

BAB II

TINJAUAN PUSTAKA

2.1 Kriminalitas

Menurut BPS (Badan Pusat Statistik), Tindak kejahatan/kriminalitas atau pelanggaran merupakan perbuatan seseorang yang dapat diancam hukuman berdasarkan kitab Undang-Undang Hukum Pidana (KUHP) atau undang-undang serta peraturan lainnya yang berlaku di Indonesia.

Menurut (Statistik, 2019), Indikator yang biasa digunakan untuk mengukur kejahatan adalah angka jumlah kejahatan (*crime total*), angka kejahatan per 100.000 penduduk (*crime rate*), dan selang waktu terjadinya suatu tindak kejahatan (*crime clock*). Meski demikian perlu kehati-hatian dalam memaknai angka kejahatan secara umum karena merupakan aritmetika sederhana yang menggabungkan semua jenis kejahatan dalam perhitungan tanpa mempertimbangan tingkat keseriusannya.

Rasa aman adalah kebutuhan dasar dan merupakan hak dasar seperti sandang, pangan, papan, kesehatan, dan pendidikan. UUD Republik Indonesia 1945 Pasal 28G ayat 1 menyebutkan bahwa setiap orang berhak atas perlindungan diri pribadi, keluarga, kehormatan, martabat, dan harta benda yang di bawah kekuasaannya, serta berhak atas rasa aman dan perlindungan dari ancaman ketakutan untuk berbuat atau tidak berbuat sesuatu yang merupakan hak asasi. Bahkan dalam *Universal Declaration of Human Rights* mencantumkan, bahwa setiap orang mempunyai hak untuk hidup dan hak akan kemerdekaan dan keamanan badan.

2.2 Kejahatan terhadap Nyawa

Menurut BPS (Badan Pusat Statistik), Kejahatan terhadap nyawa (pembunuhan/*homicide*) merupakan salah satu kejahatan yang paling tinggi hierarkinya dalam klasifikasi kejahatan internasional, selain itu dari sisi hukuman juga yang paling berat hukumannya dalam KUHP Indonesia.

Kejahatan terhadap nyawa dalam KUHP dapat dibedakan atau diklasterkan atas 2 dasar, yaitu :

1. Atas dasar unsur kesalahan
2. Atas dasar objeknya (nyawa)

Atas dasar kesalahan ada 2 (dua) klaster kejahatan terhadap nyawa yaitu :

1. Kejahatan terhadap nyawa yang dilakukan dengan sengaja (*dolus misdrijven*)
2. Kejahatan terhadap nyawa yang dilakukan karena kelalaian (*culpose misdrijven*)

2.2.1 Kejahatan terhadap nyawa yang dilakukan dengan sengaja (*dolus misdrijven*)

Adapun unsur-unsur tindak pidana dalam pasal 338 KUHP yang dikemukakan oleh Andi Abu Ayyub Saleh adalah sebagai berikut :

A. Dengan sengaja

1. Unsur sengaja meliputi tindakannya dan objeknya, artinya si pembuat atau pelaku mengetahui atau menghendaki adanya orang mati dari perbuatannya tersebut. Hilangnya jiwa seseorang harus dikehendaki dan harus menjadi

tujuan, sehingga karenanya perbuatan yang dilakukan tersebut dengan suatu maksud atau tujuan yakni adanya niat untuk menghilangkan nyawa orang lain.

2. Jika timbulnya akibat hilangnya jiwa orang lain tanpa dengan sengaja atau bukan menjadi tujuan atau bukan bermaksud dan tidak pernah diniatkan tidaklah dapat dikaitkan sebagai pembunuhan tidak dapat dikenakan ketentuan tindak pidana pembunuhan tersebut tetapi mungkin dapat dikenakan tindak pidana lain yang mengakibatkan orang mati tetapi tidak dengan unsur sengaja.
3. Baik timbulnya akibat maupun perbuatan yang menimbulkannya harus dilakukan dengan sengaja, jadi pelaku atau pembuat harus mengetahui dan menghendaki bahwa dari perbuatannya itu dapat bahkan pasti mengakibatkan adanya orang mati.
4. Untuk memenuhi tindak pidana pembunuhan dengan unsur sengaja yang terkadang dalam pasal 338 KUHP ini disyaratkan bahwa perbuatan pembunuhan tersebut harus dilakukan sesegera mungkin sesudah timbulnya suatu maksud atau niat untuk membunuh tidak dengan pikir-pikir atau tidak dengan suatu perencanaan.
5. Unsur sengaja ini dalam praktek seringkali sulit untuk membuktikannya, terutama jika pemuat atau pelaku tersebut licik ingin menghindari dari perangkat tindak pidana tersebut. Karena unsur dengan sengaja adalah unsur subjektif adalah unsur batin si pembuat yang hanya dapat diketahui dari keterangan tersangka atau terdakwa di depan pemeriksaan penyidik atau di depan pemeriksaan persidangan, kecuali mudah pembuktiannya unsur ini apabila tersangka atau terdakwa tersebut memberi keterangan sebagai

“pengakuan” artinya mengakui terus terang pengakuannya bahwa kematian di korban tersebut memang dikehendaki atau menjadi tujuannya.

6. Pada umumnya kasus-kasus tindak pidana pembunuhan si tersangka atau terdakwa berusaha menghindar dari pengakuan unsur sengaja tetapi selalu berlindung bahwa kematian si korban tersebut tidak dikehendaki atau bukan menjadi niat tujuannya yakni hanya ingin menganiaya saja atau melukainya saja.

7. Untuk membuktikan unsur sengaja menurut ketentuan ini haruslah dilihat sara melakukan dalam mewujudkan perbuatan jahatnya tersebut. Sehingga memang dikehendaki atau diharapkan supaya korbannya meninggal dunia.

B. Pembunuhan yang diikuti, disertai atau didahului oleh tindak pidana lain
Pembunuhan yang dimaksud adalah sebagaimana dirumuskan dalam pasal 339 KUHP yaitu :”Pembunuhan yang diikuti, disertai atau didahului oleh suatu tindak pidana yang dilakukan dengan maksud untuk mempersiapkan atau mempermudah pelaksanaannya, atau untuk menghindarkan diri sendiri maupun peserta lainnya dari pidana dalam hal tertangkap tangan, ataupun untuk memastikan penguasaan benda yang diperolehnya secara melawan hukum, dipidana dengan pidana penjara seumur hidup atau sementara waktu paling lama 20 tahun”.

C. Pembunuhan berencana (*Moord*)

Pembunuhan dengan rencana terlebih dahulu atau disingkat pembunuhan berencana adalah pembunuhan yang paling berat ancaman pidananya dari seluruh bentuk kejahatan terhadap nyawa manusia, diatur dalam pasal 340 yang

rumusannya adalah “Barangsiapa dengan sengaja dan dengan rencana terlebih dahulu menghilangkan nyawa orang lain, dipidana karena pembunuhan dengan berencana dengan pidana mati atau pidana penjara seumur hidup atau selama waktu tertentu, paling lama 20 tahun”.

2.2.2 Kejahatan terhadap nyawa yang dilakukan karena kelalaian (*culpose misdrijven*)

Kejahatan terhadap nyawa yang dilakukan karena kelalaian adalah kejahatan yang dirumuskan dalam pasal 359 KUHP yaitu “Barangsiapa karena kesalahannya (kealpaannya) menyebabkan orang lain mati, dipidana kurungan paling lama 1 tahun”.

2.3 Kejahatan terhadap Fisik/Badan

Menurut BPS (Badan Pusat Statistik), kejahatan terhadap fisik termasuk di dalamnya adalah penganiayaan ringan, penganiayaan berat, dan kekerasan dalam rumah tangga. Penganiayaan adalah perbuatan dengan sengaja merusak kesehatan fisik orang lain baik menimbulkan penyakit (luka/cacat/sakit) atau halangan untuk menjalankan pekerjaan sehari-hari (luka berat) maupun tidak. Pemukulan, penamparan, pengeroyokan, termasuk kategori penganiayaan.

2.4 Kejahatan terhadap Kesusilaan

Menurut BPS (Badan Pusat Statistik), dalam klasifikasi kejahatan terhadap kesusilaan terdiri dari jenis kejahatan perkosaan dan pencabulan. Pelecehan seksual adalah perbuatan yang menyerang kehormatan kesusilaan (cubitan, colekan, tepukan, sentuhan di bagian tubuh tertentu atau gerakan) maupun perbuatan cabul yang dilakukan secara sepihak dan tidak diharapkan oleh korban. Perbuatan

tersebut dapat dilakukan dengan kekerasan fisik, ancaman kekerasan, maupun tidak. Perkosaan (perbuatan yang dilakukan dengan kekerasan atau ancaman kekerasan, memaksa korban untuk bersetubuh dengannya) termasuk kategori pelecehan seksual.

2.5 Kejahatan terhadap Kemerdekaan Orang

Menurut BPS (Badan Pusat Statistik), dalam klasifikasi kejahatan terhadap kemerdekaan orang terdiri dari jenis kejahatan penculikan dan mempekerjakan anak dibawah umur. Perniagaan budak belian adalah perniagaan yang barang dagangannya terdiri dari orang-orang yang akan dipergunakan untuk dijadikan budak atau hamba belian. Menurut pasal 324 KUHP yang dinamakan perdagangan budak yaitu :”barang siapa dengan ongkos sendiri atau ongkos orang lain menjalankan perniagaan budak belian atau melakukan perbuatan perniagaan budak belian atau dengan sengaja turut campur dalam segala sesuatu itu, baik dengan langsung maupun tidak langsung “.

2.6 Kejahatan terhadap Hak Milik/Barang dengan Penggunaan Kekerasan

Menurut BPS (Badan Pusat Statistik), dalam klasifikasi kejahatan ini yang termasuk kejahatan terhadap hak/milik dengan penggunaan kekerasan adalah pencurian dengan kekerasan menggunakan senjata api (senpi) dan pencurian dengan kekerasan menggunakan senjata tajam (sajam).

Menurut (Statistik, 2019), Pencurian dengan kekerasan adalah mengambil sesuatu barang atau ternak, yang sama sekali atau sebagian termasuk kepunyaan orang lain dengan maksud akan memiliki barang barang itu dengan melawan hak, yang didahului, disertai, atau diikuti dengan kekerasan atau ancaman kekerasan

terhadap orang, dengan maksud akan menyiapkan atau memudahkan pencurian itu, atau jika tertangkap tangan (terpergok) supaya dia dan kawannya yang turut melakukan kejahatan itu sempat melarikan diri atau supaya barang yang dicuri tetap ada di tangannya. Perampokan, penodongan, pemalakan, penjambretan, termasuk dalam kategori pencurian dengan kekerasan.

2.7 Kejahatan terhadap Hak Milik/Barang tanpa Penggunaan Kekerasan

Menurut BPS (Badan Pusat Statistik), dalam klasifikasi kejahatan ini yang termasuk adalah pencurian, pencurian dengan pemberatan, pencurian kendaraan bermotor, pengrusakan/penghancuran barang, pembakaran dengan sengaja, dan penadahan.

2.8 Kejahatan terkait Narkotika

Menurut BPS (Badan Pusat Statistik), dalam klasifikasi kejahatan ini adalah yang menggunakan narkotika dan psikotropika. Kejahatan narkotika merupakan suatu kejadian transnasional yang saat ini menjadi suatu kejahatan lintas negara yang terjadi di Indonesia. Pada 2018, jenis kejahatan yang termasuk kejahatan luar biasa (*extraordinary crime*) ini memiliki pola yang berbeda dengan kejahatan lain.

Dalam pasal 114 ayat (2) UU Narkotika menyebutkan, “Dalam hal perbuatan menawarkan untuk dijual, menjual, membeli, menerima, menjadi perantara dalam jual beli, menukar, menyerahkan, atau menerima narkotika golongan 1 sebagaimana dimaksud pada ayat (1) yang dalam bentuk tanaman beratnya melebihi 1 (satu) kilogram atau melebihi 5 (lima) batang pohon atau dalam bentuk bukan tanaman beratnya 5(lima) gram, pelaku pidana penjara paling singkat 6

(enam) tahun dan paling lama 20 (dua puluh) tahun dan pidana denda maksimum sebagaimana dimaksud pada ayat (1) ditambah $\frac{1}{3}$ (sepertiga)”.

2.9 Kejahatan terkait penipuan, Penggelapan, dan Korupsi

Menurut BPS (Badan Pusat Statistik), dalam klasifikasi kejahatan ini yang termasuk adalah penipuan, penggelapan, dan korupsi. Tindak pidana penggelapan diatur dalam pasal 372 KUHP yang berbunyi “Barang siapa dengan sengaja menguasai secara melawan hukum sesuatu benda yang seharusnya atau sebagian merupakan kepunyaan orang lain yang berada padanya bukan karena kejahatan, karena bersalah melakukan penggelapan, dipidana dengan pidana penjara selamalamanya 4 (empat) tahun atau dengan pidana denda setinggi-tingginya 900 (sembilan ratus) rupiah”.

Tindak pidana korupsi diatur dalam pasal 2 ayat (1) UU No. 31 tahun 1999 tentang pemberantasan tindak pidana korupsi sebagaimana telah diubah dengan UU No. 20 tahun 2001 yang berbunyi “Setiap orang yang secara melawan hukum melakukan perbuatan memperkaya diri sendiri atau orang lain atau suatu korporasi yang dapat merugikan keuangan negara atau perekonomian negara, dipidana penjara dengan penjara seumur hidup atau pidana penjara paling singkat 4 (empat) tahun dan paling lama 20 (dua puluh) tahun dan denda paling sedikit Rp. 200.000.000,00 (dua ratus juta rupiah) dan paling banyak Rp. 1000.000.000,00 (satu milyar rupiah)”.

2.10 Analisis Klaster

2.10.1 Pengertian Klaster

Cluster atau 'klaster' dapat diartikan 'klaster;' dengan demikian, pada dasarnya analisis klaster akan menghasilkan sejumlah klaster (klaster) (Santoso, 2014 : 125). Analisis klaster yaitu analisis untuk mengelompokkan elemen yang mirip sebagai objek penelitian menjadi klaster yang berbeda dan *mutually exclusive* (Supranto, 2004 : 26).

Sekilas memang terdapat kemiripan antara analisis faktor dengan analisis klaster, keduanya disebut dengan '*factor analysis*'. Namun, terdapat perbedaan penting diantara keduanya. Jika analisis faktor (disebut dengan *R Factor Analysis*) bertujuan mereduksi sejumlah variabel menjadi satu atau beberapa faktor, maka analisis klaster (disebut dengan *Q Factor Analysis*) lebih bertujuan untuk mengelompokkan isi variabel, walaupun bisa juga disertai dengan pengelompokan variabel. Jika dikaitkan dengan pengolahan data menggunakan SPSS, analisis faktor adalah perlakuan terhadap isi dari KOLOM data, sedangkan analisis klaster adalah perlakuan terhadap isi dari BARIS data (Santoso, 2014 : 126).

Tujuan utama analisis klaster ialah mengklasifikasi objek (kasus/elemen) seperti orang, produk (barang), toko, perusahaan kedalam klaster-klaster yang relatif homogen didasarkan pada suatu set variabel yang dipertimbangkan untuk diteliti (Supranto, 2004 : 141).

2.10.2 Ciri Sebuah klaster yang baik

Menurut (Santoso, 2014) Analisis klaster akan membagi sejumlah data pada satu atau beberapa klaster tertentu. Sebuah klaster yang baik adalah klaster yang mempunyai :

- a. Homogenitas (kesamaan) yang tinggi antara anggota dalam satu *Cluster* (*within- Cluster*)
- b. Heterogenitas (perbedaan) yang tinggi antara *Cluster* yang satu dengan *Cluster* lainnya (*betweenCluster*)

Dari dua hal di atas dapat disimpulkan bahwa sebuah klaster yang baik adalah klaster yang mempunyai anggota-anggota yang semirip mungkin satu dengan yang lain, namun sangat tidak mirip dengan anggota-anggota klaster yang lain. Di sini, 'mirip' diartikan sebagai tingkat kesamaan karakteristik antara dua data.

2.10.3 Asumsi pada analisis klaster

Menurut (Santoso, 2014) Asumsi untuk analisis klaster yaitu :

- a. Sampel yang diambil benar-benar bisa mewakili populasi yang ada. Memang tidak ada ketentuan jumlah sampel yang representatif, namun tetaplah diperlukan sejumlah sampel yang cukup besar agar proses *clustering* bisa dilakukan dengan benar.
- b. Multikolinieritas, yakni kemungkinan adanya korelasi antar-objek. Sebaliknya tidak ada atau seandainya ada, besar multikolinieritas tersebut tidaklah tinggi (misal di atas 0,5). Jika sampai terjadi multikolinieritas, dianjurkan untuk menghilangkan salah satu variabel dari dua variabel yang mempunyai korelasi cukup besar.

2.10.4 Proses Analisis klaster

Menurut (Santoso : 2014), proses analisis klaster adalah sebagai berikut :

1. Proses Standarisasi

Bila terdapat data (variabel) yang mempunyai perbedaan ukuran satuan yang besar perlu dilakukan proses standarisasi. Perbedaan satuan yang mencolok akan menyebabkan bias dalam analisis klaster, sehingga data asli harus ditransformasi (standarisasi) sebelum bisa dianalisis. Transformasi dilakukan terhadap variabel yang relevan ke dalam bentuk z skor, seperti berikut :

$$Z = \frac{x - \bar{x}}{\sigma} \quad (2.2)$$

Keterangan :

x = nilai data

\bar{x} = nilai rata-rata

σ = standar deviasi

2. Mengukur Kesamaan Antar Objek (*Similarity*)

Konsep kesamaan adalah hal yang fundamental dalam analisis klaster. Kesamaan antar objek merupakan ukuran korespondensi antar objek. Ada tiga metode yang dapat diterapkan, yaitu ukuran korelasi, ukuran jarak dan ukuran asosiasi. Sesuai prinsip dasar klaster yaitu mengelompokkan objek yang mempunyai kemiripan (Supranto, 2004).

a. Ukuran Korelasi

Ukuran ini dapat diterapkan pada data dengan skala metrik, namun jarang digunakan karena titik beratnya pada nilai suatu pola tertentu, padahal fisik berat analisis kluster adalah besarnya objek.

Kesamaan antar objek dapat dilihat dari koefisien korelasi antar pasangan objek yang diukur dengan berat variabel (Supranto, 2004). Yang termasuk jenis ukuran kesamaan adalah :

b. Ukuran Mutlak Korelasi

Dalam ukuran kesamaan ini tidak mempertimbangkan nilai korelasi yang positif ataupun negatif dari objek yang dianalisis. Ukuran ini menggunakan nilai mutlak dari korelasinya sehingga ukuran kemiripan objek i terhadap j atau dinotasikan d_{ij} dapat dirumuskan sebagai berikut :

$$d_{ij} = 1 - |r_{ij}| \quad (2.3)$$

Dengan r_{ij} adalah koefisien korelasi antar objek ke- i dan objek ke- j dan $|r_{ij}|$ merupakan nilai mutlaknya.

c. Ukuran jarak korelasi

Dalam ukuran kesamaan ini nilai korelasi yang positif ataupun negatif dari objek yang dianalisis mempengaruhi pengukuran kesamaannya. Ukuran kemiripan objek i dengan objek j dapat dirumuskan sebagai berikut :

$$d_{ij} = 1 - r_{ij} \quad (2.4)$$

Dengan r_{ij} adalah koefisien korelasi objek ke- i dan objek ke- j .

d. Ukuran jarak

Pengukuran ada bermacam-macam, yang paling populer adalah metode *Euclidean Distance*. Pada dasarnya, cara ini akan memasukkan sebuah data kedalam kluster tertentu dengan mengukur ‘jarak’ data tersebut dengan pusat kluster. Jika data ada dalam jarak yang masih ada dalam batas tertentu, data tersebut dapat dimasukkan pada kluster tersebut (Santoso, 2014 : 128).

e. Ukuran Asosiasi

Pada dasarnya, cara ini akan mengasosiasikan sebuah data dengan cluster tertentu; dalam praktek, cara ini tidak sepopuler kedua cara sebelumnya (Santoso, 2014 : 128).

3. Membuat Kluster (*Clustering*)

Ada dua proses penting yaitu algoritma kluster dalam pembentukan kluster dan menentukan jumlah kluster yang akan dibentuk. Keduanya mempunyai implikasi substansial tidak hanya pada hasil yang diperoleh tetapi juga pada interpretasi yang akan dilakukan terhadap hasil tersebut (Supranto, 2004).

4. Interpretasi

Setelah kluster terbentuk, langkah selanjutnya melakukan interpretasi. Tahap interpretasi meliputi pengujian tiap kluster dalam term untuk menamai dan menandai dengan suatu label secara akurat dapat menjelaskan kealamian kluster. Proses ini dimulai dengan suatu ukuran yang sering digunakan yaitu centroid cluster. Membuat profil dan interpretasi kluster tidak hanya untuk memperoleh suatu gambaran saja melainkan pertama yaitu menyediakan suatu rata-rata untuk menilai korespondensi pada kluster yang terbentuk dan kedua yaitu profil kluster

memberikan arah bagi penelitian terhadap signifikansi praktis. Pada intinya memberi nama spesifik untuk menggambarkan isi kluster tersebut (Supranto, 2004).

5. Validasi dan *profiling* Kluster

Menurut (Supranto : 2004), validasi dan *profiling* Kluster yaitu sebagai berikut :

a. Validasi solusi kluster

Proses validasi bertujuan untuk menjamin bahwa solusi yang dihasilkan dari analisis kluster dapat mewakili populasi dan dapat digeneralisasi untuk objek lain. Pendekatan ini membandingkan solusi kluster dan menilai korespondensi hasil. Terkadang tidak dapat dipraktekkan karena adanya kendala waktu dan biaya atau ketidaktersediaan objek untuk analisis kluster ganda (Supranto, 2004).

b. *Profiling* solusi kluster

Tahap ini menggambarkan karakteristik tiap kluster untuk menjelaskan kluster-kluster tersebut dapat berbeda pada dimensi yang relevan. Titik bertnya pada karakteristik yang secara signifikan berbeda antar kluster dan memprediksi anggota dalam suatu kluster khusus(Supranto, 2004).

2.11 Analisis Komponen Utama

Menurut Supranto (2004), analisis faktor merupakan suatu analisis yang digunakan untuk mereduksi dan meringkas data. Faktor-faktor diekstrasi sehingga faktor pertama memberikan andil terbesar terhadap seluruh varian dari seluruh variabel asli, faktor kedua menyumbang terbesar kedua, faktor ketiga menyumbang terbesar ketiga, dan begitu seterusnya sehingga proses pencarian faktor dihentikan setelah sumbangan terhadap seluruh varian variabel mencapai 60% atau lebih. Dimungkinkan pula untuk mengekstraksi faktor sehingga faktor tidak berkorelasi

sesamanya seperti di dalam analisis komponen utama atau *Principal Component Analysis* (PCA). Jika terdapat variabel yang terjadi multikolinearitas, maka salah satu solusi untuk menangani masalah tersebut adalah dengan analisis komponen utama. Analisis komponen utama adalah suatu analisis yang bertujuan untuk mentransformasikan p variabel asal yang masih berkorelasi satu dengan yang lain menjadi satu set variabel baru yang tidak berkorelasi lagi.

Pembentukan *Principal Component* (PC) merupakan suatu kombinasi linear dari variabel-variabel asal. Pembentukan *Principal Component* (PC) terdapat dua cara yaitu dengan matriks kovarian atau matriks korelasi (Johnson dan Wichern, 2007). Tahapan pembentukan *Principal Component* (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 *Principal Component* (PC) dimulai dengan cara mencari nilai eigen $\lambda_1, \lambda_2, \dots, \lambda_p$ yang diperoleh dari persamaan :

$$|Z'Z - \lambda_1| = 0 \quad (2.5)$$

dimana jumlah nilai eigen merupakan jumlah diagonal matriks korelasi, yaitu :

$$\sum_{j=1}^n \lambda_j = tr(Z'Z) \quad (2.6)$$

Nilai eigen selalu diurutkan dari yang terbesar sampai nilai terkecil. Nilai eigen menunjukkan besarnya total varian yang dijelaskan oleh *Principal Component* (PC) yang terbentuk. PC_j saling orthogonal dan dibentuk berdasarkan persamaan :

$$PC_j = \gamma_{1j}Z_1 + \gamma_{2j}Z_2 + \dots + \gamma_{pj}Z_p \quad (2.7)$$

Vektor eigen γ_j diperoleh dari setiap nilai eigen λ_1 yang memenuhi persamaan :

$$|Z'Z - \lambda_j I| \gamma_j = 0 \quad (2.8)$$

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

2.12 Metode Analisis Klaster

Didalam analisis klaster terdapat dua metode analisis klaster yaitu metode hierarki dan metode non-hierarki. Keduanya memiliki kelebihan dan kekurangan masing-masing. Kelebihan metode hierarki yaitu cepat dalam proses pengolahan sehingga menghemat waktu, namun kekurangannya yaitu metode ini dapat menimbulkan kesalahan. Selain itu tidak baik diterapkan untuk menganalisis sampel dengan ukuran besar. Metode non-hierarki memiliki kelebihan yang lebih baik dibandingkan metode hierarki. Hasilnya memiliki sedikit kelemahan pada data *outlier*. Ukuran jarak yang digunakan dan termasuk variabel tak relevan atau variabel yang tidak tepat. Keuntungannya adalah hanya dengan menggunakan titik bakal non *random*, penggunaan metode non-hierarki untuk titik bakal *random* secara nyata lebih buruk dari pada metode hierarki. Penentuan metode ini yang akan dipakai tergantung kepada peneliti dan konteks penelitian dengan tidak mengabaikan substansi, teori dan konsep yang berlaku. Alternatif lain adalah dengan mengkombinasikan kedua metode ini. Pertama gunakan metode hierarki kemudian dilanjutkan dengan metode non-hierarki (Supranto, 2004).

2.12.1 Metode Hierarki

Metode ini memulai pengklasteran dengan dua atau lebih objek yang mempunyai kesamaan paling dekat. Kemudian proses diteruskan ke objek lain yang mempunyai kedekatan kedua. Demikian seterusnya sehingga klaster akan

membentuk semacam ‘pohon’ dimana ada hierarki (tingkatan) yang jelas antar-objek, dari yang paling mirip sampai paling tidak mirip. Secara logika semua objek pada akhirnya hanya akan membentuk sebuah klaster. *Dendogram* biasanya digunakan untuk membantu memperjelas proses hierarki tersebut (Santoso, 2014 : 129).

Andaikan d_{ij} merupakan ukuran ketidakmiripan antara klaster ke- i dengan klaster ke- j dan $d_{k(i,j)}$ merupakan ukuran ketidakmiripan antara klaster ke- k dengan klaster (i,j) yang merupakan penggabungan antara klaster ke- i dengan klaster ke- j . Berikut ini diberikan ukuran ketidakmiripan antar klaster.

1. Pautan Tunggal (*Single Linkage/Nearest Neighbour Method*)

Metode ini akan mengelompokkan dua objek yang mempunyai jarak terdekat terlebih dahulu (Santoso, 2014). Ukuran ketidakmiripan yang digunakan adalah

$$d_{k(i,j)} = \min\{d_{ki}, d_{kj}\} \quad (2.9)$$

2. Pautan Lengkap (*Complete Linkage/Furthest Neighbour Method*)

Metode ini akan mengelompokkan dua objek yang mempunyai jarak terjauh terlebih dahulu (Santoso, 2014). Ukuran ketidakmiripan yang digunakan adalah

$$d_{k(i,j)} = \max\{d_{ki}, d_{kj}\} \quad (2.10)$$

3. Pautan Rataan (*Average Linkage/Between Neighbour Method*)

Metode ini akan mengelompokkan objek berdasar jarak rata-rata yang didapat dengan melakukan rata-rata semua jarak antar objek terlebih dahulu (Santoso, 2014). Ukuran ketidakmiripan yang digunakan adalah

$$d_{k(i,j)} = \frac{n_i}{n_i + n_j} d_{ki} + \frac{n_j}{n_i + n_j} d_{kj} \quad (2.11)$$

Dengan n_i adalah jumlah item pada kluster i .

4. Metode *Centroid*

Metode ini akan mengelompokkan objek berdasar jarak rata-rata yang didapat dengan melakukan rata-rata semua jarak antar objek terlebih dahulu (Santoso, 2014). Ukuran ketidakmiripan yang digunakan adalah

$$d_{k(i,j)} = \frac{n_i}{n_i + n_j} d_{ki} + \frac{n_j}{n_i + n_j} d_{kj} - \frac{n_i + n_j}{(n_i + n_j)^2} d_{ij} \quad (2.12)$$

5. *Ward's Error Sum of Square Method*

Metode *Ward's* ditemukan oleh Ward (1963) tidak menghitung jarak antar kluster atau objek, namun metode ini membentuk kluster-kluster dengan memaksimalkan kehomogenan dalam kluster. Jumlah kuadrat dalam kluster digunakan sebagai ukuran kehomogenan. Metode *Ward's* terus mencoba meminimalkan total jumlah kuadrat dalam kluster. Kluster-kluster dibentuk pada masing-masing tahap seolah-olah telah menjadi solusi analisis kluster dan solusi terbaik adalah kombinasi kluster yang mempunyai jumlah kuadrat dalam kluster tekecil. Jumlah kuadrat dalam kluster (jarak kadrat *euclidean* masing-masing objek terhadap *mean* kluster yang memuat objek kluster tersebut) yang meminimalkan sering disebut

$$ESS = \sum_{k=1}^k \left\{ \sum_{i=1}^{nk} \sum_{j=1}^p x_{ijk}^2 - \frac{1}{nk} \left(\sum_{i=1}^{nk} x_{ijk} \right)^2 \right\} \quad (2.13)$$

Ukuran ketidakmiripan di atas dapat ditulis secara umum sebagai

$$d_{k(i,j)} = a_i d_{kj} + a_j d_{ki} + \beta a_{ij} + d^3 d_{ki} - d^3 d_{kj} \quad (2.14)$$

Yang akan menjadi ketidakmiripan pautan tunggal bila $a_i = a_j = \frac{1}{2}$, $\beta = 0$ dan $d = -\frac{1}{2}$. Akan menjadi pautan lengkap bila $a_i = a_j = \frac{1}{2}$, $\beta = 0$ dan $d = -\frac{1}{2}$. Akan menjadi pautan rata-rata bila $a_i = \frac{n_i}{n_i + n_j}$, $a_j = \frac{n_j}{n_i + n_j}$, $\beta = -a_i a_j$ dan $d = 0$.

Metode varians bertujuan untuk memperoleh kluster yang memiliki varians internal kluster yang sekecil mungkin. Metode varians yang umum dipakai adalah metode *ward* dimana rata-rata untuk setiap kluster dihitung. Lalu, dihitung jarak *Euclidean* antara setiap obyek dan nilai rata-rata itu, lalu jarak itu dihitung semua. Pada setiap tahap, dua kluster yang memiliki kenaikan '*sum of squares* dalam *cluster*' yang terkecil digabungkan (Simamora, 2005).

Metode *ward* merupakan suatu metode pembentukan kluster yang didasari oleh hilangnya informasi akibat penggabungan obyek menjadi kluster. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada *mean* kluster untuk setiap pengamatan. *Error sum of squares* (SSE) digunakan sebagai fungsi obyektif. Dua obyek akan digabungkan jika mempunyai fungsi obyektif terkecil diantara kemungkinan yang ada.

$$SSE = \sum_{j=1}^p (\sum_{i=1}^n x_{ij}^2 - \frac{1}{n} (\sum_{i=1}^n x_{ij})^2) \quad (2.15)$$

Dimana :

x_{ij} : nilai untuk obyek ke-I pada kluster ke-j

p : banyaknya variabel yang diukur

n : banyaknya obyek dalam klaster yang terbentuk

Langkah penyelesaian metode *Ward* yaitu :

1. Dimulai dengan memperhatikan N klaster yang mempunyai satu responden per klaster (semua responden dianggap sebagai klaster). Pada tahap pertama ini SSE bernilai nol.
2. Klaster pertama dibentuk dengan memilih dua dari N klaster yang memiliki nilai SSE terkecil.
3. $N-1$ kumpulan klaster kemudian diperhatikan kembali untuk menentukan dua dari klaster ini yang bisa meminimumkan keheterogenan. Dengan demikian N klaster secara sistematis dikurangi $N-1$.
4. Mengulangi langkah 3 dan 4 sampai diperoleh satu klaster atau semua responden bergabung menjadi satu klaster.

2.12.2 Metode Non-Hierarki

Berbeda dengan metode hierarki, metode ini justru dimulai dengan menentukan terlebih dahulu jumlah klaster yang diinginkan (dua klaster, tiga klaster atau yang lain). Setelah jumlah klaster diketahui, baru proses klaster dilakukan tanpa mengikuti proses hierarki. Metode ini biasa disebut dengan *K-Means cluster* (Santoso, 2014 : 129). Ada tiga prosedur dalam metode non-hierarki yaitu :

1. *Sequential Threshold*

Proses clustering dengan menggunakan metode *Sequential Threshold* dimulai dengan memilih sebuah klaster *seed* dan kemudian menggabungkan setiap objek yang ada dalam jarak yang telah ditentukan sebelumnya. Klaster tersebut akan disebut sebagai klaster pertama. Setelah klaster pertama terbentuk, maka klaster

seed kedua digunakan dan kemudian objek-objek yang mempunyai jarak terdekat akan digabungkan. Jika telah selesai, maka kluster berikutnya akan dibentuk dengan cara yang sama.

Metode ini disebut *Sequential Threshold* karena proses *clustering* dilakukan berurutan dari kluster pertama, kedua, ketiga dan berikutnya. Metode ini tidak memperbolehkan suatu objek pindah ke kluster lain.

2. *Parallel Threshold*

Proses *clustering* dengan menggunakan metode *Parallel Threshold* dimulai dengan memiliki kluster *seed* yang akan dijadikan patokan pembuatan kluster. Setiap objek akan diukur terhadap kluster *seed* tersebut. Sebuah objek akan masuk ke suatu kluster jika mempunyai nilai jarak terhadap suatu kluster *seed* lebih dekat daripada kluster *seed* yang lain. Langkah ini mengakibatkan penentuan kluster tidak berurutan.

Metode ini disebut *Parallel Threshold* karena proses *clustering* dilakukan tidak berurutan. Metode ini tidak memperbolehkan suatu objek yang sudah menjadi anggota suatu kluster berpindah ke kluster lain.

3. *Optimizing Partitioning*

Pada metode *optimizing partitioning*, objek digabungkan terakhir kedalam kluster untuk mengoptimalkan jarak kedekatan. Tahapan dari metode *partitioning* dimulai dengan menentukan banyaknya partisi yang akan dibentuk yaitu k , kemudian melakukan pengukuran jarak pada tiap objek. Selanjutnya, dilakukan *iterative relocation technique* dengan memindahkan objek dari satu kluster ke kluster yang lain.

Metode *Partitioning* yang sering digunakan yaitu *k-means* dan *k-medoids*. Menurut Hans dan Kamber (2006), algoritma *k-means* sangat sensitif terhadap pencilan, karena menggunakan nilai rata-rata (mean) sebagai pusat klasternya. Untuk mengatasi hal tersebut, digunakan metode *k-medoids* untuk mengelompokkan objek-objek pada suatu data yang mengandung pencilan.

2.13 Penentuan Jumlah klaster dengan Metode *Silhouette*

Banyak metode validasi *clustering* yang dikembangkan hingga saat ini baik menggunakan kriteria internal, eksternal, dan relatif. Pendekatan nilai rata-rata metode *silhouette* yaitu untuk menduga kualitas dari klaster yang terbentuk. Semakin tinggi nilai rata-ratanya maka semakin baik (Ahmad, 2018).

Silhouette digunakan untuk melihat kualitas dan kekuatan klaster seberapa baik suatu objek ditempatkan dalam suatu klaster. Metode ini merupakan gabungan dari metode *cohesion* dan *separation*. Tahapan perhitungan *silhouette* adalah sebagai berikut :

1. Hitunglah rata-rata dari suatu dokumen misalkan i dengan semua dokumen lain yang berada dalam satu klaster (Al-Zoubi, 2018)

$$a(i) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.5)$$

dengan j adalah dokumen lain dalam suatu klaster A dan $d(i, j)$ adalah jarak antara dokumen i dengan j .

2. Hitunglah rata-rata dari dokumen i tersebut dengan semua dokumen di klaster lain, dan diambil nilai terkecilnya (Al-Zoubi, 2018)

$$d(i, C) = \frac{1}{|A|-1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.6)$$

dengan $d(i, C)$ adalah jarak-rata-rata dokumen i dengan semua objek pada kluster lain C dimana $A \neq C$ (Han, 2012)

$$b(i) = \min_{C \neq A} d(i, C) \quad (2.7)$$

3. Nilai *sillhoutte* nya adalah :

$$s(i) = \frac{b(i) - a(i)}{\max(b(i), a(i))} \quad (2.8)$$

2.14 *K-Medoids Cluster*

K-Medoids Clustering juga dikenal sebagai *Partitioning Around Medoids (PAM)* adalah varian dari metode *k-means*. Hal ini didasarkan pada penggunaan *medoids* bukan dari pengamatan *mean* yang dimiliki oleh setiap kluster, dengan tujuan mengurangi sensitivitas dari partisi sehubungan dengan nilai ekstrim yang ada dalam dataset (Vercellis, 2009). *K-Medoids Clustering* hadir untuk mengatasi kelemahan *k-means clustering* yang sensitif terhadap *outlier* karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data (Han dan Kamber, 2006).

K-Medoids Clustering menggunakan metode pengklasteran partisi untuk mengklasterkan sekumpulan n objek menjadi sejumlah k kluster. Algoritma ini menggunakan objek pada kumpulan objek yang mewakili sebuah kluster. Objek yang mewakili sebuah kluster disebut dengan *medoids*. Kluster dibangun dengan menghitung kedekatan yang dimiliki antara *medoids* dengan objek *non medoids* (Setyawati, 2017). Kelebihan dari metode *k-medoids* adalah dapat mengelompokkan data dengan pencilan atau tanpa pencilan dan *k-medoids* dapat digunakan untuk data dengan variabel dominan kontinu maupun diskrit.

Menurut Han dan Kamber (2006), tahapan *k-medoids clustering* adalah sebagai berikut.

- a. Secara acak pilih k objek pada sekumpulan n objek sebagai *medoids*.
- b. Ulangi.
- c. Tempatkan objek non *medoids* ke dalam klaster yang paling dekat dengan *medoids*.
- d. Secara acak pilih O_{random} (sebuah objek non *medoids*).
- e. Hitung total *cost*, S , dari pertukaran *medoids* O_j dengan O_{random} .

$$S = \text{total cost baru} - \text{total cost lama} \quad (2.15)$$

Dengan :

S

: Selisih

Total *cost* baru: Total *cost* baru untuk non *medoids*

Total *cost* lama : Total *cost* lama untuk *medoids*

- f. Jika $S < 0$ maka tukar O_j dengan O_{random} , untuk membentuk sekumpulan k objek baru sebagai *medoids*.
- g. Hingga tidak ada perubahan.

2.15 Perbandingan Model Terbaik

Untuk mengetahui metode mana yang mempunyai kinerja terbaik, dapat digunakan rata-rata simpangan baku dalam klaster (σ_w) dan simpangan baku antar klaster (σ_b) (Bunkers, dkk, 1996).

Rumus rata-rata simpangan baku dalam klaster yaitu :

$$(\sigma_w) = K^{-1} \sum_{k=1}^k \sigma_k \quad (2.16)$$

dengan :

K : Banyaknya kluster yang terbentuk

σ_k : Simpangan baku kluster ke- k

Rumus simpangan baku kluster ke- k σ_k yaitu:

$$\sigma_k = \sqrt{\frac{1}{N} \sum_{i=1}^N (X_i - \mu_k)^2} \quad (2.17)$$

dengan :

N : Jumlah anggota dari setiap kluster

μ_k : Rata-rata kluster ke- k

X_i : Anggota kluster dari $i=1,2,\dots,N$

Rumus simpangan baku antar kluster σ_B yaitu :

$$\sigma_B = \left[(K)^{-1} \sum_{k=1}^K (\mu_k - \mu)^2 \right]^{1/2} \quad (2.18)$$

dengan :

μ_k : Rata-rata kluster ke- k

μ : Rata-rata keseluruhan kluster

Rumus rasio simpangan baku kluster σ yaitu :

$$\sigma = \frac{\sigma_w}{\sigma_B} \times 100\% \quad (2.19)$$

dengan :

σ_w : simpangan baku dalam kluster

σ_B : simpangan baku antar kluster

Metode yang mempunyai rasio terkecil merupakan metode terbaik. Klaster yang baik adalah klaster yang mempunyai homogenitas (kesamaan) yang tinggi antar anggota dalam satu klaster (*within cluster*) dan heterogenitas yang tinggi antar klaster yang satu dengan klaster yang lain (*between cluster*) (Santoso, 2007)



