

## **BAB III**

### **METODE PENELITIAN**

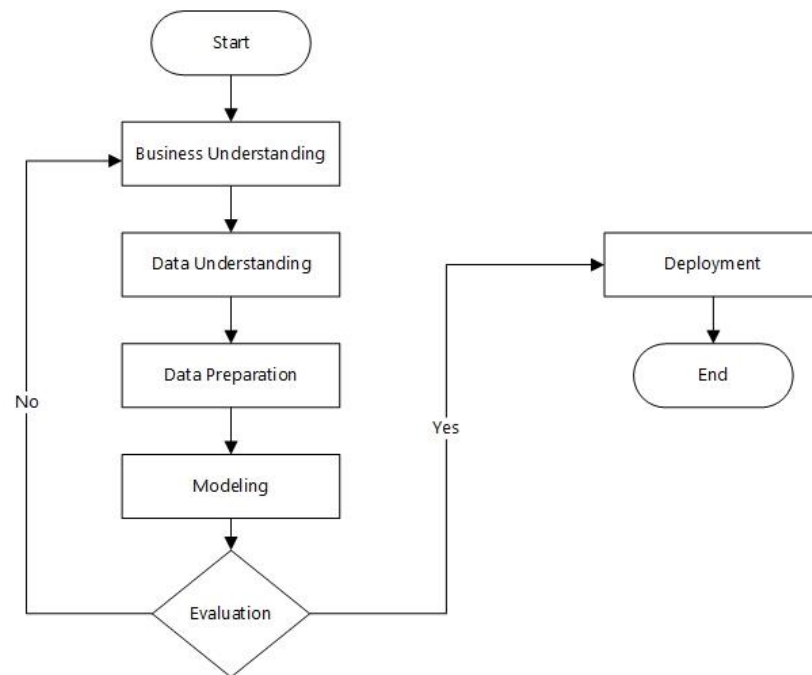
#### **3.1 Objek Penelitian**

Objek penelitian adalah variabel yang diteliti oleh di tempat penelitian dilakukan (Supriyati, 2015). Sedangkan pada sumber lain menyebutkan, objek penelitian adalah segala sesuatu yang berbentuk hal apa pun yang ditetapkan oleh peneliti untuk selanjutnya dipelajari, sehingga dapat diperoleh informasi mendalam tentang hal tersebut, kemudian ditarik kesimpulannya (Sugiyono, 2019). Dari pengertian diatas maka dapat disimpulkan bahwa objek penelitian merupakan variabel ilmiah yang diteliti dan dipelajari sehingga dapat diambil informasi dan ditarik kesimpulan. Objek pada penelitian ini merupakan *dataset* transaksi konsumen pada dua kuartal pertama tahun 2023 yakni *record* data transaksi yang terjadi mulai dari 1 Januari 2023 hingga 30 Juni 2023. Penelitian ini di lakukan di PT. XYZ yang bergerak pada industri minuman sehat di Indonesia.

#### **3.2 Desain Penelitian**

Desain penelitian merupakan rancangan penelitian yang digunakan sebagai prosedur serta pedoman dalam melakukan penelitian. Desain penelitian bertujuan untuk memberi arahan yang jelas dan terstruktur kepada peneliti dalam melakukan proses penelitian. Menurut Zakariah (Zakariah dkk., 2020) desain penelitian berisi rencana kerja lapangan yang dilakukan pada saat penelitian, sehingga diharapkan dapat memberikan gambaran dan arah mana yang dilakukan dalam melaksanakan penelitian tersebut.

Pada penelitian ini peneliti mengimplementasikan teknik analisis data menggunakan metode CRISP-DM sebagai pedoman serta alur pengerjaan proyek *data mining*, dengan tahapan yang meliputi studi literatur, fase pemahaman bisnis (*business understanding*), fase pemahaman data (*data understanding*), fase persiapan data (*data preparation*), fase pemodelan (*modeling*), fase evaluasi (*evaluation*), dan terakhir fase penyebaran (*deployment*) yang meliputi analisis karakteristik konsumen, serta pembuatan dashboard visualisasi untuk mendukung penyebaran.



**Gambar 3. 1 Diagram Alur Penelitian dengan Metode CRISP-DM**

Sumber: CRISP-DM 1.0

### 3.2.1 *Business Understanding*

Pada tahap pemahaman bisnis dilakukan beberapa hal seperti menentukan tujuan bisnis, menilai situasi, menentukan tujuan dari proyek *data mining*, dan menghasilkan rencana proyek *mining* data. Pada tahapan ini diketahui bahwa tujuan bisnis adalah untuk memberikan rekomendasi strategi pemasaran berdasarkan analisis data transaksi perusahaan dan data lainnya yang dapat meningkatkan profitabilitas perusahaan. Adapun situasi yang berlangsung dalam perusahaan saat ini bahwa perusahaan belum dapat melakukan segmentasi konsumen sehingga belum aktivitas pemasaran belum berfokus pada konsumen yang memberikan pendapatan lebih kepada perusahaan, dan konsumen yang berpotensi untuk melakukan transaksi sukses di masa depan, serta kegiatan pemasaran belum terarah dimana tidak memaksimalkan karakteristik perilaku yang dimiliki konsumen.

Tujuan *data mining* dalam tahapan ini adalah melakukan *cluster* dalam mengetahui karakteristik perilaku konsumen di setiap segmennya, selain itu menemukan pola-pola yang terbentuk berdasarkan data perusahaan. Sehingga

dihasilkan rencana proyek untuk melakukan analisis segmentasi konsumen menggunakan metode *K-Means* klustering berdasarkan model RFM sebagai rekomendasi strategi pemasaran bagi perusahaan. Sehingga perusahaan dengan tepat dapat melakukan pemasaran secara lebih efektif dan efisien. Berikut Table 3.1 memuat analisis pengetahuan data, tujuan dari setiap analisis serta *Knowledge Database Discovery* (KDD) yang dilakukan berdasarkan *database* perusahaan.

**Tabel 3. 1 Analisis Data dan Tujuan**

Analisis	Tujuan
<b>Data Segmentasi Konsumen</b>	Melakukan analisis pada data transaksi menggunakan dengan teknik <i>K-Means</i> , menggunakan model RFM. Tujuan dari analisis ini untuk menghasilkan kluster yang didasarkan pada kemiripan karakteristik data. Selanjutnya setiap kluaster akan dilakukan analisis dan disajikan rekomendasi strategi pemasaran untuk masing-masing luster.

### 3.2.1 *Data Understanding*

Tahap pemahaman data adalah langkah awal dalam proses analisis data yang melibatkan pengumpulan data yang relevan, evaluasi kualitas data, dan penemuan informasi tersembunyi yang dapat membantu membentuk hipotesis baru pada penelitian ini. Dalam konteks penelitian ini, pemahaman data dilakukan terhadap database transaksi perusahaan yang mencakup berbagai tabel. Selama tahap ini, fokus diberikan pada pemahaman tentang struktur *database*, jenis data yang terdapat dalam tabel, atribut-atribut yang relevan, jumlah data yang ada, serta analisis karakteristik data yang ada. Pemahaman data dilakukan pada database perusahaan, dimana pada database tersebut terdapat 43 tabel, namun hanya tabel yang relevan dengan tujuan penelitian yang digunakan. Sehingga, hanya 6 (enam) tabel yang digunakan pada penelitian ini yaitu tabel *users* dan *user\_transaction*. Berikut penjelasan setiap tabel dalam *database* perusahaan yang digunakan dalam penelitian:

**Tabel 3. 2 Tabel-Tabel pada Database**

No.	Nama Tabel	Jumlah <i>Records</i>	Keterangan
1	<i>users</i>	12.805 <i>records</i>	Tabel yang menyimpan seluruh informasi pengguna terdaftar
2	<i>users_transaction</i>	37.200 <i>records</i>	Tabel yang menyimpan informasi detail terkait transaksi pengguna

### 3.2.2 Data Preparation

*Data preparation* (persiapan data) merujuk pada serangkaian langkah yang dilakukan sebelum melakukan analisis data. Tujuan dari *data preparation* adalah untuk memastikan bahwa data yang digunakan dalam analisis memiliki kualitas yang tinggi, terstruktur dengan baik, tidak ada *error* dan siap untuk dianalisis. Berikut ini beberapa langkah yang biasanya dilakukan dalam data preparation.

Dalam fase persiapan data, dilakukan seleksi data yang diperlukan, penggabungan data yang terpilih, pembersihan data yang tidak diperlukan, dan pengaturan format data agar dapat diproses menggunakan teknik *data mining*. Pada fase persiapan ini, digunakan bahasa pemrograman SQL dan Python. Bahasa pemrograman SQL digunakan untuk mengintegrasikan tabel dan atribut dari database yang diperlukan untuk analisis. Sementara itu, bahasa pemrograman Python dengan *library Pandas* digunakan untuk membersihkan data seperti menghapus entri yang tidak relevan, mengolah, dan menganalisis data terstruktur. Berikut adalah tabel dan *field* yang telah dipilih dan digunakan dalam penelitian ini.

**Tabel 3. 3 Tabel dan Field**

No	Nama Tabel	Field
1	<i>users</i>	<i>id, first_name, last_name, customer_email, customer_mobile, date</i>
5	<i>users_transaction</i>	<i>id, customer_code, amount, net_value_sales, value_sales, date, status_payment</i>

### **3.2.3 Modeling**

Tahap *modeling* atau pemodelan merupakan fase yang secara langsung melibatkan teknik *data mining* yaitu dimana pada tahap ini dilakukan pemilihan teknik *data mining* dan menentukan algoritma yang digunakan. Dalam penelitian ini menggunakan algoritma *K-Means* berdasarkan model RFM (*Recency, Frequency dan Monetary*) untuk mengklusterkan konsumen yang telah melakukan transaksi dengan perusahaan, dengan menggunakan software RapidMiner versi 10.1 untuk mengolah dan mengklasterkan data tersebut.

### **3.2.4 Evaluation**

Tahap evaluasi merupakan tahap lanjutan yang digunakan untuk menilai dengan mendalam apakah *data mining* telah mencapai tujuannya. Pada tahap ini, dilakukan beberapa hal, seperti menilai sejauh mana model yang telah dibuat apakah sesuai dengan yang diharapkan, memeriksa kembali semua tahapan yang telah dilakukan untuk memastikan tidak ada yang terlewat, dan memutuskan langkah selanjutnya, apakah melanjutkan ke tahap implementasi atau kembali ke tahap awal yaitu pemahaman bisnis. Pada tahap evaluasi, Peneliti menggunakan *Davies Bouldin Index* untuk mengevaluasi hasil klustering yang didapatkan dari model sebelumnya, sehingga bisa mendapatkan jumlah kluster yang paling optimal.

### **3.2.5 Deployment**

Pada tahap penyebaran (*deployment*) ini, informasi dan pengetahuan yang telah didapatkan dari hasil penelitian kemudian direpresentasikan menjadi bentuk yang lebih mudah dipahami oleh Perusahaan. Pada tahapan ini penulis menggunakan *tool* Google Looker Studio untuk membuat dashboard visualisasi dari hasil yang didapat. Dengan penerapan Google Looker Studio dapat memberikan kemudahan dalam membaca data, penyebaran data, dan proses integrasi dashboard hasil visualisasi.

Hasil dari penelitian ini berupa analisis data *clustering* konsumen, data potensial konsumen, data produk, kategori, tipe pembayaran, dan data distribusi gender dan usia, serta memberikan rekomendasi strategi pemasaran yang dapat dilakukan perusahaan berdasarkan analisis data yang dilakukan agar dapat meningkatkan pendapatan perusahaan.

### 3.3 Metode Penelitian

#### 3.3.1 Teknik Pengumpulan Data

Pengumpulan data merupakan kegiatan yang penting bagi kegiatan penelitian, karena pengumpulan data tersebut menentukan berhasil tidaknya suatu penelitian. Sehingga dalam pemilihan teknik pengumpulan data harus cermat. Pada penelitian ini data yang digunakan adalah data sekunder, dimana data didapatkan dari subjek penelitian yaitu pihak perusahaan (Sugiyono, 2019). Teknik pengumpulan data yang digunakan dalam penelitian ini adalah.

##### 1. Studi Literatur

Pada tahap studi literatur ini dilakukan pemahaman mendalam terhadap permasalahan yang terjadi pada studi kasus dan semua referensi mengenai penyelesaian masalah yang ada pada studi kasus di PT. XYZ. Studi literatur dilakukan dengan menggali berbagai informasi melalui dokumen perusahaan, buku, *e-book*, jurnal, dan informasi pada website yang kredibel. Referensi tersebut terkait *data mining*, analisis model RFM, segmentasi pelanggan, algoritma *K-Means Clustering*, data visualisasi, dan *software RapidMiner*.

##### 2. Studi Dokumen

Studi dokumen adalah metode pengumpulan data yang digunakan sebagai pelengkap dari penggunaan metode wawancara dalam penelitian kuantitatif maupun kualitatif (Sugiyono, 2019). Pada penelitian ini studi dokumen dilakukan untuk dapat melengkapi informasi situasi perusahaan untuk mendukung analisis lebih lanjut terkait permasalahan pada perusahaan. Studi dokumen yang dilakukan adalah pembelajaran dokumen perusahaan tentang strategi pemasaran yang dilakukan saat ini dan dokumen perusahaan dalam penggunaan *database* perusahaan.

#### 3.3.2 Teknik Analisis Data

Teknik analisis data merujuk pada berbagai metode atau pendekatan yang digunakan untuk memproses, menganalisis, dan menginterpretasikan data yang dikumpulkan dalam penelitian atau studi. Tujuan dari teknik analisis data adalah untuk menggali informasi yang relevan, mengidentifikasi pola, hubungan, atau tren

yang terdapat dalam data, dan menyimpulkan temuan yang dapat digunakan untuk menjawab pertanyaan penelitian atau mencapai tujuan penelitian.

1. Analisis Deskriptif, teknik ini merujuk pada metode dan pendekatan yang digunakan untuk memberikan gambaran dan ringkasan tentang karakteristik data yang diamati. Tujuan dari analisis deskriptif adalah untuk memahami pola, tren, dan statistik dasar dari data yang ada diolah, tanpa melibatkan upaya pengembangan model prediktif atau inferensi yang lebih kompleks.
2. *Exploratory Data Analysis* (EDA), EDA atau Analisis Data Eksploratori dilakukan untuk memahami dan menganalisis *dataset* secara visual dan deskriptif dengan tujuan untuk mengungkap pola, hubungan, dan wawasan baru yang terdapat dalam data (Wickham Grolemond, 2017). EDA melibatkan eksplorasi variabel, distribusi data, dan anomali. Pada penelitian ini EDA diaplikasikan dalam tahap pemahaman data dan persiapan data menggunakan perintah bahasa Python dan *query* SQL.
3. *Clustering*, teknik data mining yang digunakan untuk mengelompokkan objek atau entitas serupa dalam dataset. Algoritma klasterisasi bertujuan dalam mencari pola atau kesamaan dalam data dan membaginya ke dalam kelompok atau klaster berdasarkan kriteria tertentu. Pada penelitian ini, proses *clustering* dilakukan pada tahap pemodelan menggunakan aplikasi RapidMiner.

### 3.4 Alat dan Bahan Penelitian

#### 3.4.1 Alat

Adapun penelitian ini menggunakan perangkat keras (*Hardware*) dan perangkat lunak (*Software*) dengan spesifikasi dan penjelasan sebagai berikut:

##### 1. Perangkat Keras

Satu Personal Computer (PC) dengan spesifikasi:

**Tabel 3. 4 Spesifikasi Perangkat Keras**

No.	Nama	Spesifikasi
1	<i>Processor</i>	Intel(R) Core(TM) i7-10750H CPU @ 2.60GHz
2	<i>Graphic</i>	NVIDIA GeForce RTX 2060
3	RAM	16,00 Gb

Deni Riswana, 2023

**ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING BERDASARKAN MODEL RFM SEBAGAI REKOMENDASI STRATEGI PEMASARAN (STUDI KASUS DI PT.XYZ)**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu

No.	Nama	Spesifikasi
4	Hardisk	1512 Gb
5	Sistem Operasi	Windows 11 Pro 64-Bit

## 2. Perangkat Lunak

### a. Google Colab

Google Colab digunakan untuk melakukan pemrograman bahasa python dalam melakukan penarikan data dari *database* untuk, *data understanding*, *data pre-processing*, *data preparation* seperti menghilangkan data yang tidak digunakan, menghapus atribut, menghitung nilai RFM, dan memeriksa kesalahan pada data yang lainnya. *Library* Python yang digunakan diantaranya *library pandas*, *numpy*, *pygsheet\_script*, dan *library mysql*.

### b. RapidMiner

RapidMiner digunakan sebagai aplikasi dalam melakukan tahap *modeling* dan analisis *data mining* menggunakan metode dan algoritma *K-Means* clustering dan beberapa metode lainnya. Pada penelitian ini penulis menggunakan *software* RapidMiner Studio Educational versi 10.1.

### c. Microsoft Excel

Microsoft Excel digunakan untuk menyimpan data yang sudah dilakukan proses pre-processing pada aplikasi Google Colab untuk dilakukan visualisasi pada Google Looker Studio dan melakukan pengunduhan data untuk dianalisis pada aplikasi RapidMiner menggunakan algoritma yang diinginkan.

### d. Google Looker Studio

Google Looker Studio digunakan untuk memvisualisasikan data berupa dashboard yang informatif agar lebih mudah dipahami dan dibagikan kepada *stakeholder*.

## 3.4.2 Bahan

Dalam penelitian ini bahan penelitian merupakan data sekunder yang diperoleh dari perusahaan yang terdiri atas *dataset* transaksi penjualan perusahaan dari tanggal 1 Januari 2023 hingga 30 Juni 2023 dan *dataset* pendukung lainnya yang dilakukan analisis sebagai bentuk rekomendasi strategi pemasaran pada PT. XYZ.

Deni Riswana, 2023

**ANALISIS SEGMENTASI KONSUMEN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING BERDASARKAN MODEL RFM SEBAGAI REKOMENDASI STRATEGI PEMASARAN (STUDI KASUS DI PT.XYZ)**

Universitas Pendidikan Indonesia | repository.upi.edu | perpustakaan.upi.edu