

BAB III

METODOLOGI PENELITIAN

3.1 Objek dan Subjek Penelitian

3.1.1 Objek Penelitian

Kata “objek penelitian” adalah topik kajian (Kamus Besar Bahasa Indonesia; 1989: 622). Menurut Supranto (2000:21), sasaran penelitian adalah sekumpulan orang, kelompok, atau benda yang akan diteliti. Sejalan dengan itu (Dajan, 1986: 21), tujuan penelitian adalah subjek yang akan diteliti guna mengumpulkan data yang lebih spesifik. Adapun Objek penelitian dalam penelitian ini yaitu pengelompokan *Corporate Customer* (pelanggan) di PT Telkom Divisi *Enterprise* segmen *Regional Banking Service*.

3.1.2 Subjek Penelitian

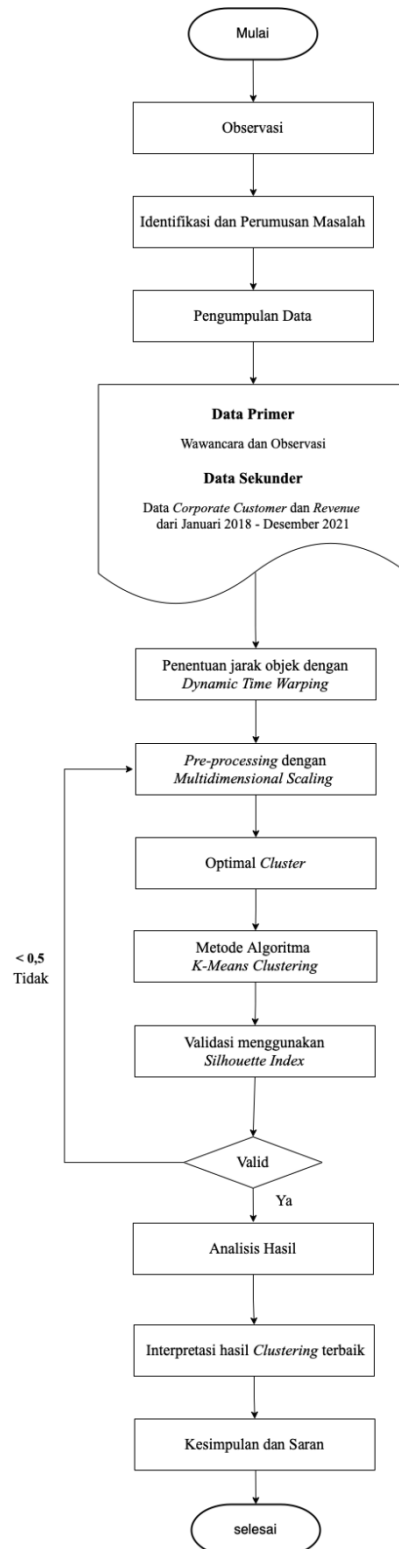
Yang dimaksud dengan 'subjek penelitian' adalah individu, lokasi, atau hal-hal yang diamati dalam rangka penelusuran sebagai tujuan (Kamus Besar Bahasa Indonesia, 1989: 862). Pelanggan Korporasi menjadi fokus kajian dalam penelitian ini (pelanggan) di PT Telkom Divisi *Enterprise* segmen *Regional Banking Service*

3.2 Alur Penelitian

Pada penelitian ini terdapat beberapa tahapan yang akan dilakukan. Dimulai dari melakukan observasi pada perusahaan lalu mengidentifikasi permasalahan yang terjadi pada objek penelitian. Tahap selanjutnya, dilakukan perumusan masalah untuk menentukan fokus penelitian yang dilakukan. Pada tinjauan pustaka meliputi kajian literatur penelitian sebelumnya dan dasar teori yang mendukung fokus penelitian.

Tahap selanjutnya, pengumpulan data dengan fokus penelitian. Data tersebut digunakan sebagai dasar dalam pengolahan data. Hasil kemudian akan dianalisis dan melakukan *clustering* dengan menggunakan tahapan proses perhitungan jarak objek pada data *time series* menggunakan *Dynamic Time Warping* lalu proses preprocessing penentuan proximitas dan konfigurasi dengan multidimensional scaling setelah itu dilanjutkan dengan proses menentukan *clustering* dengan memakai algoritma *K-means clustering* hingga pada akhir penelitian dapat

menjawab rumusan masalah pada penelitian yang dilakukan. Tahapan penelitian dapat dilihat pada Gambar 3.1 dibawah.



Gambar 3.1 Diagram Alur Penelitian

3.3 Teknik Pengumpulan Data

Berikut adalah sumber data dan metode pengumpulan penelitian ini,

1. Sumber primer

Data primer meliputi informasi yang diperoleh melalui pengalaman langsung, wawancara, dan observasi Segmen Layanan Perbankan Regional PT Telekomunikasi Indonesia. Wawancara tambahan dilakukan untuk memperoleh data primer untuk penelitian ini, yang digunakan untuk menentukan volume klien perbankan dan efisiensi prosedur bisnis.

2. Sumber kedua

Buku, data, catatan, bukti, dan arsip, baik yang diterbitkan maupun yang tidak diterbitkan, berfungsi sebagai media perantara untuk memperoleh akses ke sumber informasi sekunder. Analisis ini menggunakan kumpulan data sekunder termasuk Daftar Pelanggan Perusahaan (dari 2018 hingga 2021) untuk Layanan Perbankan Daerah, Data Pendapatan Daerah Tahunan (dari 2018 hingga 2021) untuk Populasi Proses Bisnis, dan Data Permintaan (dari September 2020). hingga Februari 2021) untuk Populasi Proses Bisnis.

Sejumlah teknik digunakan untuk mendapatkan informasi yang diperlukan untuk penelitian ini seperti,

a. Penelitian Lapangan (*Field Research*)

Peneliti melakukan kerja lapangan untuk mendapatkan data yang akurat. Teknik pengumpulan data berikut digunakan untuk penelitian ini:

b. Pengamatan

Pengamatan atau pengumpulan informasi melalui mengamati petunjuk tentang topik yang sedang dibahas.

c. Wawancara

Memperoleh informasi dengan meminta orang menjawab pertanyaan tentang masalah yang sedang dipelajari. Para ahli di Segmen Layanan Perbankan Regional PT Telekomunikasi Indonesia diwawancarai untuk penelitian ini untuk memberikan wawasan dan pendapat mereka.

d. Pencatatan

Teknik pengumpulan informasi dengan mengklasifikasikan dan

mengatur dokumen dan catatan tertulis yang relevan, seperti memo, surat, dan hasil wawancara.

e. Tinjauan literatur yang relevan.

Penelitian ini termasuk meninjau buku, artikel, dan laporan untuk mengumpulkan informasi. Penelitian perpustakaan membutuhkan perjalanan ke ruang baca.

3.4 Teknik Analisis Data

Pada tahap analisis data dan pembahasan disini akan menganalisa bagaimana memetakan dan mengidentifikasi karakteristik pelanggan di PT Telkom Divisi *Enterprise Service* segmen *Banking* bertujuan untuk menentukan metode yang sesuai untuk klusterisasi pelanggan sehingga mampu menghasilkan pemetaan *cluster* pelanggan yang optimal dan mampu merekomendasikan strategi untuk masing-masing *cluster* yang terbentuk. Data yang dipelajari kemudian ditafsirkan dalam perdebatan. Pembahasan disajikan dalam bentuk data yang telah diolah dan dievaluasi untuk mencapai pengelompokan dan pemetaan temuan *Corporate Customer* di PT Telekomunikasi Indonesia segmen *Regional Banking Service*.

Analisis dilakukan dengan membandingkan hasil penggerombolan data berdasarkan data *Revenue* (pendapatan) yang dihasilkan oleh *Corporate Customer* dari rentang waktu januari 2018 hingga desember 2021 menggunakan metode *Clustering Time series* dan memakai *software Python*.

1. Analisis *Cluster Time series*

Analisis kluster adalah teknik untuk menemukan kelompok dalam kumpulan data dengan tujuan agar data yang termasuk dalam satu kelompok memiliki kesamaan yang erat dan memiliki perbedaan yang jelas dari data yang dimiliki kelompok lain (Kaufman & Rousseuw, 1990). Pendekatan dan prosedur pengelompokan yang berbeda diperlukan untuk analisis kluster deret waktu dibandingkan dengan analisis kluster data statis. Data deret waktu adalah sekumpulan nilai observasi yang direkam pada berbagai periode waktu dengan interval waktu yang sama, dan urutan data tersebut seharusnya saling terhubung. Akibatnya, metode yang berbeda dijalankan. Dalam

perkembangan pengelompokan pada data deret waktu, beberapa strategi dibuat, diantaranya adalah penggunaan jarak pengelompokan yang sesuai dengan sifat-sifat data deret waktu (Liao, 2005).

a. Jarak *Euclid*

Jarak *Euclidean* merupakan tipe pengukuran jarak dalam analisis *cluster* untuk mengukur jarak dari obyek data ke pusat *cluster*. Jarak Euclid merupakan jarak geometris antar dua obyek data. Semakin dekat jarak maka semakin mirip suatu obyek data tersebut (Johnson & Wichern, 2002). *Euclidean Distance* merupakan jarak geometris antar dua objek data. *Euclidean Distance* dapat diperoleh dengan menggunakan Persamaan sebagai berikut,

$$d_{EUCL} = \sqrt{\sum_{t=1}^n (x_t - y_t)^2}$$

dengan:

d_{EUCL} = *Euclidean Distance*

x_t = data x pada waktu ke- t

y_t = data y pada waktu ke- t

b. *Dynamic Time Warping* (DTW)

Distance DTW merupakan metrik *dissimilarity* yang tidak membutuhkan model tertentu. Jarak DTW adalah jarak terpendek antara dua pasang titik yang memperhitungkan potensi pergerakan titik (Berndt & Clifford, 1994). Karena data bervariasi dan berkembang dari waktu ke waktu, algoritme pengelompokan tidak dapat mengukur kedekatan antar item menggunakan jarak *Euclidean* dasar. Jarak *Euclidean* cocok untuk data statis atau fungsi yang tidak berubah seiring waktu. Selama pengembangan matriks kedekatan jarak antar objek dalam data deret waktu, juga diperlukan untuk menghitung jarak antara deret waktu dinamis, memerlukan strategi pengelompokan yang

berbeda dari yang diperlukan untuk data statis (Adnyani & Sihombing, 2021). *Dynamic Time Warping* karenanya merupakan metode pengukuran alternatif yang dapat digunakan (DTW). Metrik ini memungkinkan perbandingan periode dengan durasi yang bervariasi.

c. Analisis *Multidimensional Scaling (MDS)*

Pemetaan persepsi multidimensi (peta perseptual) adalah tujuan dari analisis skala multidimensi. Data jarak yang digunakan dalam penskalaan adalah data interval atau rasio. Tidak penting bagi metrik Penskalaan Multidimensi apakah data input mewakili jarak sebenarnya atau tidak. Pendekatan ini membangun bentuk geometris berdasarkan susunan titik objek sedekat mungkin dengan jarak input.

d. Penentuan jumlah *centroid* awal yang optimal

Jumlah *centroid* awal ditentukan menggunakan metode *non-hierarchical clustering* menggunakan kriteria *Elbow* dan *Silhouette*. Metode *non-hierarchical clustering* adalah pendekatan penggabungan yang mengelompokkan setiap item ke dalam banyak *cluster*. Pendekatan non-hierarkis hanya digunakan untuk menghitung *centroid* awal, yang kemudian dinilai berdasarkan kriteria *Silhouette* dan *Elbow*. Menghitung kriteria siluet berdasarkan jarak keterpaduan dan pemisahan *cluster*. Rumus siluet setiap objek berdasarkan persamaan.

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

Di mana $a(i)$ mewakili jarak rata-rata antara objek- i dan semua objek lain dalam *cluster* yang sama. $b(i)$ adalah jarak rata-rata antara objek- i dan semua objek lain di *cluster* terdekat. Kriteria *Silhouette* ditentukan dengan rata-rata semua $s(i)$. *Centroid* awal yang dipilih berdasarkan kriteria *Silhouette* ditentukan oleh persamaan di bawah ini.

$$S = \max S(n, p)$$

Di mana $S(n, p)$ adalah hasil dari kriteria *Silhouette* diterapkan pada algoritme non-hierarkis p membuat n *cluster*. Romanuke juga

telah mengimplementasikan kriteria *Silhouette* untuk mengidentifikasi jumlah *cluster* yang optimal untuk memaksimalkan hasil *clustering*. Kriteria *Elbow* mencari K *cluster* yang ideal berdasarkan penambahan informasi terbesar, sehingga jika dimodifikasi menjadi $K+1$ *cluster*, penambahan informasi tidak begitu substansial. Masuknya data ini ditunjukkan dengan jarak objek ke masing-masing *centroid*, yang dihitung dengan menggunakan persamaan berikut.

$$SSE = \sum_{k=1}^K \sum_{\mathbf{x}_i \in S_k} \|\mathbf{x}_i - \mathbf{c}_k\|_2^2,$$

Dengan K mewakili jumlah *cluster*, S_k mewakili himpunan objek yang terdapat dalam *cluster* k , dan \mathbf{c}_k mewakili *centroid cluster* k . Pemilihan k *cluster* menggunakan kriteria *Elbow* didasarkan pada selisih nilai SSE terbesar antara pembangkitan k *cluster* dan $k-1$ *cluster*. *Centroid* awal yang dipilih berdasarkan kriteria *Elbow* ditentukan dengan persamaan berikut.

$$SSE = \max SSE(n, p),$$

Dimana $SSE(n, p)$ adalah hasil penerapan kriteria *Elbow* ke n *cluster* yang dihasilkan oleh metode p . Kriteria *Elbow* juga digunakan untuk mengidentifikasi jumlah *cluster* yang ideal dalam klasifikasi yang menggunakan segmentasi citra peta warna.

e. Validasi Clustering

Validasi diperlukan untuk menilai kualitas hasil pengelompokan. Menurut Rendon, dkk., validasi pengelompokan terdiri dari validasi eksternal dan internal. Karena kurangnya informasi *cluster* awal dalam data, validasi internal digunakan dalam penyelidikan ini. Validasi internal telah terbukti lebih unggul daripada validasi eksternal untuk memverifikasi temuan pengelompokan (Randon E, 2011). Indeks *Silhouette* berfungsi sebagai validasi internal dalam penelitian ini. Validasi ini dipilih karena kinerjanya lebih baik daripada validasi internal lainnya. Indeks *Silhouette* telah digunakan dalam studi

pertanian dan medis. Nilai *Silhouette* setiap item diturunkan dari $a(i)$, jarak rata-rata antara objek I dan semua objek lain dalam *cluster* yang sama. $b(i)$ adalah jarak rata-rata antara objek- i dan semua objek lain di *cluster* terdekat. Indeks siluet adalah rata-rata dari semua $s(i)$ yang nilainya berada dalam kisaran -1 hingga d-1. Semakin dekat nilai $s(i)$ dengan 1, semakin masuk akal untuk memasukkan item ke dalam *cluster* berdasarkan nilai $s(i)$. Item dengan $s(i)$ 0 memiliki tingkat keunikan relatif tinggi terhadap *cluster* saat ini; karenanya, disarankan agar objek membuat *cluster* sendiri. Berdasarkan nilai $s(i)$ ini, jika perhitungan menghasilkan $s(i)$ 0, ini menunjukkan bahwa item tersebut salah diklasifikasi atau termasuk dalam *cluster* yang salah. Pada investigasi ini, kesalahan klasifikasi setiap item juga akan dievaluasi berdasarkan nilai $s(i)$ pada temuan *clustering*.

Menurut Kaufman dan Rausseeuw (1990) interpretasi koefisien *Silhouette* dapat dinyatakan seperti pada Tabel 3.1

Tabel 3.1 Kategori Penggerombolan Berdasarkan Koefisien *Silhouette*

No	Koefisien <i>Silhouette</i>	Kategori Penggerombolan
1.	0.71-1.00	Strong Classification
2.	0.51-0.70	Good Classification
3.	0.26-0.50	Weak Classification
4.	0.00-0.25	Bad Classification

f. *K-Means Clustering*

Metode analisis *cluster K-means* digunakan untuk mengelompokkan lebih dari 200 contoh secara lebih efektif. Algoritma *K-Means* untuk pengelompokan pelanggan. Pengelompokan *K-Means* adalah pendekatan yang efektif untuk menangani masalah pengelompokan data berdasarkan pembelajaran mesin. Metode ini adalah salah satu algoritme tanpa pengawasan yang cocok untuk tantangan segmentasi konsumen. Pendekatan ini didasarkan pada penyortiran *centroid* terdekat, yang melibatkan *instance*

pengelompokan berdasarkan kedekatannya dengan pusat grup. Aspek penting dari strategi ini adalah,

1. Pendekatan ini membutuhkan pengguna untuk menentukan jumlah grup terlebih dahulu, dan analisis hierarkis dapat digunakan untuk mengidentifikasi jumlah grup.
2. Metode ini juga dapat digunakan untuk mengkategorikan data yang baru dikumpulkan ke dalam kelompok terdekat.

Dengan menggunakan *K-means*, metode berikut mengelompokkan data baru ke dalam *cluster*.

1. Dokumentasikan pengelompokan dan variabel terkaitnya. k menunjukkan jumlah *cluster* yang dibuat.
2. Tentukan *centroid* dari k *center cluster* pertama yang dipilih secara acak

Centroid awal dihasilkan secara acak dari objek yang tersedia di k *cluster* pertama, dan rumus berikut digunakan untuk mencari *centroid cluster* ke- i :

$$v = \frac{\sum_{i=1}^n x_i}{n} \quad ; i = 1, 2, 3, \dots, n$$

Dimana,

v : *centroid* pada *cluster*

x_i : objek ke- i

n : banyaknya objek/jumlah objek yang menjadi anggota *cluster*

3. Berikut ini dihitung nilai jarak setiap objek ke setiap *centroid* setiap *cluster* dengan menggunakan *Euclidean*, yang bertujuan untuk menghitung jarak setiap data di setiap grup, kemudian jarak yang berdekatan menunjukkan yang paling besar kesamaan

$$d(x, y) = \|x - y\| = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad ; i = 1, 2, 3, \dots, n$$

dengan kelompok lain, yang mengarah pada kesimpulan bahwa data milik kelompok itu..

Dimana;

x_i : objek x ke-i

y_i : daya y ke-i

n : banyaknya objek

4. Sebagai hasil dari dimasukkannya data baru, perhitungan mean dari *centroid* grup akan berubah, dan rata-rata juga akan berubah.
5. Mengalokasikan setiap item ke *centroid* terdekat Secara umum, *hard k-means* dapat digunakan untuk mengalokasikan objek ke dalam setiap *cluster* selama iterasi, di mana setiap objek ditetapkan sebagai anggota *cluster* berdasarkan kedekatannya dengan pusat *cluster*.
6. Lakukan iterasi, lalu temukan lokasi *centroid* baru.
7. Ulangi langkah 3 jika lokasi centroid yang baru tidak sama dengan *centroid* sebelumnya.

3.5 Jadwal Kegiatan

Penelitian akan dilaksanakan dalam waktu empat bulan terhitung mulai bulan oktober sampai dengan bulan Februari 2023. Jadwal penelitian meliputi pengumpulan data, pengolahan data, analisis hasil, pembuatan laporan dan pembuatan paper. Jadwal kegiatan maksimal 4 bulan dengan rincian sebagai berikut:

Tabel 3.2 Jadwal Kegiatan

Kegiatan	Bulan ke-1	Bulan ke-2	Bulan ke-3	Bulan ke-4
Pengumpulan Data	■			
Pengolahan Data		■	■	■
Analisis Hasil			■	■
Pembuatan Laporan			■	■
Pembuatan Paper			■	■