*
4. Menghitung fungsi objektif
5. Menghitung perubahan matriks keanggotaan
6. Cek kondisi berhenti
 - a. Jika $|J_t - J_{t-1}| < \varepsilon$ atau $t > \maxIter$ maka berhenti
 - b. Jika tidak : $t=t+1$ ulangi langkah 3.

Output yang dihasilkan FCM merupakan *centroid cluster* dan derajat keanggotaan untuk tiap-tiap objek.

2.5 Perbandingan Hasil Pengelompokan

Membandingkan hasil pengelompokan yang baik dengan melihat nilai simpangan baku dalam kelompok dan antar kelompok yang terbentuk. Semakin kecil nilai simpangan baku maka semakin bagus hasil pengelompokan tersebut. Rata-rata simpangan baku dalam kelompok didapatkan dengan rumus berikut:

$$S_w = \frac{1}{C} \sum_{i=1}^C S_i$$

$$S_i = \sqrt{\frac{1}{n_i-1} \sum_{i=1}^{n_i} (X_i - \bar{X})^2}$$

dengan:

S_w = rata-rata simpangan baku didalam kelompok

S_i = simpangan baku kelompok ke- i

C = jumlah kelompok yang akan dibentuk

n_i = jumlah objek yang menjadi anggota kelompok ke- i

Sedangkan untuk simpangan baku antar kelompok dapat diketahui dengan menggunakan rumus sebagai berikut:

$$S_B = \sqrt{\frac{1}{C-1} \sum_{i=1}^C (\bar{X}_i - \bar{X})^2}$$

dengan:

S_B = rata-rata simpangan baku antar kelompok

\bar{X}_i = rata-rata objek pada kelompok ke- i

\bar{X} = rata-rata objek pada semua kelompok

C = jumlah kelompok yang akan dibentuk

2.6 Demam Berdarah Dengue (DBD)

Demam berdarah dengue (DBD) adalah penyakit yang disebabkan oleh virus dengue yang tergolong *Arthropod-Borne Virus*, genus *Flavivirus* dan *Flaciciridae*. DBD ditularkan melalui gigitan nyamuk genus *Aedes*, terutama *Aedes Aegypti* atau *Aedes albopictus*. Pada wilayah tropis dan subtropis penyakit DBD merupakan endemik yang muncul sepanjang tahun, terutama saat musim hujan karena nyamuk berkembang biak secara optimal. Demam berdarah dengue memiliki beberapa inveksi virus dengan gejala yang dibagi menjadi 3 yaitu: (a). Demam dengue tanpa gejala spesifik (b). Demam dengue dengan demam di tambah 2 gejala sepesifik yaitu pendarahan dan tanpa pendarahan (c). Demam Berdarah Dengue dengan atau tanpa *shock syndrome*. Variabel yang digunakan dalam penelitian:

a. Jumlah Penderita Penyakit DBD

Demam berdarah dengue (DBD) adalah penyakit yang disebabkan oleh virus dengue dan ditularkan melalui gigitan nyamuk *genus Aedes*, terutama *Aedes Aegypti* atau *Aedes albopictus*. Penyakit ini sebagian besar penyakit ini meyerang anak berumur kurang dari 15 tahun, namun dapat juga menyerang seluruh kelompok umur.

b. Kepadatan penduduk

Kepadatan penduduk adalah perbandingan jumlah penduduk dengan luas wilayahnya. Kepadatan penduduk menunjukkan rasio banyaknya penduduk per kilometer persegi. Kepadatan penduduk dipengaruhi oleh beberapa faktor antara lain: fisiografis, keamanan, pertumbuhan penduduk, biologis, psikologis dan lain-lain.

c. Tempat tinggal yang tidak memenuhi syarat kesehatan

Rumah merupakan salah satu kebutuhan manusia yang berfungsi sebagai tempat tinggal atau hunian dan sebagai sarana pembinaan keluarga. Rumah harusnya sehat dan nyaman agar penghuninya dapat berkarya untuk meningkatkan produktivitas. Kontruksi rumah dan lingkungan yang tidak memenuhi syarat kesehatan merupakan faktor resiko penularan berbagai penyakit khususnya penyakit berbasis lingkungan seperti Demam Berdarah Dengue, Malaria, Flu Burung, TBC, ISPA dan lain-lain.

d. Penduduk miskin

Penduduk miskin adalah penduduk yang memiliki rata-rata pengeluaran per kapita per bulan di bawah garis kemiskinan. Garis kemiskinan dapat dibedakan menjadi garis kemiskinan makanan dan garis kemiskinan non-makanan.

i) Garis Kemiskinan Makanan (GKM)

Garis kemiskinan makanan (GKM) adalah nilai pengeluaran kebutuhan minimum makanan yang disetarakan dengan 2.100 kalori per kapita per hari.

ii) Garis Kemiskinan Non-Makanan (GKMNM)

Garis kemiskinan non-makanan adalah kebutuhan minimum untuk perumahan, sandang, pendidikan, kesehatan, dan kebutuhan dasar lainnya.

e. Perilaku hidup bersih dan sehat (PHBS)

Perilaku hidup bersih dan sehat merupakan upaya terhadap rumah tangga agar memberdayakan anggota keluarga agar sadar, mau dan mampu melakukan PHBS dalam memelihara dan meningkatkan kesehatan, mencegah terjadinya resiko penyakit dan melindungi diri dari ancaman penyakit serta berperan aktif dalam kesehatan masyarakat.

f. Penduduk usia kurang dari 15 tahun

Penduduk adalah sekelompok manusia yang mendiami suatu wilayah suatu wilayah tertentu pada waktu tertentu yang memiliki hak hukum untuk

tinggal di wilayah tersebut. Penduduk usia kurang dari 15 tahun merupakan penduduk yang masih tergolong dalam usia anak-anak.

g. Rumah sakit umum

Dalam upaya meningkatkan derajat kesehatan masyarakat sehingga diperlukan upaya kuratif dan rehabilitatif selain upaya promotif dan preventif. Upaya kuratif dan rehabilitatif dapat diperoleh dari rumah sakit yang berfungsi sebagai penyedia pelayanan kesehatan. Undang-Undang No. 44 Tahun 2009 tentang Rumah Sakit mengelompokkan rumah sakit berdasarkan jenis pelayanan yang diberikan dibedakan menjadi rumah sakit umum dan rumah sakit khusus. Rumah sakit umum adalah rumah sakit yang memberikan pelayanan kesehatan pada semua bidang dan jenis penyakit.

h. Dokter

Dokter adalah orang yang karena keilmuannya dapat menyembuhkan orang yang sakit. Tidak semua yang menyembuhkan penyakit dapat dikatakan dokter. Untuk menjadi dokter biasanya diperlukan pendidikan dan pelatihan khusus serta mempunyai gelar dalam bidang kedokteran.

i. Balita gizi buruk

Gizi buruk merupakan salah satu klasifikasi status gizi dimana mengalami kurang gizi yang diketahui berdasarkan antropometri seperti penambahan berat badan, tinggi badan/panjang badan, lingkar kepala, lingkar lengan dan lain-lain.