

## **BAB II KAJIAN TEORI**

### **2.1 Kemiskinan**

Kemiskinan secara etimologis berasal dari kata “miskin” yang artinya tidak berharta benda dan serba kekurangan. Menurut Smeru (2016:2) kemiskinan adalah kondisi keterbatasan kemampuan untuk memenuhi kebutuhan hidup secara layak seperti keterbatasan dalam pendapatan, keterampilan, kondisi kesehatan, penguasaan aset ekonomi, ataupun akses informasi. Menurut BPS (2017:58) kemiskinan adalah ketidakmampuan dari sisi ekonomi untuk memenuhi kebutuhan dasar makanan dan bukan makanan yang diukur dari sisi pengeluaran perkapita perbulan. Secara garis besar kemiskinan merupakan suatu kondisi dimana seseorang tidak mampu untuk memenuhi kebutuhan dasarnya seperti, sandang, pangan, tempat tinggal, pendidikan, dan kesehatan yang baik.

BPS memiliki konsep kemampuan memenuhi kebutuhan dasar (*basic needs approach*) dalam mengukur kemiskinan. Dengan pendekatan ini, kemiskinan dipandang sebagai ketidakmampuan ekonomi masyarakat untuk memenuhi kebutuhan dasar pangan dan non-pangan yang diukur dari sisi pengeluaran. Sedangkan untuk kategori penduduk miskin menurut BPS adalah penduduk yang memiliki rata-rata pengeluaran perkapita perbulan di bawah garis kemiskinan (GK). GK didapatkan dari hasil survey konsumsi Susenas yang ditetapkan dalam rupiah perorang perbulan. GK ditetapkan setiap tiga tahun sekali baik untuk tingkat nasional maupun tingkat provinsi. Beberapa konsep GK yang digunakan BPS sebagai berikut:

- a. Garis kemiskinan merupakan penjumlahan dari garis kemiskinan makanan (GKM) dan garis kemiskinan non-makanan (GKNM). Penduduk yang memiliki rata-rata pengeluaran perkapita per bulan di bawah GK dikategorikan sebagai penduduk miskin.
- b. GKM merupakan nilai pengeluaran kebutuhan minimum makanan yang disetarakan dengan 2.100 kilokalori perkapita perhari. Paket komoditi kebutuhan dasar makanan diwakili oleh 52 jenis komoditi (padi-padian, umbi-umbian, ikan, daging, telur dan susu, sayuran, kacang-kacangan, buah-buahan, minyak dan lemak, dan lain-lain).
- c. GKNM adalah kebutuhan minimum untuk perumahan, sandang, pendidikan dan kesehatan. Paket komoditi kebutuhan dasar non-makanan diwakili oleh 51 jenis komoditi di perkotaan dan 47 jenis komoditi di pedesaan.

Metode yang digunakan untuk menghitung GK yang terdiri dari dua komponen yaitu GKM dan GKNM secara matematis yaitu:

$$GK = GKM + GKNM \quad (1)$$

(BPS, 2017:58)

Menurut Foster-Greer-Thorbecke (dalam BPS, 2017:62) merumuskan suatu ukuran yang digunakan untuk mengukur kemiskinan yaitu:

$$P_{\alpha} = \frac{1}{n} \sum_{i=1}^q \left[ \frac{z-y_i}{z} \right]^{\alpha} \quad (2)$$

Keterangan:

$\alpha = 0, 1, 2$

$z =$  garis kemiskinan

$y_i$  = rata-rata pengeluaran perkapita sebulan penduduk yang berada di bawah garis kemiskinan.

$q$  = banyaknya penduduk yang berada di bawah garis kemiskinan

$n$  = jumlah penduduk

## 2.2 Pendidikan

Kemiskinan dan ketidaksetaraan dapat diatasi melalui investasi pada *human capital*. *Human capital is the skills the labor force possesses and is the regarded as a resource or asset (human capital adalah keterampilan yang dimiliki oleh angkatan kerja dan dianggap sebagai sumber daya atau aset) (Goldin, 2016:1).* Sumber daya atau aset yang menjadi salah satu indikatornya yaitu pendidikan yang nantinya akan meningkatkan produktivitas individu.

Rendahnya produktivitas tenaga kerja kaum miskin salah satu tak lain disebabkan oleh karena rendahnya akses mereka untuk memperoleh pendidikan dan kesehatan (Sitepu, 2010). Pendidikan sangat berguna untuk pembangunan sumber daya manusia disuatu negara. Melalui pendidikan akan terjadi suatu proses pengubahan sikap, menumbuhkan, mengembangkan, mendewasakan, menata dan mengarahkan seseorang atau sekelompok orang yang diharapkan menjadi sumber daya manusia yang lebih baik dan mampu bersaing di dunia kerja. Orang dengan pendidikan yang lebih tinggi memiliki peluang untuk mendapatkan pekerjaan yang baik dan gaji yang tinggi. Oleh karena itu, kualitas dari sumber daya manusia sangat bergantung pada kualitas pendidikan yang diterimanya.

Salah satu upaya pemerintah untuk meningkatkan kualitas pendidikan yaitu dengan program wajib belajar selama 9 tahun dengan harapan tingkat pendidikan dari masyarakat akan lebih baik. Untuk melihat keadaan pendidikan di Indonesia BPS mengkaji beberapa indikator antara lain seperti pendidikan yang ditamatkan, angka melek huruf dan angka partisipasi sekolah:

#### 1. Pendidikan yang ditamatkan

Pendidikan yang ditamatkan adalah ijazah/STTB tertinggi yang dimiliki oleh penduduk miskin berumur 15 tahun ke atas. Pada publikasi ini BPS mengelompokkan pendidikan yang ditamatkan menjadi tiga yaitu SD ke bawah (tidak mempunyai ijazah), mempunyai ijazah SD atau SMP dan mempunyai ijazah SMA atau ijazah perguruan tinggi.

#### 2. Angka Melek Huruf

Angka melek huruf adalah proporsi penduduk miskin yang dapat membaca dan menulis kalimat sederhana dalam aksara tertentu, yaitu huruf latin, huruf arab dan lain-lain. Pada publikasi ini BPS menyajikan angka melek huruf penduduk miskin untuk kelompok umur 15-24 tahun dan 15-55 tahun.

#### 3. Angka Partisipasi Sekolah

Angka partisipasi sekolah adalah proporsi dari penduduk miskin yang masih bersekolah. Pada publikasi ini BPS menyajikan angka partisipasi sekolah penduduk miskin pada kelompok umur 7-12 tahun dan 13-15 tahun yang berguna untuk memantau pelaksanaan program wajib belajar 9 tahun diantara penduduk miskin berusia sekolah. Pada penelitian

ini indikator pendidikan yang digunakan sebagai variabel penelitian adalah penduduk yang tidak tamat SD dan penduduk yang dapat membaca pada usia 15-55 tahun.

### **2.3 Ketenagakerjaan**

Ketenagakerjaan merupakan aspek penting dalam pembangunan ekonomi di suatu negara. Penduduk merupakan suplai bagi pasar tenaga kerja. Karena untuk mendapatkan penghasilan yang akan digunakan untuk memenuhi kebutuhan dasar penduduk diharuskan untuk bekerja. Bekerja adalah kegiatan ekonomi yang dilakukan oleh seseorang dengan maksud memperoleh atau membantu memperoleh pendapatan atau keuntungan, paling sedikit 1 jam (tidak terputus) dalam seminggu yang lalu (BPS, 2018). Penduduk usia kerja dibagi menjadi dua golongan yaitu penduduk yang termasuk angkatan kerja dan penduduk yang termasuk bukan angkatan kerja (BPS, 2018):

1. Penduduk yang termasuk angkatan kerja adalah penduduk usia kerja (15 tahun keatas) yang bekerja atau punya pekerjaan namun sementara tidak bekerja dan pengangguran.
2. Penduduk yang termasuk bukan angkatan kerja adalah penduduk usia kerja (15 tahun keatas) yang masih sekolah, mengurus rumah tangga atau melaksanakan kegiatan lainnya selain kegiatan pribadi.

Penduduk miskin yang berada dalam angkatan kerja dibagi menjadi penduduk miskin yang tidak bekerja dan penduduk miskin yang bekerja. Untuk penduduk miskin yang tidak bekerja, dapat dilihat bahwa kemiskinannya terjadi karena ketiadaan penduduk tersebut untuk bekerja sehingga tidak dapat memenuhi

kebutuhan hidupnya. Penduduk miskin yang bekerja dapat dikelompokkan menjadi tiga kategori yaitu kelompok setengah penganggur, pekerja paruh waktu dan kelompok yang bekerja penuh (BPS:2018):

1. Setengah penganggur adalah mereka yang bekerja dibawah jam kerja normal (kurang dari 35 jam seminggu) dan masih mencari pekerjaan atau masih bersedia menerima pekerjaan.
2. Pekerja paruh waktu adalah mereka yang bekerja dibawah jam kerja normal (kurang dari 35 jam seminggu), tetapi tidak mencari pekerjaan atau tidak bersedia menerima pekerjaan lain.
3. Pekerja yang bekerja penuh adalah pekerja dengan jam kerja diatas 35 jam seminggu.

Pada dimensi indikator ketenagakerjaan dapat diketahui berapa besar penduduk yang bekerja, mengetahui jumlah pengangguran dan pencari kerja, jika dilihat dari statusnya dapat diketahui berapa jumlah penduduk yang bekerja di sektor formal dan sektor informal. Untuk melihat masalah ketenagakerjaan BPS mengkaji beberapa indikator antara penduduk miskin yang berusia 15 tahun keatas berdasarkan status bekerja dan sektor berkerja (BPS, 2017:5):

1. Bekerja di sektor informal adalah penduduk miskin yang mempunyai status/kedudukan dalam pekerjaan utamanya adalah berusaha sendiri, berusaha dibantu buruh tidak tetap/buruh tidak dibayar, pekerja bebas, atau pekerja keluarga/tidak dibayar.

2. Bekerja di sektor formal adalah penduduk miskin yang mempunyai status/kedudukan dalam pekerjaan utamanya adalah bekerja dibantu buruh tetap/buruh dibayar atau buruh/karyawan/pegawai.
3. Bekerja di sektor pertanian adalah penduduk miskin yang bekerja di sektor pertanian, seperti pertambangan dan penggalian, industri pengolahan, listrik dan gas, konstruksi/bangunanm perdagangan, hotel dan rumah makan, transportasi, keuangan, jasa atau lainnya.
4. Bekerja di sektor bukan pertanian adalah penduduk miskin yang bekerja selain di sektor pertanian.

Pada penelitian ini indikator pada dimensi ketenagakerjaan yang digunakan sebagai variabel penelitian antara lain: penduduk yang bekerja di sektor non pertanian dan penduduk yang bekerja di sektor pertanian.

#### **2.4 Fasilitas Perumahan**

Rumah merupakan salah satu bangunan yang dijadikan sebagai tempat tinggal selama jangka waktu tertentu. Rumah harus menjamin kepentingan keluarga seperti untuk tumbuh, memberi kemungkinan untuk hidup bergaul dengan tetangganya dan juga rumah harus memberi ketenangan, kesenangan, kebahagiaan dan kenyamanan pada segala peristiwa hidupnya. Karena rumah merupakan suatu hal yang sangat penting maka fasilitas-fasilitas yang ada dalam rumah sangat berpengaruh kepada penghuninya. Untuk penduduk yang dikategorikan miskin secara umum, rumah yang mereka tinggali jauh dari kata layak.

Rumah penduduk miskin kebanyakan berada pada daerah yang kurang bersih seperti di sekitar pantai, kolong jembatan, di sekitar sungai dan lain-lain. Menurut BPS (2017:6) pada dimensi fasilitas perumahan terdapat dua indikator yang dikaji yaitu rumah tangga pengguna air layak dan rumah tangga pengguna jamban sendiri:

1. Rumah tangga pengguna air layak adalah rumah tangga miskin yang menggunakan sumber utama air minum dari air tidak sustain (air hujan), air terlindung maupun tidak terlindung dengan syarat sumber mandi/cuci dan lain-lain yang digunakan berasal dari air terlindung.
2. Rumah tangga pengguna jamban sendiri/bersama adalah rumah tangga yang menggunakan fasilitas tempat pembuangan air besar yang digunakan oleh rumah tangga sendiri atau bersama dengan rumah tangga tertentu.

Pada penelitian ini, pada dimensi fasilitas perumahan yang digunakan sebagai variabel penelitian adalah rumah tangga pengguna air layak dan rumah tangga miskin yang menggunakan jamban sendiri/bersama.

## **2.5 Fertilitas dan Keluarga Berencana**

Fertilitas dalam perspektif demografi merupakan salah satu faktor yang mempengaruhi jumlah penduduk. Fertilitas merupakan kemampuan seorang wanita untuk menghasilkan kelahiran hidup. Kelahiran hidup yaitu terlepasnya bayi dari rahim seorang wanita dengan adanya tanda-tanda kehidupan seperti: bernapas, berteriak, jantung berdenyut dan lain-lain. Banyaknya anak yang dilahirkan hidup sangat erat kaitannya dengan kesejahteraan rumah tangga. Semakin banyak jumlah anak maka akan semakin banyak pula tanggungan yang



ditanggung oleh kepala rumah tangga. Untuk rumah tangga yang kondisi ekonominya lemah pembatasan jumlah anak merupakan salah satu cara tercapainya suatu rumah tangga yang sejahtera.

Pemerintah melalui program KB (Keluarga Berencana) yang memiliki tujuan khusus yaitu menurunkan jumlah angka kelahiran bayi, meningkatkan kesehatan keluarga berencana dengan cara penjarangan kelahiran dan meningkatkan jumlah penduduk sebagai pengguna alat kontrasepsi. KB adalah upaya peningkatan kepedulian dan peran serta masyarakat melalui pendewasaan usia perkawinan, pengaturan kelahiran, pembinaan ketahanan keluarga, peningkatan kesejahteraan keluarga untuk mewujudkan keluarga kecil, bahagia dan sejahtera (Undang-undang No.10/1992).

Tujuan tersebut dapat terealisasi salah satunya dengan penggunaan alat kontrasepsi oleh penduduk. Pengguna alat kontrasepsi adalah perempuan di rumah tangga berumur 15-49 tahun berstatus kawin, cerai hidup, cerai mati atau yang memiliki pasangan menggunakan alat KB atau cara tradisional guna menunda kehamilan (BPS, 2017:5). Alat KB atau cara tradisional seperti:sterilisasi wanita/tubektomi, sterilisasi pria/vasektomi, IUD/AKDR, suntikan, susuk KB/implan, PIL dan lain-lain. Pada penelitian ini untuk dimensi fertilitas dan KB yang digunakan sebagai variabel penelitian adalah perempuan berstatus miskin usia 15-49 tahun yang menggunakan alat KB.

## **2.6 Analisis Klaster**

Beberapa metode dalam analisis multivariat yang dapat digunakan untuk memahami dan mempermudah interpretasi data multivariat di antaranya adalah

analisis kluster, analisis diskriman, analisis komponen utama dan analisis faktor. Analisis kluster merupakan pengelompokan objek atau kasus menjadi kelompok-kelompok yang lebih kecil dimana setiap kelompok berisi objek yang mirip satu sama lain (Supranto, 2004). Dalam pengelompokan digunakan adalah ukuran jarak *Euclidean* (Johnson & Wichern, 1982).

Analisis kluster dibagi menjadi dua metode yaitu metode hirarki dan metode non-hirarki. Dalam metode hirarki jumlah kelompok yang akan diperoleh belum diketahui, sedangkan dalam metode non-hirarki diasumsikan ada  $k$  kelompok terlebih dahulu. Metode hirarki dibagi menjadi dua, yaitu metode *agglomerative* (pemusatan) dan metode *divise* (penyebaran). Pada metode *agglomerative*, langkah pertama masing-masing obyek pengamatan dijadikan sebagai kelompok yang memiliki satu anggota setiap kelompok. Langkah selanjutnya dua kelompok (atau obyek) yang memiliki jarak terdekat dikombinasikan ke dalam satu kelompok. Sedangkan pada metode *divisive* dilakukan hal yang sebaliknya, jadi semua obyek pengamatan dianggap sebagai satu kelompok kemudian dipisah sampai terbentuk kelompok-kelompok dengan anggota satu.

Hasil dari metode *agglomerative* dapat ditampilkan dalam bentuk diagram yang disebut dendogram. Dendogram menggambarkan proses pembentukan kluster yang dinyatakan dalam bentuk gambar. Garis mendatar di atas dendogram menunjukkan skala yang menggambarkan tingkat kemiripan, semakin kecil nilai skala menunjukkan semakin mirip individu tersebut. Ada beberapa kelebihan dan kelemahan dari analisis kluster (Raharto, 2008), yaitu:

Kelebihan analisis klaster antara lain:

1. Dapat mengelompokkan data observasi dalam jumlah besar dan variabel yang relatif banyak, sedemikian sehingga data yang direduksi dengan kelompok akan mudah dianalisis.
2. Dapat dipakai dalam skala data ordinal, interval dan rasio.

Sedangkan kelemahan analisis klaster antara lain:

1. Pengelompokan bersifat subjektifitas peneliti karena hanya melihat gambar dendogram.
2. Untuk data yang terlalu heterogen antara objek penelitian yang satu dengan yang lain akan sulit bagi peneliti untuk menentukan jumlah kelompok yang akan dibentuk.
3. Metode-metode yang dipakai memberikan perbedaan yang signifikan sehingga dalam perhitungan biasanya masing-masing metode dibandingkan.
4. Semakin besar observasi, biasanya tingkat kesalahan pengelompokan akan semakin besar.

## **2.7 Tahapan Analisis Klaster**

Terhadap yang dilakukan dalam analisis klaster secara umum adalah sebagai berikut:

### **2.7.1 Perumusan masalah**

Hal yang paling penting di dalam masalah analisis klaster adalah pemilihan variabel-variabel yang akan dipergunakan untuk pembentukan klaster (Supranto, 2004).

### 2.7.2 Standarisasi Data

Jika satuan peubah yang digunakan sama, maka standarisasi peubah asal  $X$  dapat menggunakan rumus sebagai berikut:

$$Z_{ij} = x_{ij} - \bar{X}_j; \text{ untuk } i=1,2,\dots,n \text{ dan } j=1,2,\dots,p$$

Dengan

$Z_{ij}$  = nilai baku objek ke- $i$  pada peubah ke- $j$

$x_{ij}$  = nilai pengamatan objek ke- $i$  pada peubah ke- $j$

$\bar{X}_j$  = rata-rata peubah ke- $j$

Namun jika satuan peubah yang digunakan tidak sama, maka dapat dilakukan standarisasi pada matriks data awal. Menurut Matjik & Sumertajaya (2011: 124), standarisasi peubah asal  $X$  dapat menggunakan rumus:

$$Z_{ij} = \frac{(x_{ij} - \bar{X}_j)}{s_j}; \text{ untuk } i=1,2,\dots,n \text{ dan } j=1,2,\dots,p \quad (3)$$

Dengan

$Z_{ij}$  = nilai baku objek ke- $i$  pada peubah ke- $j$

$x_{ij}$  = nilai pengamatan objek ke- $i$  pada peubah ke- $j$

$\bar{X}_j$  = rata-rata peubah ke- $j$

$S_j$  = simpangan baku peubah ke- $j$

## 2.8 Asumsi Analisis Kluster

### A. Sampel *Representatif*

Penggunaan sampel dalam penelitian harus dapat mewakili populasi atau representatif. Penggunaan sampel yang representatif akan memberikan

hasil yang maksimal dan sesuai dengan kondisi populasi yang ada. Namun apabila penelitian menggunakan populasi maka dapat disimpulkan bahwa asumsi representatif terpenuhi. (Hair, dkk., 1998).

#### B. Tidak Terdapat Multikolinieritas

Multikolinieritas adalah adanya hubungan linear yang sempurna atau pasti di antara beberapa atau semua variabel (Gujarati, 2009). Sebaiknya multikolinieritas ini tidak terjadi atau tidak terdapat multikolinieritas diantara variabel-variabel. Salah satu cara indentifikasi adanya multikolinieritas adalah menghitung nilai *Varians Inflation Factor* (VIF) dengan rumus sebagai berikut (Gujarati, 2009):

1. *Varians Inflation Factor* (VIF). Nilai  $VIF > 10$  menunjukkan adanya multikolinieritas. Rumus VIF adalah sebagai berikut:

$$VIF = \frac{1}{1-R^2} \quad (4)$$

Dengan  $R^2$  adalah nilai koefisien determinasi variabel dependen dengan variabel independen. Multikolinieritas terindikasi apabila nilai  $VIF > 10$ . Selain menggunakan nilai VIF juga bisa menggunakan matriks korelasi. Menurut Yamin dan Kurniawan (dalam Puspitasari dan Susanti, 2016) dikatakan terjadi multikolinieritas apabila nilai korelasi antar variabel  $> 70$ . Apabila terjadi multikolinieritas maka harus dilakukan tindakan perbaikan multikolinieritas. Salah satu cara yang dapat dilakukan bila terjadi multikolinieritas adalah dengan melakukan Analisis Komponen Utama.

## 2.9 *Principal Component Analysis (PCA)*

*Principal Component Analysis (PCA)* atau Analisis Komponen Utama (AKU) merupakan salah satu solusi dalam analisis kluster jika terjadi multikolinearitas dalam data. *PCA* bertujuan untuk mereduksi variabel menjadi lebih sedikit dari jumlah variabel sebelumnya dimana  $K$  variabel baru  $< K$  variabel lama. *Principal Component (PC)* merupakan suatu kombinasi linear dari variabel-variabel asal. Pembentukan *PC* berdasarkan dua cara yaitu matriks kovarian atau matriks korelasi (Johnson dan Wichern, 2007). Tahapan menentukan *PC* berdasarkan matriks korelasi adalah sebagai berikut:

1. Membuat matriks  $Z$  yang berisi data dari variabel  $X$  yang telah distandarisasi.
2. Membuat matriks korelasi dari  $Z$  yaitu  $Z'Z$ . Pereduksian *PC* dimulai dengan cara mencari nilai eigen  $\lambda_1, \lambda_2, \dots, \lambda_p$  yang diperoleh dari rumus

$$|Z'Z - \lambda I| = 0$$

Dimana jumlahan nilai eigen merupakan jumlah diagonal matriks korelasi, yaitu:

$$\sum_{j=1}^n \lambda_j = tr(Z'Z) \quad (5)$$

Nilai eigen selalu diurutkan dari yang terbesar sampai nilai terkecil. Nilai eigen menunjukkan besarnya total varian yang dijelaskan oleh *PC* yang terbentuk.

$PC_i$  saling orthogonal dan dibentuk berdasarkan persamaan:

$$PC1 = \gamma_{1j}z_1 + \gamma_{2j}z_2 + \dots + \gamma_{pj}z_p \quad (6)$$

Vektor eigen  $\gamma_j$  diperoleh dari setiap nilai eigen  $\lambda_i$  yang memenuhi persamaan:

$$(Z'Z - \lambda_j I)\gamma_j = 0 \quad (7)$$

Jumlah *Principal Component (PC)* yang terpilih berdasarkan nilai eigen ( $\lambda$ ) dimana nilai  $\lambda > 1$  maka PC tersebut akan dipilih (Hardika et al, 2013)

## 2.10 Ukuran Jarak atau Ketakmiripan

Metode analisis gerombol membutuhkan suatu ukuran ketakmiripan (jarak) yang didefinisikan untuk setiap pasang objek yang akan dikelompokkan. Untuk menentukan ketakmiripan antar objek dilihat berdasarkan kedekatan jarak antar objek tersebut. Makin dekat jaraknya, maka makin mirip karakteristik kedua objek tersebut. Jarak yang paling umum digunakan adalah jarak *Euclid*. Jarak *Euclid* digunakan bila peubah-peubah yang digunakan tidak berkorelasi dan memiliki satuan yang sama.

Metode ini memiliki beberapa keuntungan, antara lain jarak dari 2 objek apa saja tidak dipengaruhi oleh penambahan dari objek baru untuk di analisis, yang mungkin saja objek baru tersebut merupakan pencilan. Selain itu metode ini digunakan untuk meminimalkan variansi (ragam) antar objek yang ada dalam satu gerombol dan memaksimalkan variansi (ragam) antar gerombol. Dengan menggambarkan vektor  $x$  dan  $y$  berukuran  $n \times 1$  sebagai titik pada ruang yang berdimensi  $n$ , maka didefinisikan jarak antara  $x$  dan  $y$  sebagai norma dari vektor  $(x - y)$  dengan demikian jarak  $d(x, y)$  :

$$\begin{aligned} d(x, y) &= \sqrt{(x - y)'(x - y)} \\ &= \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \end{aligned} \quad (8)$$

Dengan  $x_i$  dan  $y_i$  berurutan adalah unsur ke- $i$  dari vektor  $x$  dan  $y$ . Jika terjadi korelasi antara peubah, maka dilakukan transformasi terhadap data awal dengan menggunakan Analisis Komponen Utama. Peubah-peubah baru yang dihasilkan dari AKU merupakan kombinasi linear dari peubah-peubah asal dan saling bebas linear (tidak berkorelasi). Pada penelitian ini, penulis menggunakan jarak *Euclidean* sebagai ukuran dari kemiripan suatu objek.

### 2.11 Metode Pengklasteran

Proses pengklasteran pada analisis kluster dilakukan menggunakan prosedur pengklasteran. Klasifikasi prosedur pengklasteran analisis kluster ini ditampilkan dalam bagan Gambar 1 (Simamora, 2005). Pada tahap ini, dapat dipilih salah satu prosedur atau metode untuk digunakan dalam pengklasteran. Menurut Hair *et al.* (2010) terdapat dua prosedur pengklasteran yaitu metode berhierarki dan metode tak berhierarki.

### 2.12 Metode Non Hirarki

Metode tak berhierarki digunakan apabila banyak gerombol yang akan dibentuk sudah diketahui terlebih dahulu. Metode tak berhierarki bertujuan untuk mengelompokkan  $n$  objek ke dalam  $k$  kelompok ( $k < n$ ).

### 2.13 Metode Hirarki

Metode berhierarki digunakan jika peneliti belum mengetahui banyaknya gerombol yang akan dibentuk. Terdapat dua prosedur pada metode berhierarki yaitu *agglomerative hierarchical clustering* dan *divisive hierarchical clustering*. Pada metode *agglomerative* prosesnya dimulai dengan masing-masing objek dianggap satu gerombol kemudian antar gerombol yang jaraknya berdekatan



bergabung menjadi satu gerombol besar yang beranggotakan semua objek. Sebaliknya, pada metode *divisive* prosesnya dimulai dengan semua objek berada dalam satu gerombol besar. Gerombol besar tersebut kemudian dipisah menjadi gerombol yang lebih kecil, hingga terbentuk satu gerombol hanya beranggotakan satu objek saja. Metode *agglomerative* terdapat beberapa metode perbaikan jarak yang dapat digunakan, yaitu *Single Linkage*, *Complete Linkage*, *Average Linkage* dan lain-lain (Matjik & Sumertajaya, 2011 : 199).

### 2.13.1 Metode Ward

Metode *Ward* adalah metode klastering hierarki yang bersifat aglomeratif untuk memperoleh klaster yang memiliki varian internal sekecil mungkin. Metode *Ward* berusaha untuk meminimalkan variasi antar objek yang ada dalam satu klaster. Jarak antara dua klaster yang terbentuk pada metode *Ward* adalah *sum of squares* di antara dua klaster tersebut. Metode *Ward* didasarkan pada kriteria *sum square error* (SSE) dengan ukuran kehomogenan antara dua objek berdasarkan jumlah kuadrat kesalahan yang paling minimal. SSE hanya dapat dihitung jika klaster memiliki elemen lebih dari satu objek. SSE klaster yang hanya memiliki satu objek adalah nol (Gudono, 2011). Formula untuk SSE adalah sebagai berikut (Rencher, 2002):

$$SSE = \sum_{i=1}^n (X_i - \bar{X})' (X_i - \bar{X}) \quad (9)$$

Dimana  $x_i$  adalah vektor kolom yang berisikan nilai objek  $i$ ,  $\bar{X}$  adalah vektor kolom yang entrinya rata-rata nilai objek dalam klaster,  $n$  adalah banyaknya objek. Jika  $UV$  adalah klaster yang diperoleh dengan

menggabungkan kluster U dan V, maka jumlah kuadrat di dalam kluster adalah:

$$SSE_U = \sum_{i=1} n_u (X_i - \bar{X}_u)' (X_i - \bar{X}_u)$$

$$SSE_V = \sum_{i=1} n_v (X_i - \bar{X}_v)' (X_i - \bar{X}_v)$$

$$SSE_{UV} = \sum_{i=1} n_{uv} (X_i - \bar{X}_{uv})' (X_i - \bar{X}_{uv})$$

Dengan  $SSE_U$ ,  $SSE_V$  dan  $SSE_{UV}$  adalah jumlah kuadrat kesalahan kluster U, V, dan UV,  $X_U$ ,  $X_V$  dan  $X_{UV}$  adalah vektor kolom yang berisikan nilai objek dari kluster U, V, dan UV,  $n_U$ ,  $n_V$  dan  $n_{UV}$  adalah banyaknya objek pada kluster U, V, dan UV. Metode *Ward* menggabungkan dua kluster U dan V dengan meminimalkan peningkatan SSE didefinisikan sebagai jarak kluster U dan kluster V yaitu sebagai berikut:

$$I_{UV} = SSE_{UV} - (SSE_U + SSE_V) \quad (10)$$

Hal itu dapat ditunjukkan bahwa peningkatan pada  $I_{UV}$  memiliki bentuk ekuivalen sebagai berikut:

$$\begin{aligned} I_{UV} &= n_u (\bar{X}_U - \bar{X}_{UV})' (\bar{X}_U - \bar{X}_{UV}) + n_v (\bar{X}_V - \bar{X}_{UV})' (\bar{X}_V - \bar{X}_{UV}) \\ &= \frac{n_u n_v}{n_u + n_v} (\bar{X}_U - \bar{X}_V)' (\bar{X}_U - \bar{X}_V) \end{aligned} \quad (11)$$

Dari persamaan di atas, meminimalkan peningkatan SSE ekuivalen dengan meminimalkan jarak antar objek. Jika U hanya terdiri dari  $X_i$  dan V hanya terdiri dari  $X_j$ , maka  $SSE_U$ ,  $SSE_V$  dan  $SSE_{UV}$  adalah nol. Selanjutnya dari persamaan (14) dan (15) diperoleh persamaan jarak antar dua objek yang

merupakan rumus yang digunakan untuk menentukan jarak dengan metode sebagai berikut:

$$\begin{aligned}
 I_{ij} = SSE_{ij} &= \frac{1}{2} (X_i - X_j)' (X_i - X_j) \\
 &= \frac{1}{2} d^2(X_i, X_j) \\
 &= \frac{1}{2} \sum_{k=1}^p (X_{ik} - X_{jk})^2
 \end{aligned} \tag{12}$$

Dengan  $I_{ij}$  adalah jarak antar objek  $i$  dan objek  $j$ ,  $x_{ik}$  adalah nilai objek  $i$  pada variabel ke- $k$ ,  $x_{jk}$  adalah nilai objek  $j$  pada variabel ke- $k$ ,  $d^2(X_i, X_j)$  adalah jarak *Squared Euclidean* antara objek  $i$  dan objek  $j$ . Dari hasil jarak antar objek, diperoleh kluster pertama yang memiliki jarak terdekat yang merupakan kluster yang paling homogen diantara kluster lain.

### 2.13.2 Metode *Single Linkage*

Untuk menentukan jarak antar kluster dengan menggunakan metode *single linkage* dapat dilakukan dengan melihat jarak antar dua kluster yang ada kemudian memilih jarak paling dekat atau aturan tetangga dekat (*nearest-neighbour rule*). Dihitung dengan cara:

$$d_{(UV)W} = \min\{d_{UV}, d_{VW}\} \tag{13}$$

Dimana nilai  $d_{UV}$  dan  $d_{VW}$  merupakan jarak minimum antara kluster  $U$  dan  $W$  serta kluster  $V$  dan  $W$  (Johnson dan Wichern, 2007). Berikut langkah penyelesaian dengan metode *single linkage*:

- a. Bentuk matriks jarak untuk matriks data yang diberikan, misalkan

$$D(1)_{n \times m} = \begin{bmatrix} d_{11} & d_{12} & \dots & d_{1n} \\ d_{21} & d_{22} & \dots & d_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ d_{n1} & d_{n2} & \dots & d_{nn} \end{bmatrix} \quad (14)$$

- b. Asumsikan setiap data dianggap sebagai klaster, kemudian tentukan klaster yang mempunyai jarak terdekat, misal klaster U dan klaster V mempunyai jarak terdekat kemudian digabungkan, hasil gabungannya adalah klaster UV.
- c. Dari klaster UV yang telah terbentuk cari jarak minimum antar klaster UV dengan klaster (objek) lainnya yang belum bergabung, matriks jarak baru yang diperoleh yaitu D(2). Misalkan  $d_{(UV)W} = \min(d_{UV}, d_{VW})$  maka klaster yang baru terbentuk adalah (UVW).
- d. Ulangi langkah (b) sampai semua objek bergabung menjadi satu kelompok.

### 2.12.3 Metode *Average Linkage*

*Average linkage* memperlakukan jarak antara dua klaster sebagai jarak rata-rata antara semua pasangan item-item di mana satu anggota dari pasangan tersebut dimiliki tiap klaster. Mulai dengan mencari matriks jarak  $D = (d_{ik})$  untuk memperoleh objek-objek paling dekat misalnya U dan V. Objek-objek ini digabungkan untuk membentuk klaster (UV). Dihitung dengan cara:

$$d_{(UV)W} = \frac{\sum_i \sum_k d_{ik}}{N_{(UV)} N_W} \quad (15)$$

Dimana  $d_{ik}$  adalah jarak antara objek  $i$  dalam klaster (UV) dan objek  $k$  dalam klaster  $W$ ,  $N_{UV}$  dan  $N_W$  berturut-turut adalah banyaknya item-item dalam klaster (UV) dan  $W$ .

### 2.12.3 Metode *Complete Linkage*

*Complete linkage* memberikan kepastian bahwa semua item-item dalam satu klaster berada dalam jarak paling jauh (similaritas terkecil) satu sama lain. Algoritma *agglomerative* pada umumnya dimulai dengan menentukan *entri* (elemen matriks) dalam  $D = (d_{ik})$  dan menggabungkan objek-objek yang bersesuaian misalnya  $U$  dan  $V$  untuk mendapatkan klaster (UV). Untuk langkah dari algoritma di atas jarak jarak antara klaster (UV) dan klaster  $W$  yang lain dihitung dengan cara:

$$d_{(UV)W} = \max\{d_{UW}, d_{VW}\} \quad (16)$$

Disini besaran-besaran  $d_{UW}$  dan  $d_{VW}$  berturut-turut adalah jarak antara tetangga terdekat klaster-klaster  $U$  dan  $W$  dan juga klaster-klaster  $V$  dan  $W$ .

(Johnson & Wichern, 1992)

### 2.14 Pemilihan Metode Terbaik

Pemilihan metode terbaik dimaksudkan untuk melihat kinerja masing-masing pengelompokan khususnya dalam mengelompokkan Kabupaten/Kota di Aceh. Untuk mengukur kinerja metode analisis yang digunakan, menurut Bunkers (dalam Komariah dan Akbar, 2011) dilakukan evaluasi kelompok dengan kriteria dua nilai simpangan baku, yaitu simpangan baku dalam kelompok ( $S_w$ ) dan antar kelompok ( $S_B$ ) sehingga diperoleh metode terbaik, berikut persamaan  $S_w$ :

$$S_w = K^{-1} \sum_{k=1}^k S_k \quad (17)$$

$S_w$  = rata-rata simpangan baku dalam klaster

$S_k$  = simpangan baku klaster ke-k

Dimana  $K$  adalah banyaknya kelompok yang terbentuk dan  $S_k$  merupakan simpangan baku kelompok ke-k. Nilai  $S_B$  dicari dengan persamaan berikut:

$$S_B = [(K - 1)^{-1} \sum_{k=1}^k k (\bar{X}_k - \bar{X})^2]^{1/2} \quad (18)$$

$S_B$  = simpangan baku antar kelompok

Dimana  $\bar{X}_k$  adalah rata-rata kelompok ke-k dan  $\bar{X}$  adalah rata-rata keseluruhan kelompok. Semakin kecil nilai  $S_w$  dan semakin besar nilai  $S_B$  maka metode tersebut memiliki kinerja yang baik, artinya mempunyai homogen yang tinggi (Komariah dan Akbar, 2011).

