

BAB II

LANDASAN TEORI

2.1 Analisis Multivariat

Secara umum, analisis multivariat atau metode multivariat berhubungan dengan metode-metode statistika yang secara bersama-sama melakukan analisis terhadap lebih dari dua variabel pada setiap objek atau orang. Variabel di dalam analisis multivariat dibagi menjadi dua, yaitu variabel independen dan variabel dependen. Variabel independen adalah variabel yang digunakan untuk mengestimasi atau memprediksi nilai variabel lain yaitu variabel dependen. Sedangkan variabel dependen adalah variabel yang nilainya ditentukan oleh variabel lain yaitu variabel independen (Santoso, 2018).

Data variabel independen maupun dependen di dalam analisis multivariat dikelompokkan menjadi dua jenis data yaitu kuantitatif dan kualitatif. Variabel kuantitatif adalah variabel yang disajikan dalam bentuk angka atau metrik. Contoh variabel kuantitatif seperti harga mobil, keuntungan perusahaan, harga saham, berat badan seseorang, dan lain sebagainya. Sedangkan variabel kualitatif adalah variabel yang disajikan dengan tidak dalam bentuk angka atau nonmetrik. Data variabel kualitatif dinyatakan dalam bentuk atribut atau merupakan karakteristik dari variabel. Contohnya seperti jenis kelamin dan tingkat pendidikan (Widarjono, 2002).

Jika data multivariat bersifat independen, metode analisis yang digunakan adalah analisis faktor, analisis *cluster*, *Multidimensional Scaling* (MDS) dan *Categorical Analysis* (CA). Sedangkan data multivariat yang bersifat dependen, metode yang digunakan adalah regresi berganda, regresi logistik, analisis deskriminan, SEM, MANOVA, dan korelasi kanonikal (Santoso, 2018).

2.2 Analisis Cluster

Analisis *cluster* merupakan suatu teknik yang digunakan untuk mengelompokkan objek atau kasus ke dalam kelompok yang relatif homogen yang disebut *cluster*. Objek dalam setiap kelompok cenderung mirip satu sama

lain dan berbeda jauh dengan objek dari *cluster* lainnya. Analisis *cluster* meneliti seluruh hubungan interdependensi, tidak ada perbedaan variabel independen dan variabel dependen (Supranto, 2008).

Analisis *cluster* membagi sejumlah data pada satu atau beberapa *cluster* tertentu. Sebuah *cluster* yang baik adalah *cluster* yang memiliki (Prasetyo, 2012):

1. Homogenitas (kesamaan) yang tinggi antar-anggota dalam satu *cluster*.
2. Heterogenitas (perbedaan) yang tinggi antar-*cluster* yang satu dengan *cluster* yang lainnya.

Dari dua hal diatas, dapat disimpulkan bahwa sebuah *cluster* yang baik adalah *cluster* yang mempunyai anggota-anggota yang semirip mungkin satu dengan yang lain (Santoso, 2018).

Cluster pada dasarnya mencari dan mengelompokkan data yang mirip satu dengan yang lain, maka kriteria mirip adalah dasar dari metode *cluster*. Proses pengelolaan data sehingga sekumpulan data mentah dapat dikelompokkan menjadi satu atau beberapa *cluster* adalah sebagai berikut (Simamora, 2005):

1. Menetapkan ukuran jarak antar-objek

Mengukur kesamaan antar-objek sesuai dengan prinsip dasar *cluster* yang mengelompokkan objek yang mempunyai kemiripan, maka proses pertama adalah mengukur seberapa jauh ada kesamaan antar-objek. Salah satu metode yang dapat digunakan yaitu mengukur jarak (*distance*) antara dua objek. Terdapat tiga ukuran kedekatan untuk mengukur kesamaan jarak antara pasangan objek, yaitu:

- 1) Jarak *Euclidean* mengukur jumlah kuadrat perbedaan nilai masing-masing variabel. Dirumuskan sebagai berikut:

$$d_{ij} = \sqrt{\sum_{l=1}^p (x_{il} - x_{jl})^2} \quad (2.1)$$

- 2) Jarak *Squared Euclidean* yang merupakan variasi dari jarak *Euclidean*. Jika pada jarak *Euclidean* diakarkan, maka pada jarak *Square Euclidean* tidak diakarkan. Dirumuskan sebagai berikut:

$$d_{ij} = \sum_{l=1}^p (x_{il} - x_{jl})^2 \quad (2.2)$$

- 3) Jarak *Cityblock* atau yang biasa disebut dengan jarak *Manhattan*. Pada jarak *Cityblock*, jarak antara dua objek merupakan jumlah perbedaan mutlak di dalam nilai untuk setiap variabel. Dirumuskan sebagai berikut:

$$d_{ij} = \sum_{l=1}^p |x_{il} - x_{jl}| \quad (2.3)$$

dimana,

- d_{ij} : Jarak antar objek ke- i dengan objek ke- j
- p : Banyak variabel
- x_{il} : Nilai objek ke- i pada variabel ke- l
- x_{jl} : Nilai objek ke- j pada variabel ke- l
- l : $1, 2, \dots, p$

2. Melakukan proses standarisasi data jika diperlukan.

Setelah cara mengukur jarak ditetapkan, yang perlu diperhatikan adalah apakah satuan data mempunyai perbedaan yang besar. Sebagai contoh, jika variabel penghasilan mempunyai satuan puluhan juta dan usia hanya memiliki satuan puluhan, maka perbedaan yang mencolok ini akan membuat perhitungan jarak menjadi tidak valid.

3. Melakukan proses *cluster*.

Setelah data yang dianggap mempunyai satuan yang sangat berbeda diseragamkan, langkah selanjutnya membuat *cluster*. Proses inti dari *cluster* adalah pengelompokan data, yang bisa dilakukan dengan dua metode:

- 1) Metode hirarki. 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

hingga *cluster* akan membentuk semacam pohon dimana ada tingkatan yang jelas antar-objek, dari yang paling mirip sampai paling tidak mirip.

- 2) Metode non-hirarki. Dalam metode ini justru dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan. Setelah jumlah *cluster* diketahui, baru proses *cluster* dilakukan tanpa mengikuti tingkatan.
4. Melakukan penamaan *cluster-cluster* yang terbentuk.
Melakukan interpretasi terhadap *cluster* yang telah terbentuk, yang pada intinya memberikan nama spesifik untuk menggambarkan isi *cluster* tersebut.
5. Melakukan validasi.

Cluster yang terbentuk kemudian dilakukan validasi apakah hasil tersebut valid.

2.3 Asumsi Analisis Cluster

Analisis *cluster* menetapkan adanya asumsi-asumsi yang harus dipenuhi. Asumsi dalam analisis *cluster* yaitu sebagai berikut (Santoso, 2018):

1. Sampel yang diambil harus benar-benar dapat mewakili populasi.
2. Tidak terjadi multikolinearitas (Non-Multikolinearitas).

Multikolinearitas adalah adanya hubungan linear atau korelasi yang tinggi, keberadaan hubungan linear yang sempurna atau tepat, di antara sebagian atau seluruh variabel independen dalam sebuah model regresi. Multikolinearitas biasanya terjadi ketika sebagian besar variabel yang digunakan saling terkait dalam suatu model regresi. Multikolinearitas disebabkan oleh beberapa faktor seperti metode pengumpulan data yang digunakan, batasan yang ada pada model atau populasi yang diambil sampelnya, spesifikasi model, dan model yang “*overdetermined*”. Untuk mengetahui ada tidaknya multikolinearitas, dapat dilihat dari nilai-nilai koefisien korelasi *pearson*. Koefisien korelasi *pearson* digunakan untuk mengukur ada tidaknya hubungan antar satu variabel terhadap variabel lainnya. Dirumuskan sebagai berikut:

$$r_{XY} = \frac{n \sum_{i=1}^n X_i - (\sum_{i=1}^n X_i)(\sum_{i=1}^n Y_i)}{\sqrt{\{n \sum_{i=1}^n X_i^2 - (\sum_{i=1}^n X_i)^2\}\{n \sum_{i=1}^n Y_i^2 - (\sum_{i=1}^n Y_i)^2\}}} \quad (2.4)$$

dimana,

$$\begin{aligned}
 n &= \text{Banyak pasangan data} \\
 X_i &= \text{variabel } X \text{ ke } i \\
 Y_i &= \text{variabel } Y \text{ ke } i \\
 X_i^2 &= \text{Kuadrat dari variabel } X \text{ ke } i \\
 Y_i^2 &= \text{Kuadrat dari variabel } Y \text{ ke } i \\
 i &= 1, 2, 3, \dots, n
 \end{aligned}$$

Dua variabel dikatakan multikolinearitas apabila nilai korelasinya lebih dari 0,80. Cara untuk menghilangkan multikolinearitas yaitu dengan menghilangkan salah satu variabel yang berkorelasi, melakukan transformasi variabel, dan penambahan data (Nugroho,2008).

2.4 Metode Hirarki

Metode hirarki adalah metode suatu model pada analisis *cluster* yang membentuk tingkat tertentu seperti pada struktur pohon karena proses *cluster* dilakukan secara bertahap dan bertingkat. Hasil dari metode hirarki dapat disajikan dalam bentuk dendogram. Dendogram adalah representasi visual dari langkah – langkah dalam analisis *cluster* yang menunjukkan bagaimana *cluster* terbentuk dan nilai koefisien jarak ada setiap langkah (Simamora, 2005).

Metode-metode yang digunakan dalam metode hirarki adalah metode *agglomeratif* dan metode *divisif*. Metode *agglomeratif* dimulai dengan menganggap bahwa setiap objek adalah sebuah *cluster*. Kemudian dua objek dengan jarak terdekat digabungkan menjadi satu *cluster*. Selanjutnya objek ketiga akan bergabung dengan *cluster* yang ada atau bersama objek lain dan membentuk *cluster* baru dengan tetap memperhitungkan jarak kedekatan antar objek. Proses akan berlanjut hingga akhirnya terbentuk satu *cluster* yang terdiri dari keseluruhan objek. Ada empat metode *agglomeratif* dalam pembentukan *cluster*, yaitu metode *single linkage*, metode *complete linkage*, metode *average linkage*, dan metode *Ward* (Simamora, 2005).

Proses dalam *divisif* berkebalikan dari metode *agglomeratif*. Metode ini dimulai dengan satu *cluster* besar yang mencakup semua objek pengamatan.

Selanjutnya, secara bertahap objek yang mempunyai ketidakmiripan cukup besar dipisahkan ke dalam *cluster* yang berbeda. Proses dilakukan sehingga terbentuk sejumlah *cluster* yang diinginkan (Simamora, 2005).

2.4.1 Metode *Single Linkage*

Metode ini mengelompokkan dua objek yang mempunyai jarak terdekat lebih dahulu. Metode *single linkage* atau disebut juga dengan metode pautan tunggal, jarak antara dua *cluster* dapat ditentukan dari dua objek berpasangan yang memiliki kesamaan atau memiliki jarak terdekat dalam *cluster* yang berbeda.

Hal ini berlaku jika pengelompokkan itu mempunyai makna atau tujuan yang jelas, terdapat kumpulan-kumpulan objek yang bersama-sama membentuk *cluster* serta kecenderungan dari hasil pengelompokkan itu menggambarkan rantai “*chain*” yang panjang (Mattjik dan Sumertajaya, 2011).

Pertama kali yang harus dilakukan untuk mendapatkan pautan tunggal ini adalah harus menemukan jarak terdekat antar *cluster* tersebut. Kemudian menggabungkan objek-objek yang sesuai, misalkan objek tersebut dilambangkan dengan *cluster* U dan *cluster* V untuk mendapatkan *cluster* gabungan (UV) (Mattjik dan Sumertajaya, 2011). Dirumuskan sebagai berikut:

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

2.4.2 Metode *Complete Linkage*

Pada metode *complete linkage*, jarak antara satu *cluster* dan *cluster* lain diukur berdasarkan objek yang mempunyai jarak terjauh (Simamora, 2005). Jarak antar *cluster* pada metode *complete linkage* atau disebut juga metode pautan lengkap, ditentukan dari jarak terjauh antara dua objek pada *cluster* yang berbeda. Metode ini dapat digunakan dengan baik untuk kasus dimana objek yang ada berasal dari kelompok yang benar-benar berbeda (Mattjik dan Sumertajaya, 2011).

Langkah pertama yang harus dilakukan adalah sama seperti kasus *single linkage* yaitu harus menemukan jarak terdekat antar *cluster* tersebut. Kemudian menggabungkan objek yang sesuai, misalkan objek tersebut dilambangkan dengan

cluster U dan *cluster* V untuk mendapatkan *cluster* gabungan (UV). Dirumuskan sebagai berikut (Mattjik dan Sumertajaya, 2011):

$$d_{UV(W)} = \max\{d_{UW}, d_{VW}\} \quad (2.6)$$

2.4.3 Metode *Average Linkage*

Average linkage merupakan variasi dari algoritma *single linkage* dan *complete linkage* yaitu menghitung jarak antara dua *cluster* yang disebut sebagai jarak rata-rata dimana jarak tersebut dihitung pada masing-masing custer dengan meminimumkan rata-rata jarak antara pasangan *cluster* yang digabungkan.

Langkah-langkah pengelompokkan menggunakan metode *average linkage* adalah sebagai berikut (Simamora, 2005):

1. Dimulai dengan n *cluster*, dimana masing-masing memuat satu kesatuan. Jika terdapat matriks $n \times n$ dengan jarak $D = \{d_{(ij)}\}$.
2. Mencari matriks jarak untuk pasangan *cluster* terdekat. Misalkan dipilih pasangan *cluster* paling mirip yaitu objek U dan V maka $D = \{d_{(UV)}\}$.
3. Menggabungkan *cluster* U dan V menjadi sebuah *cluster* yang baru (UV) dengan cara sebagai berikut:
 - 1) Menghapus baris dan kolom sesuai dengan *cluster* yang mirip yaitu *cluster* U dan V.
 - 2) Menambahkan baris dan kolom baru dengan memberikan nilai jarak antara *cluster* (UV) dengan menggunakan rumus sebagai berikut:

$$d_{(UV)W} = \text{average}(d_{UW}, d_{VW}) \quad (2.7)$$

4. Diulangi langkah (2) dan (3) sebanyak $n - 1$ kali, dimana n adalah jumlah objek.

2.4.4 Metode *Ward*

Metode *Ward* merupakan salah satu metode hirarki dalam analisis *cluster*. Metode *ward* bertujuan untuk memperoleh *cluster* yang memiliki varians dalam *cluster* sekecil mungkin. Ukuran kesamaan yang digunakan adalah jarak *square euclidean* (Supranto, 2004). Metode *ward* mengelompokkan objek didasarkan

pada kenaikan *Sum Square Error* (SSE). Pada tiap tahap, dua *cluster* yang memiliki kenaikan SSE paling kecil digabungkan (Simamora, 2005).

Metode *ward* merupakan suatu metode pembentukan *cluster* yang didasari oleh hilangnya informasi akibat penggabungan objek menjadi *cluster*. Hal ini diukur dengan menggunakan jumlah total dari deviasi kuadrat pada mean untuk setiap pengamatan. *Sum Square Error* (SSE) digunakan sebagai fungsi objektif. Dua objek akan digunakan jika mempunyai fungsi objektif terkecil di antara kemungkinan yang ada (Simamora, 2005).

2.5 Metode Non-hirarki

Metode non-hirarki dimulai dengan menentukan terlebih dahulu jumlah *cluster* yang diinginkan dan *centroid* di tiap *cluster*. Pada beberapa *software*, *centroid* yang digunakan adalah k pengamatan pertama namun ada juga *software* yang menentukan *centroid* secara acak. Pada langkah selanjutnya, jarak antara tiap objek dengan tiap *centroid* dihitung, dan tiap objek dimasukkan ke suatu *cluster* berdasarkan jarak terdekat dengan *centroid cluster*. Kemudian, tiap *centroid* yang terbentuk dihitung kembali. Begitu seterusnya hingga tidak ada lagi pemindahan objek dalam *cluster*. Metode ini biasa disebut *K-Means* (Nugroho, 2008).

Metode *K-Means* pertama kali dipopulerkan oleh Hartigan pada tahun 1975. Penggunaan metode tersebut untuk menjelaskan algoritma dalam penentuan suatu objek ke dalam *cluster* tertentu berdasarkan rataan terdekat.

2.6 Indeks Pembangunan Manusia

Konsep pembangunan manusia pertama kali diperkenalkan oleh *United Nations Development Programme* (UNDP) pada tahun 1990 melalui laporan yang berjudul *Human Development Report* (HDR). Munculnya gagasan tentang pembangunan manusia yang dikemukakan oleh UNDP tersebut bukan hanya menempatkan manusia sebagai *input* dari pembangunan tetapi juga menjadi tujuan akhir pembangunan, yaitu menciptakan lingkungan yang memungkinkan bagi masyarakat untuk produktif sehingga dapat memiliki umur panjang dan hidup sehat, menguasai pengetahuan, dan memenuhi standar hidup yang layak. HDR

juga memberikan rekomendasi pengukuran pembangunan manusia yang disebut sebagai IPM. Sejak tahun 2014, Indonesia mengimplementasikan metode perhitungan IPM terbaru untuk menghitung IPM yang disajikan secara periodik setiap tahun pada tingkat nasional, provinsi, dan kabupaten/kota (Badan Pusat Statistik, 2021).

UNDP menetapkan tiga dimensi pembentukan IPM. Ketiga dimensi ini merupakan pendekatan yang dipilih dalam penggambaran kualitas hidup manusia dan tidak mengalami perubahan hingga saat ini. Dimensi tersebut mencakup (Badan Pusat Statistik, 2021):

1. Umur panjang dan hidup sehat
2. Pengetahuan
3. Standar hidup layak.

2.6.1 Dimensi Umur Panjang Dan Hidup Sehat

Dimensi umur panjang dan hidup sehat diwakili oleh indikator Umur Harapan Hidup (UHH) saat lahir. Pentingnya umur harapan hidup terletak pada kepercayaan umum bahwa umur panjang merupakan hal yang berharga dan kenyataan bahwa terdapat berbagai faktor yang secara tidak langsung berkaitan erat dengan umur harapan hidup, seperti nutrisi yang cukup dan kesehatan yang baik (Badan Pusat Statistik, 2021).

UHH merupakan rata-rata perkiraan lamanya waktu yang dapat dijalani oleh seseorang selama hidupnya. Perhitungan UHH dilakukan melalui pendekatan tidak langsung. Data yang digunakan adalah Anak Lahir Hidup (ALH) dan Anak Masih Hidup (AMH) yang bersumber dari hasil sensus penduduk. Metode estimasi yang digunakan adalah *Trussel* dengan model *West*, yang sesuai dengan sejarah kependudukan dan kondisi Indonesia dan negara-negara Asia Tenggara umumnya (Badan Pusat Statistik, 2021).

UHH yang mewakili dimensi umur panjang dan hidup sehat di Indonesia secara konsisten terus meningkat dari tahun ke tahun. Hal ini menunjukkan bahwa bayi yang baru lahir mempunyai harapan untuk dapat menjalani hidup lebih panjang. Secara tidak langsung, hal ini juga mencerminkan semakin baiknya

derajat kesehatan masyarakat di Indonesia. Derajat kesehatan dapat diukur dari angka kematian (mortalitas) dan angka kehidupan (morbidity). Pada tahun 2020 capaian UHH meningkat sebesar 0,18 persen, walaupun terjadi peningkatan tetapi terjadi sedikit mengalami perlambatan dibandingkan dengan tiga tahun sebelumnya yang masing-masing tumbuh sebesar 0,23 persen, 0,20 persen, dan 0,20 persen (Badan Pusat Statistik, 2021).

Berdasarkan teori Henrik L. Blum terdapat empat faktor yang dapat mempengaruhi morbidity dan mortalitas penduduk yaitu faktor lingkungan, perilaku kesehatan, pelayanan kesehatan, dan keturunan. Keempat faktor tersebut saling terkait dan berinteraksi dengan faktor lingkungan dan perilaku kesehatan yang paling besar pengaruhnya terhadap derajat kesehatan manusia (Badan Pusat Statistik, 2021).

2.6.2 Dimensi Pengetahuan

Dimensi pengetahuan dalam perhitungan IPM terdiri dari dua indikator, yaitu Harapan Lama Sekolah (HLS) dan Rata-rata Lama Sekolah (RLS). HLS adalah perkiraan lamanya sekolah yang akan dijalani oleh anak yang berumur tujuh tahun. HLS dihitung berdasarkan penduduk usia tujuh tahun ke atas agar konsisten dengan referensi umur pada program wajib belajar yang direncanakan oleh pemerintah. Sementara itu, RLS adalah rata-rata lamanya waktu yang digunakan oleh penduduk dalam menjalani pendidikan formal. Cakupan penduduk untuk menghitung RLS adalah penduduk berumur dua puluh lima tahun ke atas, dengan asumsi bahwa proses pendidikan seseorang sudah berakhir sebelum mencapai usia dua puluh lima tahun (Badan Pusat Statistik, 2021).

Selama tahun 2010 hingga 2020, kedua indikator selalu mengalami peningkatan. Meskipun terus mengalami peningkatan tetapi sejak tahun 2014 mengalami perlambatan. Pada tahun 2020 hanya meningkat 0,23 persen, sedikit lebih lambat dibandingkan dengan peningkatan sebelumnya sebesar 0,31 persen. Peningkatan pada tahun 2020 dipengaruhi oleh peningkatan Angka Partisipasi Sekolah (APS) di seluruh kelompok umur pendidikan. Indikator Rata-rata Lama Sekolah (RLS) di masa pandemi juga meningkat dibandingkan tahun sebelumnya

tetapi juga masih mengalami perlambatan hanya meningkat sebesar 1,68 persen dari peningkatan tahun sebelumnya sebesar 2,08 persen. Peningkatan IRL juga dipengaruhi oleh APS (Badan Pusat Statistik, 2021).

2.6.3 Dimensi Standar Hidup Layak

Dimensi standar hidup layak dicerminkan oleh indikator pengeluaran per rill per kapita per tahun yang disesuaikan. Indikator ini menggambarkan kemampuan daya beli masyarakat selama periode tertentu. Selama tahun 2010-2019, indikator ini selalu tumbuh dengan rata-rata pertumbuhan sebesar 2,02 persen per tahun, tetapi pada periode 2010-2020 turun menjadi 1,73 persen per tahun akibat dampak Covid-19. Pandemi Covid-19 menyebabkan penurunan aktivitas ekonomi yang mengakibatkan penurunan penyerapan tenaga kerja dan peningkatan Tingkat Pengangguran Terbuka (TPT). Sehingga berdampak terjadinya meningkatnya kemiskinan (Badan Pusat Statistik, 2021).