

BAB II

LANDASAN TEORI

2.1 Indikator Kemiskinan

Kemiskinan merupakan masalah dalam pembangunan yang bersifat multidimensi, dimana ditandai oleh pengangguran dan keterbelakangan yang selanjutnya meningkat menjadi pemicu ketimpangan pendapatan dan kesenjangan antargolongan penduduk (Mahendra, 2019). Menurut Ginandjar Kartasasmita (dalam Hermawan & Yulianti, 2021), bahwa merinci indikator-indikator kemiskinan diantaranya adalah:

1. Rendahnya derajat kesehatan

Taraf kesehatan dan gizi yang rendah menyebabkan rendahnya daya tahan fisik, daya pikir, dan prakarsa. Rendahnya derajat kesehatan dapat ditinjau dari Angka Harapan Hidup (AHH) dan keluhan kesehatan.

2. Kondisi keterisolasian

Banyaknya penduduk miskin, secara ekonomi tidak berdaya karena terpencil dan terisolasi sehingga sulit atau tidak dapat terjangkau oleh pelayanan pendidikan, kesehatan, dan gerak kemajuan yang dinikmati masyarakat lainnya. Kondisi tersebut bisa ditinjau dari garis kemiskinan, Produk Domestik Regional Bruto (PDRB), dan laju pertumbuhan penduduk.

3. Rendahnya taraf pendidikan

Taraf pendidikan yang rendah mengakibatkan kemampuan pengembangan diri terbatas dan menyebabkan sempitnya lapangan kerja yang dapat dimasuki. Dalam bersaing untuk mendapatkan lapangan kerja yang ada, taraf pendidikan menjadi faktor yang sangat menentukan. Taraf pendidikan yang rendah ini bisa ditinjau dari Harapan Lama Sekolah (HLS) dan Indeks Pembangunan Manusia (IPM).

4. Terbatasnya lapangan kerja

Keadaan kemiskinan karena kondisi pendidikan dan kesehatan diperberat oleh terbatasnya lapangan pekerjaan. Selama ada lapangan pekerjaan atau kegiatan usaha, selama itu pula ada harapan untuk memutuskan lingkaran kemiskinan

itu. Terbatasnya lapangan pekerjaan dapat ditinjau dari Tingkat Partisipasi Angkatan Kerja (TPAK), Tingkat Pengangguran Terbuka (TPT), dan Upah Minimum Kabupaten/Kota (UMK).

2.2 Analisis Cluster

Analisis *cluster* merupakan teknik multivariat yang digunakan untuk mengelompokkan objek/kasus (responden) menjadi beberapa kelompok dimana setiap kelompok berisi objek/kasus yang mirip satu sama lain (Asiska, Satyahadewi, & Perdana, 2019). Analisis ini diawali dengan pemahaman bahwa sejumlah data tertentu sebenarnya mempunyai kemiripan di antara anggotanya. Karena itu, dimungkinkan untuk mengelompokkan anggota-anggota yang mempunyai karakteristik serupa tersebut dalam satu atau lebih dari satu kelompok/*cluster* (Alfira, Hermin, & Wiraningsih, 2017).

Talakua, Leleury, dan Taluta (2017) menyatakan bahwa ciri-ciri sebuah *cluster* yang baik adalah yang mempunyai:

- 1) Homogenitas (kesamaan) yang tinggi antaranggota dalam satu *cluster* (*within cluster*).
- 2) Heterogenitas (perbedaan) yang tinggi antara *cluster* yang satu dengan *cluster* lainnya (*between-cluster*).

Dari dua hal diatas dapat disimpulkan bahwa *cluster* yang baik adalah sebuah *cluster* yang mempunyai tingkat kesamaan karakteristik yang tinggi antara objek satu dengan objek yang lainnya, namun sangat tidak mirip dengan *cluster* lainnya.

Sebelum proses *clustering*, dapat dilakukan standardisasi data jika diperlukan. Hal ini perlu diperhatikan apabila satuan data mempunyai perbedaan yang besar. Sebagai contoh, jika variabel garis kemiskinan mempunyai satuan ratusan ribu, sedangkan angka harapan hidup seseorang hanya mempunyai satuan puluhan, maka perbedaan yang mencolok ini akan membuat perhitungan jarak menjadi tidak valid. Jika ada perbedaan yang cukup signifikan, maka data harus di standardisasi dengan mengubah data yang ada ke *Z-Score*. Proses standardisasi menjadikan dua data dengan perbedaan satuan yang lebar akan otomatis menjadi menyempit

(Wicaksono, 2017). Cara menentukan nilai standardisasi adalah dengan menggunakan persamaan berikut (Lestari, Hayati, & Amijaya, 2020).

$$Z_{i,j} = \frac{x_{ij} - \mu_j}{\sigma_j} \quad (2.1)$$

Dimana:

$Z_{i,j}$: standardisasi untuk data ke- i variabel ke- j

x_{ij} : data dari objek ke- i pada variabel ke- j

μ_j : rata-rata variabel ke- j

σ_j : standar deviasi variabel ke- j

Setelah data yang dianggap mempunyai satuan yang sangat berbeda diseragamkan, maka langkah selanjutnya adalah menentukan ukuran jarak. Ada 3 metode yang dapat dipilih, untuk menentukan ukuran jarak antardata.

1. Mengukur korelasi antara sepasang objek pada beberapa variabel. Cara ini sebenarnya sederhana, jika beberapa objek memang akan tergabung menjadi satu *cluster*, tentulah diantara objek tersebut ada hubungan yang erat, atau disebut berkorelasi satu dengan yang lain. Metode ini berdasarkan besaran korelasi antar objek untuk mengetahui kemiripan objek satu dengan yang lain.
2. Mengukur asosiasi antar objek. Pada dasarnya, cara ini akan mengasosiasikan sebuah objek dengan *cluster* tertentu. Namun pada pengaplikasiannya metode ini jarang dipakai.
3. Mengukur jarak antar dua objek. Pengukuran ada bermacam-macam, yang paling populer adalah metode jarak *euclidean*. Pada dasarnya, cara ini akan memasukkan sebuah objek ke dalam *cluster* tertentu yang mengukur jarak objek tersebut dengan pusat *cluster*. Jika objek ada dalam jarak yang masih ada dalam batas tertentu, objek tersebut dapat dimasukkan dalam *cluster* tersebut. Berikut adalah persamaan jarak *euclidean* (Wicaksono, 2017).

$$d_{hi} = \sqrt{\sum_{j=1}^c (x_{hj} - x_{ij})^2} \quad (2.2)$$

Dimana:

d_{hi} : jarak antara objek ke- h dan objek ke- i

c : banyaknya variabel

x_{hj} : data dari objek ke- h pada variabel ke- j

x_{ij} : data dari objek ke- i pada variabel ke- j

Setelah cara menentukan jarak ditetapkan, selanjutnya yaitu membuat *cluster*.

Ada dua metode dalam pengelompokan data yaitu:

1. *Hierarchical Method*

Metode ini memulai pengelompokan dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga *cluster* akan membentuk semacam ‘pohon’ dimana ada hierarki (tingkatan) yang jelas antar objek, dari yang paling mirip karakteristiknya sampai yang paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah *cluster*. Dendogram biasanya digunakan untuk membantu memperjelas proses hierarki tersebut (Wicaksono, 2017). Ada 3 metode dalam analisis *hierarchical clustering* yang digunakan dalam penelitian ini, diantaranya adalah sebagai berikut.

a. *Single Linkage*

Untuk menentukan jarak antar *cluster* dengan menggunakan *single linkage*, maka dipilih jarak yang paling dekat atau aturan tetangga dekat (*nearest neighbour rule*). Menghitung jarak *cluster* dengan *cluster-cluster* yang lain dapat dirumuskan dengan (Yusfar, Tiro, & Sudarmin, 2021):

$$d_{(UV)W} = \min(d_{UW}, d_{VW}) \quad (2.3)$$

keterangan:

$d_{(UV)W}$: jarak antara *cluster* (UV) dan *cluster* W

$\min(d_{UW}, d_{VW})$: jarak tetangga terdekat antara *cluster* U dan W atau antara *cluster* V dan W .

b. *Complete Linkage*

Pada metode *complete linkage*, jarak antar *cluster* ditentukan oleh jarak terjauh (*furthest-neighbour*) antara dua objek pada *cluster* yang berbeda. Dapat dirumuskan dengan (Yusfar, Tiro, & Sudarmin, 2021):

$$d_{(UV)W} = \max(d_{UW}, d_{VW}) \quad (2.4)$$

keterangan:

$\max(d_{UW}, d_{VW})$: jarak paling jauh antara *cluster U* dan *W* atau antara *cluster V* dan *W*.

c. *Average Linkage*

Metode ini akan mengelompokkan objek berdasarkan jarak antara dua *cluster* dianggap sebagai jarak rata-rata antara semua anggota dalam satu *cluster* dengan semua anggota *cluster* lain. Pada berbagai keadaan, metode ini dianggap lebih stabil dibandingkan dengan kedua metode sebelumnya, dimana dapat dirumuskan dengan (Asiska, Satyahadewi, & Perdana, 2019):

$$d_{(UV)W} = \text{average}(d_{UW}, d_{VW}) \quad (2.5)$$

keterangan:

$\text{average}(d_{UW}, d_{VW})$: rata-rata antara *cluster U* dan *W* dengan *cluster V* dan *W*.

2. *Non-Hierarchical Method*

Berbeda dengan metode hierarki, metode ini justru dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan. Setelah jumlah *cluster* diketahui, baru proses *cluster* dilakukan tanpa mengikuti proses hierarki (Wicaksono, 2017). Salah satu metodenya adalah *K-means cluster*.

Setelah sejumlah *cluster* terbentuk, maka langkah selanjutnya adalah melakukan penamaan/interpretasi terhadap *cluster-cluster* yang telah terbentuk, yaitu memberi nama spesifik untuk menggambarkan isi *cluster* tersebut (Wicaksono, 2017).

2.3 *Calinski-Harabasz Pseudo-F Statistic*

Pseudo-F statistic adalah salah satu metode yang umum digunakan untuk menentukan banyaknya kelompok yang optimum. Penelitian oleh Milligan dan Cooper tahun 1985 menunjukkan bahwa *Pseudo-F statistic* yang selanjutnya disebut *Pseudo-F*, memberikan hasil terbaik diantara 30 metode dan merupakan metode yang dapat digunakan secara global. *Pseudo-F* dirumuskan oleh Calinski dan Harabasz yang ditulis dalam persamaan berikut (Mauludina & Setiawan, 2012).

$$Pseudo F = \frac{\frac{R^2}{p-1}}{\frac{1-R^2}{N-p}} \quad (2.6)$$

$$R^2 = \frac{SST-SSW}{SST} \quad (2.7)$$

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_j)^2 \quad (2.8)$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2 \quad (2.9)$$

Keterangan:

SST (<i>Sum Square Total</i>)	:	total jumlah dari kuadrat jarak sampel terhadap rata-rata keseluruhan
SSW (<i>Sum Square Within</i>)	:	total jumlah dari kuadrat jarak sampel terhadap rata-rata <i>cluster</i> -nya
n	:	banyaknya sampel pada setiap <i>cluster</i>
c	:	banyaknya variabel
p	:	banyaknya <i>cluster</i>
x_{ijk}	:	sampel ke- i pada variabel ke- j <i>cluster</i> ke- k
\bar{x}_j	:	rata-rata seluruh sampel pada variabel ke- j
\bar{x}_{jk}	:	rata-rata sampel pada variabel ke- j dan <i>cluster</i> ke- k
N	:	banyaknya sampel

Nilai *Pseudo-F statistic* tertinggi menunjukkan bahwa jumlah *cluster* yang digunakan untuk mempartisi data telah optimum.