

BAB II

LANDASAN TEORI

2.1 Analisis Cluster

Analisis *cluster* adalah metode yang tepat untuk mengidentifikasi objek- objek yang homogen ke dalam kelompok-kelompok yang disebut *cluster*. *Cluster-cluster* yang terbentuk memiliki homogenitas internal yang tinggi dan heterogenitas eksternal yang tinggi (Hair, dkk., 2010:486). Secara garis besar metode *clustering* dibagi menjadi dua, yaitu: metode *clustering* hirarki dan metode *clustering* nonhirarki. Metode *clustering* hirarki dibagi menjadi dua, yaitu *agglomerative* (penyatuan) dan metode *divisive* (pembagian). Dalam metode *agglomerative*, proses pengelompokan dimulai dengan objek-objek yang individual. Jadi, banyaknya *cluster* sama dengan banyaknya objek. Objek-objek yang paling mirip pertama kali bergabung membentuk *cluster*, demikian seterusnya sampai membentuk satu *cluster*. Sedangkan metode *divisive* merupakan kebalikan dari metode *agglomerative*. Metode ini dimulai dari satu *cluster* yang mencakup semua objek pengamatan. Kemudian objek yang memiliki ketidakmiripan cukup besar akan dipisahkan menjadi kelompok baru, demikian seterusnya sampai terbentuk *cluster* yang jumlahnya sama dengan jumlah objek yang dikelompokkan (Johnson & Wichern, 2007:680).

Secara umum ada dua metode dalam analisis *cluster* yaitu:

2.1.1 Metode Tak Hirarki

Metode tak hirarki

merupakan metode pengelompokan dimana klaster yang ingin dibentuk ditentukan terlebih dahulu, sehingga objek-objek akan dikelompokkan kedalam k kelompok yang telah ditentukan.

2.1.2 Metode Hirarki

Metode hirarki dapat digunakan jika jumlah kelompok yang diinginkan tidak diketahui, dan metode ini biasanya digunakan dalam objek dengan pengamatan yang tidak besar (Simamora, 2005).

Metode perbaikan jarak yang dapat digunakan pada metode hirarki adalah

(Supranto, 2004):

1. Metode pautan tunggal (*Single Linkage*)

Cluster dibentuk dari individu objek dengan jelas menggabungkan jarak terdekat. Pada setiap tahap, setelah terbentuk *cluster* baru (*UV*), maka jarak antara (*UV*) dan *cluster* lainnya, misal *W* adalah:

$$d_{(uv)w} = \min \{d_{uw}, d_{vw}\} \quad (2.1)$$

Keterangan:

d_{uw} = jarak antara *cluster* *U* dan *W*

d_{vw} = jarak antara *cluster* *V* dan *W*

$d_{(uv)w}$ = jarak antara *cluster* (*UV*) dan *W*

2. Metode pautan lengkap (*Complete Linkage*)

Metode pautan lengkap dilakukan sama seperti metode pautan tunggal, dengan suatu pengecualian yaitu, pada setiap tahap jarak antara *cluster* ditentukan oleh jarak antara dua objek, satu dari setiap *cluster* yang paling jauh. Setelah *cluster* *U* dan *V* digabung menjadi *cluster* (*UV*) jarak antara *cluster* (*UV*) dan *cluster* lain, misal *W* adalah:

$$d_{(uv)w} = \max \{d_{uw}, d_{vw}\} \quad (2.2)$$

Keterangan:

d_{uw} = jarak antara *cluster* U dan W

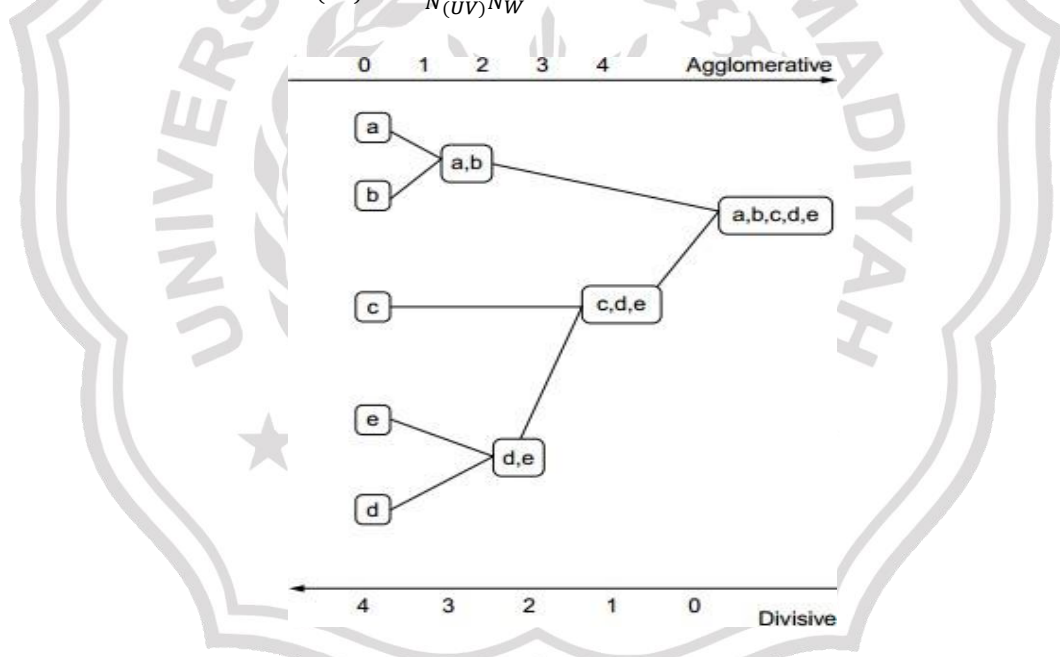
d_{vw} = jarak antara *cluster* V dan W

$d_{(uv)w}$ = jarak antara *cluster* (UV) dan W

3. Metode pautan rata-rata (*Average Linkage*)

Metode pautan rata-rata dilakukan hampir sama dengan metode sebelumnya, perbedaannya yaitu, Dengan d_{ik} (Johnson dan Wichern, 1996:594) merupakan jarak antara obyek i dalam *cluster* (UV) dan obyek k dalam *cluster* W . Sedangkan $N_{(UV)}$ dan N_w berturut-turut merupakan jumlah obyek dalam *cluster* (UV) dan (W)

$$d_{(uv)w} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} N_w} \quad (2.3)$$



Gambar 2.1 Contoh Struktur Pohon Metode *Clustering* Hirarki (Sumber Gambar: Kauffman & Rousseeuw (1970) dalam Everitt, dkk. (2011:72))

2.2 Vektor Mean dan Matriks Varian-Kovarian Data Multivariat

Data multivariat adalah data yang diperoleh dengan mengukur lebih dari dua variabel kriteria pada setiap individu anggota sampel. Data multivariat dinotasikan dengan X_{ij} , yang menunjukkan nilai tertentu yang diamati pada objek ke- i dari

variabel ke- j , dengan $i = 1, 2, 3, \dots, n$ dan $j = 1, 2, 3, \dots, p$. Data multivariat dapat diilustrasikan sebagai berikut :

Tabel 2.1 Data Multivariat

Pengamatan	Var. 1	Var. 2	...	Var. j	...	Var. p
Objek 1	X_{11}	X_{12}	...	X_{1j}	...	X_{1p}
Objek 2	X_{21}	X_{22}	...	X_{2j}	...	X_{2p}
⋮	⋮	⋮		⋮		⋮
Objek i	X_{i1}	X_{i2}	...	X_{ij}	...	X_{ip}
⋮	⋮	⋮		⋮		⋮
Objek n	X_{n1}	X_{n2}	...	X_{nj}	...	X_{np}

atau dapat juga ditulis dalam bentuk matriks X dengan n baris dan p kolom

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1j} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2j} & \cdots & X_{2p} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{i1} & X_{i2} & \cdots & X_{ij} & \cdots & X_{ip} \\ \vdots & \vdots & & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nj} & \cdots & X_{np} \end{bmatrix} \quad (2.4)$$

Secara umum mean sampel untuk n objek dan p variabel adalah

$$\bar{X}_j = \frac{1}{n} \sum_{i=1}^n X_{ij}; j = 1, 2, \dots, p \quad (2.5)$$

Sehingga vektor mean sampel $\bar{X} = \begin{bmatrix} \bar{X}_1 \\ \bar{X}_2 \\ \bar{X}_3 \\ \bar{X}_4 \\ \vdots \\ \bar{X}_p \end{bmatrix}$ (2.6)

Varian Sampel untuk variabel ke- j adalah

$$S_{jj} = S_j^2 = \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)^2$$

$$= \frac{1}{n-1} (\sum_{i=1}^n (X_{ij} - \bar{X}_j)^2); j = 1, 2, \dots, p \quad (2.7)$$

Sedangkan kovarian sampel untuk variabel ke-j dan ke-k adalah

$$\begin{aligned} S_{jk} &= \frac{1}{n-1} \sum_{i=1}^n (X_{ij} - \bar{X}_j)(X_{ik} - \bar{X}_k) \\ &= \frac{1}{n-1} \sum_{i=1}^n X_{ij}X_{ik} - n\bar{X}_j\bar{X}_k \end{aligned} \quad (2.8)$$

Sehingga matriks varian-kovarian data multivariat p variabel ditulis

$$\Sigma = \begin{bmatrix} S_{1^2} & S_{12} & \cdots & S_{1p} \\ S_{21} & S_{2^2} & \cdots & S_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ S_{p1} & S_{p2} & \cdots & S_{p^2} \end{bmatrix} \quad (2.9)$$

2.3 Uji Validitas Cluster

Setelah mendapatkan hasil/solusi *cluster* dari proses *clustering* data, hal yang penting untuk dilakukan adalah menguji validitas *cluster*. Uji validitas *cluster* diperlukan untuk melihat kebaikan (*goodness*) atau kualitas (*quality*) hasil analisis *cluster*. Salah satu ukuran yang dapat digunakan untuk menguji validitas hasil *clustering* metode hirarki adalah *Sillhouetes Index*. Koefisien korelasi *cophenetic* merupakan koefisien korelasi antara elemen-elemen asli matriks ketidakmiripan (matriks jarak Euclidean) dan elemen-elemen yang dihasilkan oleh dendrogram (matriks *cophenetic*) (Silva & Dias, 2013:589-590).

Sillhoutte Index adalah salah satu metode yang digunakan untuk memvalidasi baik sebuah data, *cluster* tunggal (satu *cluster* dari sejumlah *cluster*), atau bahkan keseluruhan *cluster*. Metode ini paling banyak digunakan untuk memvalidasi *cluster* yang menggabungkan nilai kohesi dan separasi (Eko Prasetyo, 2014). Untuk menghitung SI dari sebuah data ke- i , ada 2 komponen yaitu a_i dan b_i setelah didapatkan nilai a dan b baru bisa dihitung nilai *Sillhoutte Index*-nya. Dengan formula sebagai berikut

Pertama formula untuk mencari nilai a_i^j sebagai berikut:

$$\alpha_i^j = \frac{1}{m_j-1} \sum_{\substack{r=1 \\ r \neq i}}^{m_j} d(x_i^j, x_r^j) \quad (2.10)$$

Keterangan :

α_i^j = rata-rata jarak data ke- i terhadap semua data dalam satu *cluster*

m_j = jumlah data dalam *cluster* ke- j

$d(x_i^j, x_r^j)$ = jarak data ke- i dengan data ke- r dalam satu *cluster* j

Kemudian formula untuk mencari nilai b_i^j sebagai berikut:

$$b_i^j = \min_{\substack{n=1, \dots, k \\ n \neq j}} \left\{ \frac{1}{m_n} \sum_{\substack{r=1 \\ r \neq i}}^{m_n} d(x_i^j, x_r^n) \right\} \quad (2.11)$$

Keterangan :

b_i^j = rata-rata jarak data ke- i terhadap semua data dari *cluster* lain

m_n = jumlah data dalam *cluster* ke- n

$d(x_i^j, x_r^n)$ = jarak data ke- i dengan data ke- r dengan *cluster* ke- n

Untuk mendapatkan nilai SI data ke- i digunakan formula sebagai berikut

$$SI_i^j = \frac{b_i^j - \alpha_i^j}{\max \{ \alpha_i^j, b_i^j \}} \quad (2.12)$$

Keterangan :

SI_i^j = *Sillhouette index* data ke- i

α_i^j = rata-rata jarak data ke- i terhadap semua data dalam satu *cluster*

b_i^j = rata-rata jarak data ke- i terhadap semua data dari *cluster* lain yang tak

Untuk mendapatkan nilai SI dari sebuah *cluster* didapatkan dengan menghitung nilai rata-rata dari sebuah *cluster* menggunakan formula sebagai berikut

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_i^j \quad (2.13)$$

Keterangan :

SI_j = *Sillhoute index cluster ke-j*

m_j = Jumlah data dalam *cluster ke-j*

Sementara nilai SI Global didapatkan dengan menghitung rata-rata nilai SI dari semua *cluster* menggunakan persamaan berikut

$$SI = \frac{1}{k} \sum_{j=1}^k SI_j \quad (2.14)$$

Keterangan :

SI = *Sillhoute index global*

k = Jumlah *cluster*

(Eko Prasetyo, 2014)

2.4 Pembangunan Manusia

Manusia adalah kekayaan bangsa yang sesungguhnya. Tujuan utama dari pembangunan adalah menciptakan lingkungan yang memungkinkan bagi rakyatnya untuk menikmati umur panjang, sehat, dan menjalankan kehidupan produktif. Hal ini tampaknya merupakan suatu kekayaan yang sederhana. Tetapi hal ini seringkali terlupakan oleh berbagai kesibukan jangka pendek untuk mengumpulkan harta dan uang. Pada tahun 1990 *United Nations Development Program* (UNDP) dalam laporannya “*Global Human Development Report*” memperkenalkan konsep “Pembangunan Manusia (*Human Development*)” sebagai paradigma baru model pembangunan. Menurut UNDP, pembangunan manusia dirumuskan sebagai perluasan pilihan bagi penduduk (*enlarging the choices of people*), yang dapat dilihat sebagai proses upaya ke arah “perluasan pilihan” dan sekaligus sebagai taraf yang dicapai dari upaya tersebut. Pada saat yang sama pembangunan manusia dapat

dilihat juga sebagai pembangunan (*formation*) kemampuan manusia melalui perbaikan taraf kesehatan, pengetahuan, dan keterampilan, sekaligus sebagai pemanfaatan (*utilization*) kemampuan/ketrampilan mereka. Konsep pembangunan tersebut jauh lebih luas pengertiannya dibandingkan konsep pembangunan ekonomi yang menekankan pada pertumbuhan (*economic growth*), kebutuhan dasar, masyarakat, atau pengembangan sumber daya manusia.

Menurut UNDP (1995), untuk memperluas pilihan-pilihan manusia, konsep pembangunan manusia harus dibangun dari empat dimensi yang tidak terpisahkan. Berdasarkan konsep di atas maka untuk menjamin tercapainya tujuan pembangunan manusia, ada empat unsur pokok yang perlu diperhatikan yaitu:

a. Produktivitas (*Productivity*)

Masyarakat harus mampu untuk meningkatkan produktifitas mereka dan berpartisipasi penuh dalam proses mencari penghasilan dan lapangan pekerjaan. Oleh karena itu, pembangunan ekonomi merupakan bagian dari model pembangunan manusia.

b. Pemerataan (*equality*)

Masyarakat harus mempunyai akses untuk memperoleh kesempatan yang adil. Semua hambatan terhadap peluang ekonomi dan politik harus dihapuskan sehingga masyarakat dapat berpartisipasi di dalam dan memperoleh manfaat dari peluang-peluang yang ada.

c. Kestinambungan (*Sustainability*)

Akses untuk memperoleh kesempatan harus dipastikan bahwa tidak hanya untuk generasi sekarang tetapi juga untuk generasi yang akan datang. Semua jenis pemodalannya baik itu fisik, manusia, dan lingkungan hidup harus dilengkapi.

d. Pemberdayaan (*Empowerment*)

Pembangunan harus dilakukan oleh masyarakat, dan bukan hanya untuk mereka. Masyarakat harus berpartisipasi penuh dalam mengambil keputusan dan proses-proses yang mempengaruhi kehidupan mereka.

2.5 Indeks Pembangunan Manusia

Indeks Pembangunan Manusia (IPM) merupakan suatu indikator yang menjelaskan bagaimana penduduk suatu wilayah mempunyai kesempatan untuk mengakses hasil dari suatu pembangunan sebagai bagian dari haknya dalam memperoleh pendapatan, kesehatan, pendidikan, dan sebagainya. Nilai IPM ini menunjukkan seberapa jauh wilayah tersebut telah mencapai sasaran yang ditentukan yaitu umur harapan hidup 85 tahun, pendidikan dasar bagi semua lapisan masyarakat (tanpa kecuali), dan tingkat pengeluaran dan konsumsi yang telah mencapai standar hidup layak.

IPM merupakan indeks komposit yang dihitung sebagai rata-rata sederhana dari tiga indeks yang terdiri dari indeks harapan hidup yang diukur dengan harapan hidup pada saat lahir, indeks pendidikan yang diukur dengan kombinasi antara angka melek huruf pada penduduk dewasa dan rata-rata lama sekolah, serta indeks standar hidup layak yang diukur dengan pengeluaran perkapita yang telah disesuaikan atau paritas daya beli. Dalam konteks pembangunan daerah Indeks Pembangunan Manusia (IPM) ditetapkan sebagai salah satu ukuran utama yang dicantumkan dalam pola dasar pembangunan daerah. Hal ini menandakan bahwa IPM menduduki satu posisi penting dalam manajemen pembangunan daerah.

Fungsi IPM dan indikator pembangunan manusia lainnya akan menjadi kunci bagi terlaksananya perencanaan dan pembangunan yang terarah. Peran IPM sebagai alat ukur pembangunan akan lebih terlihat bila dilengkapi dengan data basis dan hitungan yang benar sampai ke wilayah terkecil tanpa membedakan daerah miskin atau tidak sehingga diharapkan perencanaan pembangunan akan benar-benar memihak masyarakat tanpa terkecuali.

2.6 Dimensi Indeks Pembangunan Manusia

Menurut Suparman terdapat 3 dimensi IPM, dan secara rinci ketiga dimensi tersebut dapat dijelaskan sebagai berikut:

1. Umur Harapan Hidup adalah indikator yang mengukur *longevity* (panjang umur) dari seseorang di suatu wilayah atau negara. *Longevity* ini bukan hanya upaya perorangan tetapi merupakan upaya masyarakat secara keseluruhan untuk menggunakan sumber daya yang ada sehingga dapat memperpanjang hidupnya. Dapat dikatakan seseorang akan bertahan hidup lebih panjang apabila selalu sehat, atau jika menderita sakit secepatnya dapat berobat untuk membantu mempercepat kesembuhannya.
2. Harapan Lama Sekolah dan Rata Lama Sekolah adalah HLS merupakan salah satu output yang dapat digunakan untuk memotret pemerataan pembangunan pendidikan di Indonesia. Karena HLS mengukur kesempatan pendidikan seorang penduduk di mulai pada usia tujuh tahun. Secara sederhana, HLS dapat didefinisikan sebagai angka partisipasi sekolah menurut umur tunggal. HLS merupakan indikator yang menggambarkan lamanya sekolah (dalam tahun) yang diharapkan akan dirasakan oleh anak pada umur tertentu di masa mendatang. Angka ini diperoleh dengan cara membagi banyaknya partisipasi sekolah penduduk pada usia a pada tahun t dengan jumlah penduduk yang bersekolah pada usia a pada tahun t .
3. Paritas Daya Beli adalah indikator yang mengukur tentang besarnya daya beli masyarakat di suatu wilayah atau negara. Dengan menggunakan indikator konsumsi riil yang disesuaikan.

Kualitas pembangunan manusia yang telah dicapai oleh suatu wilayah dapat dilakukan dengan mengukur mutu pembangunan tersebut dengan menggunakan parameter dengan 3 (tiga) komponen antara lain; (1) Keberhasilan dalam kesehatannya yaitu dilihat dari kemampuan hidup secara fisik yaitu dengan melihat umur harapan hidup; (2) Kemampuan untuk merefleksikan keberhasilan

pengembangan pendidikan dengan melihat harapan lama sekolah dan rata lama sekolah; (3) Besarnya barang dan jasa yang dapat disediakan oleh masyarakat bagi warganya yaitu dengan melihat paritas daya beli masyarakat. Dengan kata lain indeks pembangunan manusia diukur dengan tiga dimensi, yaitu 1) indek kesehatan, 2) pendidikan dan 3) ekonomi. Indek kesehatan diukur dari umur harapan hidup, biasanya umur harapan hidup bayi yang lahir. Indek pendidikan salah satunya dapat diukur dari angka melek huruf. Kemudian dimensi ekonomi diukur dari indek daya beli masyarakat.

Dalam melakukan pembangunan manusia tentunya tidak hanya memperhatikan pada tiga komponen yang menjadi tolok ukur dalam penentuan Indeks Pembangunan Manusia (IPM) saja, disamping hal-hal tersebut tentunya masih banyak aspek lain yang juga mempengaruhi yaitu dari berbagai aspek pembangunan manusia yang tidak dapat diukur seperti; moral, mental, spiritual, tanggung jawab dan lain sebagainya. Untuk itu dalam upaya mencapai kesejahteraan secara utuh kita jangan hanya terjebak untuk memprioritaskan pada peningkatan di tiga aspek yang menjadi tolok ukur IPM saja, tetapi juga perlu meningkatkan kualitas manusia dari aspek-aspek yang lainnya.

2.7 Penelitian Terdahulu

Sebagai bahan pertimbangan dalam penelitian tentang klasifikasi tentang metode topos, maka penulis akan cantumkan beberapa hasil penelitian terdahulu oleh beberapa peneliti yang pernah penulis baca:

Pertama, penelitian sebelumnya dari Lisna Zahrotun (2015) yang berjudul “Analisis Pengelompokan Jumlah Penumpang Bus Trans Jogja Menggunakan Metode *Clustering K-Means* dan *Agglomerative Hierarchical Clustering (AHC)*”. Pada penelitian ini permasalahan timbul karena adanya peningkatan pengguna bus tapi tak ada peningkatan jumlah halte, sehingga banyak penumpang yang tidak terangkut bus trans jogja. Disisi lain juga terdapat halte-halte yang hanya terdapat sedikit calon penumpang pada jam tertentu, sehingga banyak diantaranya yang ingin menggunakan jasa bus Trans Jogja, tetapi tidak terpenuhi karena tidak adanya *shelter*. Maka dibuatlah penelitian tersebut yang diharapkan dapat membantu

jumlah halte yang tepat pada tiap daerah yang dapat memenuhi kebutuhan pengguna jasa bus.

Penelitian ini menggunakan 3 variabel untuk menghitung kemiripan yaitu jam kedatangan bus, jalur bus, dan jumlah penumpang. Dari perhitungan yang menggunakan 113 data diperoleh hasil Sebaran dari data terhadap ketiga cluster menghasilkan jumlah data pada masing-masing cluster sebagai berikut: cluster 0 = 2 data, cluster 1 = 57 data, dan cluster 2 = 54 data. (Lisna Zahrotun, 2015)

Kedua, penelitian sebelumnya dari Gangga Anuraga (2015) dengan judul “*Hierarchical Clustering Multiscale Bootstrap* Untuk Pengelompokan Kemiskinan Di Jawa Timur”. Pada Penelitian ini variabel yang digunakan sejumlah 16 variabel yang dapat digolongkan menjadi 3 kelompok yaitu ekonomi, SDM, dan kesehatan. Berdasarkan hasil perhitungan yang menggunakan 38 data yang terdiri dari 29 kabupaten dan 9 kota, terbentuk 5 cluster yang signifikansi nilai AU (*P-values*) yang mendekati 0.95 dengan nilai standar *error bootstrap* yang kecil. Terdapat lima kelompok (*cluster*) yang terbentuk dengan kemiripan satu dengan yang lain dan signifikan dengan nilai AU (*Approximately Unbiased*) $p_value \geq 0,95$. Dan kelompok ke-5 yang terdiri dari Kabupaten Jember, Kabupaten Bondowoso, Kabupaten Situbondo, Kabupaten Probolinggo, Kabupaten Pasuruan dapat dikatakan sebagai daerah yang masih cukup tinggi persentase kemiskinannya dibandingkan dengan kelompok 1, 2, 3, dan 4. (Gangga Anuraga, 2015)