

BAB III

METODOLOGI PENELITIAN

3.1 Gambaran Umum Objek Penelitian

Sumber data yang digunakan berfokus pada data transaksi pelanggan retailer XYZ selama 2 tahun terakhir. Perusahaan bisnis ritel XYZ yang merupakan bisnis ritel yang menjual bahan sembako. Dalam setiap harinya terdapat kisaran pelanggan sejumlah 150 hingga 200 pelanggan yang melakukan transaksi dan setiap harinya perusahaan ini mengalami peningkatan, namun saat ini data transaksi berupa informasi waktu transaksi dan besaran transaksi yang dilakukan oleh pelanggan hanya disimpan sebagai arsip biasa. Informasi tentang waktu dan besaran transaksi yang dilakukan oleh pelanggan adalah variabel data transaksi yang digunakan. Untuk membandingkan analisis segmentasi pelanggan, algoritma *clustering K-means* dan *Fuzzy C-means* digunakan dalam *data mining*.

3.2 Metode Penelitian

Untuk membandingkan hasil pengolahan data, akan dibandingkan dua algoritma yaitu *K-means clustering* dan *Fuzzy C-means* pada penelitian ini, yang selanjutnya variabel data yang digunakan yaitu waktu transaksi serta besaran transaksi yang dilakukan pelanggan. Algoritma *K-means* dan *Fuzzy C-means* dipilih berdasarkan penelitian terdahulu, serta dipilih karena algoritma *K-means* itu algoritma interaktif dan mudah diinterpretasikan, diterapkan, serta bersifat dinamis pada data yang tersebar, sedangkan algoritma *Fuzzy C-means* dipilih karena pusat *cluster* yang ditempatkan lebih tepat jika membandingkannya dengan algoritma *cluster* lainnya.

Tabel 3.1 Perbandingan *K-Means* dan *Fuzzy C-Means*

Perbedaan	Algoritma	
	<i>K-means</i>	<i>Fuzzy C-means</i>
Pengelompokan terhadap data	Pengelompokkan data berdasarkan perhitungan	Pengelompokkan data derajat

Perbedaan	Algoritma	
	<i>K-means</i>	<i>Fuzzy C-means</i>
	jarak terdekat antara data ke <i>centroid</i> .	keanggotaan yang berbeda antara 0 hingga 1.
Ukuran Dataset	Cocok untuk dataset berukuran besar.	Cocok untuk dataset berukuran kecil
Penetapan K di awal proses	Penetapan jumlah K <i>cluster</i> diawal proses	Penetapan jumlah K <i>cluster</i> diawal proses
Proses	Algoritma <i>K-means clustering</i> adalah algoritma yang efisien yang dapat mengelompokkan banyak data dan waktu komputasi yang cukup cepat.	<i>Fuzzy C-means</i> merupakan teknik untuk menentukan <i>cluster</i> optimal dalam suatu ruang vektor.

Sumber: Sania Fitri Octavia dan Mustakim [41], Rizqina Rahmati dan Arie Wahyu Wijayanto [42]

3.3 Variabel Penelitian

Saat melakukan pengolahan data akan mempengaruhi hasil dari *clustering* data transaksi yang dipengaruhi oleh variabel yang ada dalam dataset. Selanjutnya, variabel informasi yang digunakan pada penelitian ini ialah waktu dan besaran transaksi yang dilakukan oleh pelanggan, selain itu, algoritma *clustering K-means* dan *fuzzy C-means clustering* membantu menentukan keakuratan hasil analisis. Variabel penelitian terbagi menjadi dua kategori yang berdasarkan konteks hubungan, yaitu variabel bebas (*independen*) dan variabel terikat (*dependent*).

- a) Variabel bebas atau *independent* merupakan variabel yang nilainya mempengaruhi variabel lainnya yaitu variabel terikat [43]. Berdasarkan data transaksi terhadap segmentasi pelanggan, maka transaksi pelanggan sebagai variabel bebas dikarenakan transaksi pelanggan dapat berubah-ubah.

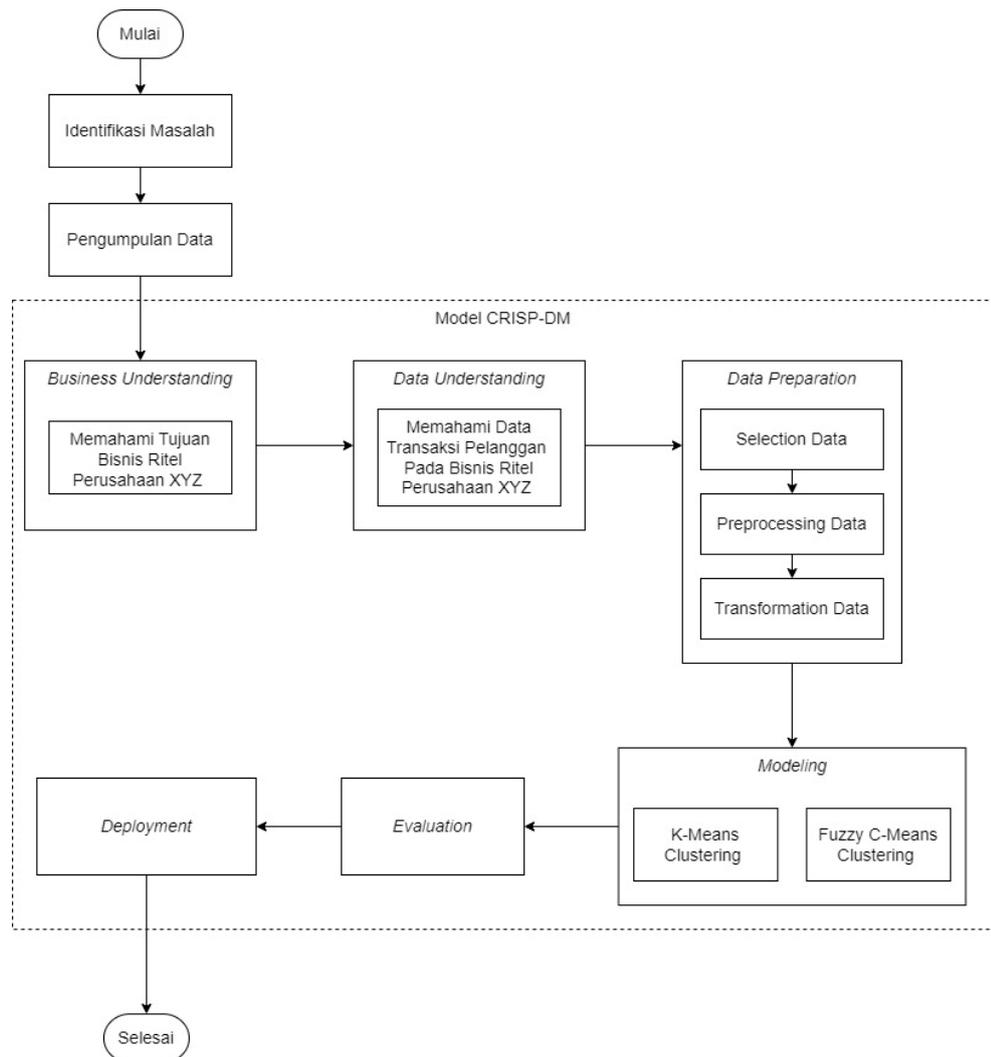
- b) Variabel terikat merupakan variabel yang nilainya tergantung dari nilai variabel lainnya [43]. Berdasarkan data transaksi terhadap segmentasi pelanggan, maka segmentasi pelanggan adalah variabel terikat, dikarenakan nilainya tergantung dari variabel bebas yaitu transaksi pelanggan.

3.4 Teknik Pengumpulan Data

Dalam tahap pengumpulan data, dilakukan identifikasi masalah terlebih dahulu, untuk melihat kendala yang terjadi pada Perusahaan bisnis ritel XYZ. Kemudian dilakukan observasi agar memperoleh data yang dibutuhkan untuk analisis segmentasi pelanggan. Dataset yang digunakan pada penelitian ini yang berasal dari Perusahaan bisnis ritel XYZ. Dataset yang digunakan yaitu data transaksi pelanggan perusahaan ritel XYZ selama 2 tahun terakhir pada periode 2021-2022. Data transaksi pelanggan yang digunakan dalam penelitian ini menggunakan data sebanyak 1000 data/bulan. Jadi jika dijumlahkan 1000 data dikali dengan 12 bulan yaitu sebanyak 12.000 data transaksi dalam satu tahun, jika dikali dengan 24 bulan yaitu sebanyak 24.000 data transaksi dalam periode 2021-2022.

3.5 Teknik Analisis Data

Klasterisasi pada data transaksi pelanggan akan dilakukan dalam penelitian ini dengan *data mining* menggunakan algoritma *K-Means* dan *Fuzzy C-Means*, oleh karena itu metode *CRISP-DM* digunakan untuk penelitian ini. Adapun alur dari penelitian ini yaitu pada Gambar 3.1 [44].



Gambar 3.1 Alur Penelitian

Berdasarkan Gambar 3.1 terdapat tahapan dalam melakukan penelitian ini dan dijelaskan sebagai berikut :

1. Identifikasi Masalah

Perusahaan bisnis ritel XYZ yang merupakan bisnis ritel yang menjual bahan sembako. Banyaknya pelanggan yang melakukan transaksi hingga perusahaan ini mengalami peningkatan, Namun, informasi tentang waktu dan besaran transaksi pelanggan masih belum diolah dengan baik. hanya disimpan sebagai arsip pemilik bisnis ritel, dimana hal tersebut dapat dimanfaatkan kembali untuk meningkatkan nilai perusahaan bisnis ritel XYZ dengan mengetahui kriteria pelanggan. Selain itu, persaingan antar bisnis ritel

semakin ketat, sehingga banyak yang memfokuskan untuk mempertahankan dan meningkatkan kesetiaan pelanggan.

Untuk menentukan pelanggan yang menguntungkan atau tidak, perlu dilakukan analisis segmentasi pelanggan menggunakan model *Recency, Frequency, & Monetary* (RFM). Model RFM mengelompokkan pelanggan berdasarkan interval waktu kunjungan terakhir pelanggan, frekuensi kunjungan, dan besaran nilai keluar sebagai royalti perusahaan. Ini dilakukan dengan menggunakan algoritma *k-means* dan *fuzzy c-means clustering*

2. Pengumpulan Data

Pengumpulan data dilakukan dalam tahap ini data yaitu dengan cara observasi untuk memperoleh data sehingga diperoleh data yaitu data transaksi pelanggan perusahaan ritel XYZ selama 2 tahun terakhir.

3. *Business Understanding*

Tujuan bisnisnya pada penelitian ini adalah memetakan klusterisasi pada data transaksi pelanggan yang diperlukan sebagai referensi menetapkan prioritas agar terlihat segmentasi pelanggan pada perusahaan ritel XYZ.

4. *Data understanding*

Untuk membuat klusterisasi data transaksi pelanggan berdasarkan segmentasi pelanggan, penetapan dan validasi data diperlukan pada tahap *data understanding*

5. *Data Preparation*

Tahap *data preparation*, dimana dataset disesuaikan untuk memenuhi kebutuhan tahap pemodelan. Seleksi data dilakukan untuk mempersiapkan data mentah menjadi data yang siap untuk tahap pemodelan, dan *data transformation* dilakukan menggunakan model RFM untuk mengubah data.

6. *Modeling*

Tahap ini dimana tahapan membuat model dengan menggunakan teknik *data mining clustering*, menggunakan algoritma *K-means clustering* dan *Fuzzy C-means clustering*. Proses *modeling* dilakukan dengan *dataset* yang sama dengan menggunakan algoritma tersebut secara terpisah. Pada tahap ini diimplementasikan dalam bentuk visualisasi menggunakan bahasa

pemrograman *Python* dan tools *Jupyter Notebook* dan *Microsoft Excel*.

Adapun langkah-langkah untuk algoritma *K-means* diantaranya:

- a) Menginisialisasi pusat *cluster* awal,
- b) Kemudian diulangi langkah c-e sampai tidak ada perpindahan anggota tiap *cluster*,
- c) Selanjutnya jarak antara tiap data dengan pusat *cluster* dihitung,
- d) Setelah itu data dikelompokan berdasarkan jarak minimum,
- e) Dan menghitung pusat *cluster* yang baru untuk setiap *cluster*.

Selain itu langkah-langkah untuk algoritma *Fuzzy C-means* yaitu:

- a) Menentukan matriks c berukuran $n \times p$, dengan n = jumlah data yang akan diklaster, dan p = parameter/jumlah variabel. Kemudian jumlah klaster yang akan dibentuk = c (2), serta untuk eksponen pemboot, m . Dan iterasi awal, $t = 1$
- b) Setelah itu bentuk matriks keanggotaan awal dengan bilangan acak untuk $i = 1, 2, 3, \dots, n$; $j = 1, 2, 3, \dots, c$.
- c) Kemudian cari nilai pusat klaster untuk setiap klaster dengan $i = 1, 2, 3, \dots, c$ dan $j = 1, 2, 3, \dots, p$.
- d) Selanjutnya derajat keanggotaan seluruh data masing-masing klaster diperbaiki, dengan catatan menghitung fungsi objek (formula (1)) terlebih dahulu, yang kemudian menggunakan persamaan (2.3)
- e) Tentukan kriteria berhenti atau konvergen dari perulangan, yaitu dengan melihat perubahan nilai fungsi objektif iterasi saat dengan iterasi sebelumnya atau iterasi kurang dari maksimum iterasi.

7. Evaluation

Tahap ini dimana tahap menginisialisasi pusat pengukuran *cluster* awal terhadap model klasterisasi yang dihasilkan dari tahap *modeling*. Dalam hal ini mengukur akurasi dan performansi *K-Means* dan *Fuzzy C-Means* yang digunakan dalam model, dimana pada penelitian ini menggunakan metode evaluasi *Davies Bouldin Index* (DBI), dimana perubahan jumlah *cluster* juga mempengaruhi informasi yang dihasilkan dan DBI terbukti menghasilkan

jumlah *cluster* optimal yang berisikan informasi dengan pola yang lebih baik karna memiliki nilai intra *cluster* kecil.

8. *Deployment*

Tahap ini merupakan tahap dimana hasil klasterisasi dari algoritma *K-Means* dan *Fuzzy C-Means* digunakan dalam hal segmentasi pelanggan pada perusahaan bisnis ritel XYZ.

