

BAB III

ANALISIS DAN PERANCANGAN

3.1. Pengolahan Data pada Metode *K-Means*

Pengolahan data pada metode *K-Means* dimulai dengan tahap pra-pemrosesan, di mana data yang dikumpulkan dibersihkan dan diubah ke dalam bentuk numerik agar dapat dihitung secara matematis. Setelah itu, data dinormalisasi untuk memastikan semua atribut berada dalam skala yang sebanding. Selanjutnya, algoritma *K-Means* akan mengelompokkan data ke dalam sejumlah kluster berdasarkan jarak antar titik data terhadap pusat kluster (centroid) yang dipilih secara acak pada awalnya. Proses ini berjalan secara iteratif, di mana posisi centroid akan diperbarui hingga tercapai kondisi stabil, yaitu ketika tidak ada lagi perubahan signifikan dalam pembentukan kluster. Hasil akhirnya adalah pembagian data ke dalam kelompok-kelompok yang memiliki karakteristik serupa, yang dapat dianalisis lebih lanjut untuk pengambilan keputusan bisnis.

3.1.1. Pengumpulan Data

Pengumpulan data dalam penelitian ini dilakukan dengan cara mengakses dan merekam data transaksi penjualan pakaian yang terjadi di Toko Riski selama periode tertentu. Data yang dikumpulkan mencakup informasi penting seperti jumlah transaksi yang dilakukan oleh masing-masing pelanggan, total pembelian dalam satuan nilai rupiah, serta metode pembayaran yang digunakan. Proses pengumpulan dilakukan secara sistematis agar data yang diperoleh bersifat akurat

dan dapat merepresentasikan pola perilaku pelanggan. Data ini kemudian menjadi dasar utama untuk proses analisis menggunakan metode *K-Means Clustering*.

Tabel 3. 1. Data Sampel Penelitian

No	Nama Pelanggan	Frekuensi Belanja	Total Pengeluaran	Jenis Pembayaran	Jenis Produk Favorit
1	P1	Sering	Rendah	Transfer	Jaket
2	P2	Jarang	Rendah	Transfer	Sepatu
3	P3	Jarang	Tinggi	Transfer	Aksesoris
4	P4	Jarang	Tinggi	Transfer	Atasan
5	P5	Sering	Rendah	Tunai	Sepatu
6	P6	Jarang	Tinggi	Tunai	Jaket
7	P7	Sering	Tinggi	Tunai	Aksesoris
8	P8	Jarang	Tinggi	Transfer	Sepatu
9	P9	Jarang	Tinggi	Tunai	Celana
10	P10	Jarang	Tinggi	Tunai	Aksesoris
11	P11	Sering	Rendah	Tunai	Aksesoris
12	P12	Jarang	Tinggi	Tunai	Jaket
13	P13	Jarang	Rendah	Tunai	Aksesoris
14	P14	Sering	Tinggi	Transfer	Atasan
15	P15	Sering	Tinggi	Tunai	Sepatu
16	P16	Jarang	Tinggi	Tunai	Atasan
17	P17	Sering	Rendah	Tunai	Aksesoris
18	P18	Jarang	Tinggi	Transfer	Sepatu
19	P19	Sering	Rendah	Transfer	Aksesoris
20	P20	Sering	Tinggi	Transfer	Aksesoris

Tabel di atas menampilkan data pelanggan yang terdiri dari lima atribut utama, yaitu frekuensi belanja, total pengeluaran, jenis pembayaran, dan jenis produk favorit. Dari data tersebut terlihat adanya variasi perilaku konsumen, misalnya sebagian pelanggan berbelanja dengan frekuensi sering namun dengan pengeluaran yang masih tergolong rendah, sementara yang lain meskipun berbelanja jarang tetapi memiliki total pengeluaran tinggi. Metode pembayaran juga terbagi antara tunai dan transfer, yang menunjukkan adanya perbedaan

preferensi dalam melakukan transaksi. Selain itu, jenis produk favorit pelanggan cukup beragam, mulai dari jaket, sepatu, aksesoris, atasan, hingga celana, yang mengindikasikan bahwa kebutuhan serta minat konsumen tidak terfokus pada satu kategori produk saja, melainkan tersebar pada berbagai pilihan. Hal ini memberikan gambaran pola belanja pelanggan yang bisa dianalisis lebih lanjut untuk strategi pemasaran maupun pengelolaan produk.

3.1.2. Pembersihan Data

Pada tahapan pembersihan data, merupakan proses yang dilakukan untuk memilih dan menyiapkan data agar layak digunakan dalam analisis lebih lanjut. Namun, pada penelitian ini, pembersihan data difokuskan pada transformasi data dari bentuk kategorikal menjadi bentuk numerikal. Hal ini dilakukan agar data dapat diolah secara komputasional menggunakan metode klasifikasi, karena algoritma yang digunakan dalam penelitian ini memerlukan input dalam format numerik untuk dapat melakukan perhitungan dan pengambilan keputusan secara akurat. Untuk perubahan konversi data dalam bentuk kategorikal ke numerikal yaitu sebagai berikut.

Tabel 3. 2. Atribut Data Sampel

Atribut	Partisi	Nilai
Frekuensi Belanja	Sering	2
	Jarang	1
Total Pengeluaran	Tinggi	2
	Rendah	1
Jenis Pembayaran	Transfer	2
	Tunai	1
Jenis Produk Favorit	Jaket	1
	Sepatu	2
	Aksesoris	3
	Atasan	4
	Celana	5

Tabel nilai transformasi di atas menunjukkan proses pengkodean kategori menjadi angka agar data lebih mudah dianalisis secara kuantitatif. Setiap atribut memiliki partisi yang dikonversi ke dalam bentuk numerik, misalnya pada frekuensi belanja, pelanggan yang berbelanja sering diberi nilai 2 dan yang jarang diberi nilai 1. Hal serupa berlaku untuk total pengeluaran dengan nilai 2 untuk tinggi dan 1 untuk rendah, serta jenis pembayaran di mana transfer diberi nilai 2 dan tunai diberi nilai 1. Sementara itu, pada atribut jenis produk favorit terdapat lima kategori yang dipetakan secara berurutan, mulai dari jaket (1), sepatu (2), aksesoris (3), atasan (4), hingga celana (5). Dengan transformasi ini, data kategorikal dapat diolah lebih mudah dalam analisis statistik maupun algoritma *Data Mining*.

Setelah diperoleh untuk nilai konversi yang dilakukan, maka data akan langsung diubah menjadi data numerical yaitu sebagai berikut.

Tabel 3. 3. Data Sampel Konversi

No	Nama Pelanggan	Frekuensi Belanja	Total Pengeluaran	Jenis Pembayaran	Jenis Produk Favorit
1	P1	2	1	2	1
2	P2	1	1	2	2
3	P3	1	2	2	3
4	P4	1	2	2	4
5	P5	2	1	1	2
6	P6	1	2	1	1
7	P7	2	2	1	3
8	P8	1	2	2	2
9	P9	1	2	1	5
10	P10	1	2	1	3
11	P11	2	1	1	3
12	P12	1	2	1	1
13	P13	1	1	1	3
14	P14	2	2	2	4
15	P15	2	2	1	2
16	P16	1	2	1	4
17	P17	2	1	1	3
18	P18	1	2	2	2
19	P19	2	1	2	3
20	P20	2	2	2	3

Tabel di atas merupakan hasil transformasi data dari bentuk kategorikal menjadi bentuk numerik sesuai dengan partisi nilai yang telah ditetapkan sebelumnya. Setiap atribut seperti frekuensi belanja, total pengeluaran, jenis pembayaran, dan jenis produk favorit telah dikodekan menggunakan angka agar lebih mudah diolah dalam analisis statistik maupun algoritma *Data Mining*. Misalnya, frekuensi belanja sering diberi kode 2 dan jarang diberi kode 1, total pengeluaran tinggi diberi kode 2 dan rendah diberi kode 1, sementara jenis pembayaran transfer diberi kode 2 dan tunai diberi kode 1. Pada atribut produk favorit, kode diberikan secara berurutan mulai dari jaket (1), sepatu (2), aksesoris (3), atasan (4), hingga celana (5). Dengan transformasi ini, data pelanggan sudah berada dalam format numerik sehingga dapat langsung digunakan untuk analisis lebih lanjut seperti klasifikasi, *Clustering*, atau pemodelan prediktif.

3.1.3. Perhitungan Data

Untuk melakukan pengolahan data pada metode *K-Means* dapat menggunakan data sampel seperti dibawah ini, yaitu sebagai berikut. Tetapi untuk atribut pendapatan akan diubah data nya pada pengolahan,

Tabel 3. 4. Data Sampel Penelitian

No	Nama Pelanggan	Frekuensi Belanja	Total Pengeluaran	Jenis Pembayaran	Jenis Produk Favorit
1	P1	2	1	2	1
2	P2	1	1	2	2
3	P3	1	2	2	3
4	P4	1	2	2	4
5	P5	2	1	1	2
6	P6	1	2	1	1
7	P7	2	2	1	3
8	P8	1	2	2	2
9	P9	1	2	1	5
10	P10	1	2	1	3
11	P11	2	1	1	3

12	P12	1	2	1	1
13	P13	1	1	1	3
14	P14	2	2	2	4
15	P15	2	2	1	2
16	P16	1	2	1	4
17	P17	2	1	1	3
18	P18	1	2	2	2
19	P19	2	1	2	3
20	P20	2	2	2	3

Untuk tahapan awal melakukan perhitungan pada metode *K-Means Clustering* yaitu menghitung centroid dari 3 data yang ada pada data sampel. Untuk menghitung data nya juga, yang akan diambil awalnya yaitu data pertama dan kedua. Untuk proses perhitungan juga terdapat beberapa tahapan yaitu sebagai berikut.

$$\text{Centroid 1} = 2, 1, 2, 1$$

$$\text{Centroid 2} = 1, 1, 2, 2$$

$$\text{Centroid 3} = 1, 2, 2, 3$$

$$\begin{aligned} d_{1,1} &= \sqrt{(2-2)^2 + (1-1)^2 + (2-2)^2 + (1-1)^2} \\ &= \sqrt{(0)^2 + (0)^2 + (0)^2 + (0)^2} \\ &= 0 \end{aligned}$$

$$\begin{aligned} d_{1,2} &= \sqrt{(2-1)^2 + (1-1)^2 + (2-2)^2 + (1-1)^2} \\ &= \sqrt{(1)^2 + (0)^2 + (0)^2 + (0)^2} \\ &= 1.41 \end{aligned}$$

$$\begin{aligned} d_{1,3} &= \sqrt{(2-1)^2 + (1-2)^2 + (2-2)^2 + (1-3)^2} \\ &= \sqrt{(1)^2 + (-1)^2 + (0)^2 + (-2)^2} \\ &= 2.45 \end{aligned}$$

$$\begin{aligned} d_{2,1} &= \sqrt{(10-10)^2 + (15-5)^2 + (5-5)^2 + (5-5)^2} \\ &= \sqrt{(0)^2 + (10)^2 + (0)^2} \end{aligned}$$

$$= 1.41$$

$$\begin{aligned} d_{2,2} &= \sqrt{(1-1)^2 + (1-1)^2 + (2-2)^2 + (2-2)^2} \\ &= \sqrt{(0)^2 + (0)^2 + (0)^2 + (0)^2} \\ &= 0 \end{aligned}$$

$$\begin{aligned} d_{2,3} &= \sqrt{(1-1)^2 + (1-2)^2 + (2-2)^2 + (2-3)^2} \\ &= \sqrt{(0)^2 + (-1)^2 + (0)^2 + (-1)^2} \\ &= 1.41 \end{aligned}$$

$$\begin{aligned} d_{3,1} &= \sqrt{(1-2)^2 + (2-1)^2 + (2-2)^2 + (3-1)^2} \\ &= \sqrt{(-1)^2 + (1)^2 + (0)^2 + (2)^2} \\ &= 2.45 \end{aligned}$$

$$\begin{aligned} d_{1,2} &= \sqrt{(1-1)^2 + (2-1)^2 + (2-2)^2 + (3-1)^2} \\ &= \sqrt{(0)^2 + (1)^2 + (0)^2 + (2)^2} \\ &= 1.41 \end{aligned}$$

$$\begin{aligned} d_{1,3} &= \sqrt{(1-1)^2 + (2-2)^2 + (2-2)^2 + (3-3)^2} \\ &= \sqrt{(0)^2 + (0)^2 + (0)^2 + (0)^2} \\ &= 0 \end{aligned}$$

Hasil perhitungan jarak antar titik data dengan centroid di atas menunjukkan tahapan awal dari metode *K-Means Clustering* dalam analisis pola pembelian pelanggan di Toko Riski. Setiap data pelanggan dihitung jaraknya terhadap tiga centroid yang telah ditentukan, menggunakan rumus Euclidean Distance, sehingga diperoleh nilai seperti $d_{1,1} = 0$, $d_{1,2} = 1,41$, dan seterusnya. Nilai jarak ini menunjukkan seberapa dekat masing-masing pelanggan dengan centroid tertentu, yang nantinya menjadi dasar untuk mengelompokkan pelanggan ke dalam cluster yang paling sesuai. Misalnya, pelanggan dengan jarak terdekat ke Centroid 1 akan masuk ke Cluster 1, sedangkan yang dekat dengan Centroid 2 atau 3 akan masuk

ke Cluster 2 atau 3. Dengan pendekatan ini, Toko Riski dapat mengidentifikasi pola belanja pelanggan, seperti frekuensi pembelian, total pengeluaran, dan jenis produk favorit, sehingga strategi penjualan dan promosi dapat disesuaikan dengan karakteristik tiap cluster.

Hasil *Clustering*

Tabel 3. 5. Hasil Perhitungan di Excel

No	Nama Pelanggan	X1	X2	X3	Cluster
1	P1	0	1.41	2.45	1
2	P2	1.41	0	1.41	2
3	P3	2.45	1.41	0	3
4	P4	3.32	2.24	1	3
5	P5	1.41	1.41	2	1
6	P6	1.73	1.73	2.24	1
7	P7	2.45	2	1.41	3
8	P8	1.73	1	1	2
9	P9	4.36	3.32	2.24	3
10	P10	2.65	1.73	1	3
11	P11	2.24	1.73	1.73	2
12	P12	1.73	1.73	2.24	1
13	P13	2.45	1.41	1.41	2
14	P14	3.16	2.45	1.41	3
15	P15	1.73	1.73	1.73	1
16	P16	3.46	2.45	1.41	3
17	P17	2.24	1.73	1.73	2
18	P18	1.73	1	1	2
19	P19	2	1.41	1.41	2
20	P20	2.24	1.73	1	3

Hasil *Clustering* menunjukkan bahwa setiap pelanggan telah dikelompokkan ke dalam tiga cluster berdasarkan jarak Euclidean terhadap centroid yang telah ditentukan. Misalnya, pelanggan seperti Adit Anggraini, Ahmad Suhendra, Ahmad Wahyudi, Budi Cahyani, dan Citra Halim masuk ke Cluster 1, yang menunjukkan kedekatan karakteristik mereka terhadap centroid

pertama. Cluster ini umumnya terdiri dari pelanggan dengan kombinasi frekuensi belanja dan total pengeluaran yang relatif seimbang, serta preferensi produk yang konsisten, sehingga mencerminkan kelompok pelanggan yang mungkin sering berbelanja dengan pengeluaran menengah hingga rendah.

Sementara itu, Cluster 2 terdiri dari pelanggan seperti Adit Wijaya, Aldi Maulana, Dewi Permata, dan Dewi Salim, yang lebih dekat dengan centroid kedua. Dari analisis jarak (X_1 , X_2 , X_3), pelanggan dalam cluster ini cenderung memiliki perilaku belanja yang bervariasi namun memiliki kecenderungan pada metode pembayaran tertentu dan jumlah transaksi sedang, serta sebagian memiliki pengeluaran yang tidak terlalu tinggi. Hal ini menunjukkan Cluster 2 dapat diinterpretasikan sebagai pelanggan potensial yang perlu perhatian khusus untuk ditingkatkan frekuensi pembeliannya atau diarahkan pada promosi tertentu.

Sedangkan Cluster 3, yang mencakup pelanggan seperti Agung Wijaya, Ahmad Kurniawan, Aldi Lestari, dan Dewi Suharto, terdiri dari individu yang memiliki jarak terdekat dengan centroid ketiga, biasanya memiliki total pengeluaran tinggi atau frekuensi belanja lebih jarang namun dengan nilai transaksi besar. Cluster ini menggambarkan kelompok pelanggan premium atau high-value, yang kontribusinya terhadap penjualan lebih signifikan meskipun jumlah transaksinya tidak sebanyak Cluster 1. Dengan demikian, Toko Riski dapat memanfaatkan informasi ini untuk strategi pemasaran yang berbeda pada tiap cluster, misalnya memberikan promo eksklusif kepada Cluster 3, diskon rutin untuk Cluster 1, dan pendekatan peningkatan loyalitas pada Cluster 2