

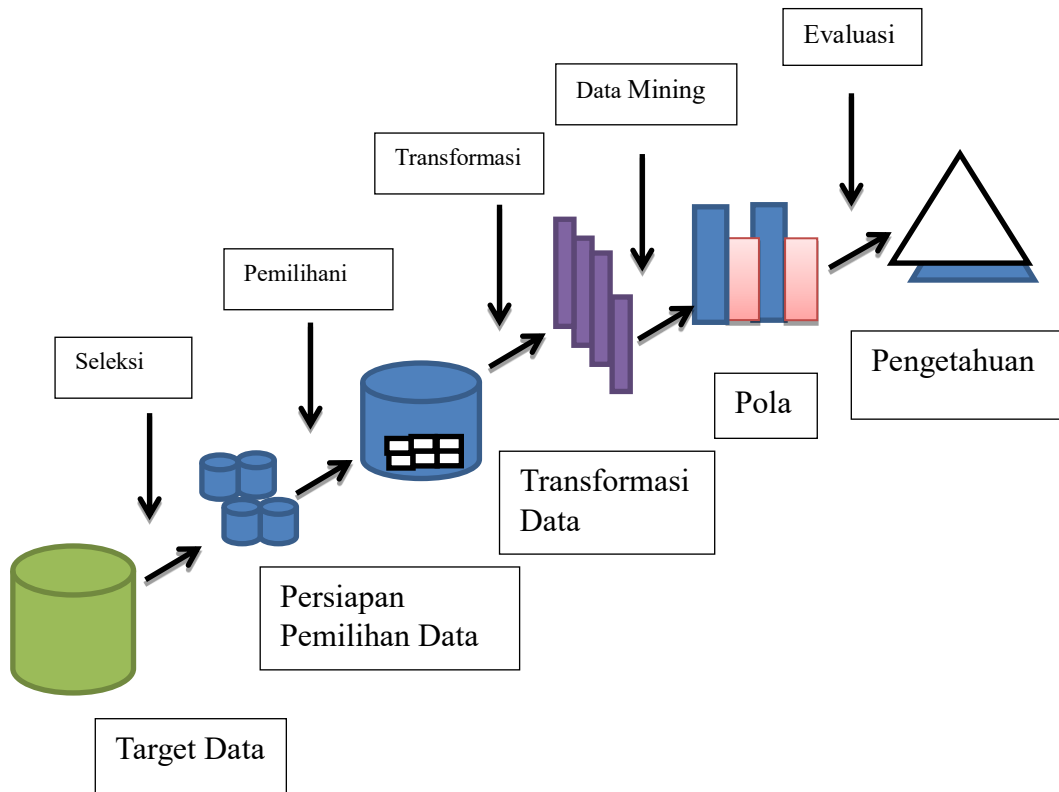
## BAB II

### LANDASAN TEORI

#### 2.1 *Knowledge Discovery In Database*

*Knowledge Discovery in Database* adalah bentuk dari suatu proses pengambilan informasi data yang tersembunyi, di mana informasi dari data tersebut sebelumnya tidak pernah dikenal. *Knowledge Discovery in Database* juga merupakan suatu metode yang sering digunakan untuk mendapat suatu pengetahuan yang berasal dari database yang ada pada data mining. Hasil dari suatu pengetahuan yang diterima dapat dimanfaatkan untuk basis pengetahuan (*knowledge base*) yang dipergunakan dalam keperluan mengambil keputusan.

*Knowledge Discovery in Database* adalah suatu teknik pembentukan pola atau rule dalam informasi. Informasi yang dihasilkan didapatkan dari suatu data yang besar atau dikenal dengan tambang data yang disimpan dalam basis data yang awalnya belum diketahui dan menghasilkan suatu data yang potensial bermanfaat [1]. *Knowledge discovery In Database* merupakan serangkaian dari proses yang tidak mudah untuk memeriksa serta mengetahui pola dalam suatu data, dimana pola itu bersifat asli, kemudian dapat dipahami dan bermanfaat untuk pengguna [2]. *Knowlegde Discovery in Database* (KKD) adalah proses yang berfokus pada penggalian atau pencarian informasi yang berguna dari kumpulan data yang besar. Proses ini melibatkan beberapa tahapan penting seperti pembersihan data (*data cleaning*), pengambilan sampel (*sampling*), penskalaan (*scaling*), pengelompokan (*clustering*), dan lainnya untuk mendapatkan wawasan yang bermakna dan dapat digunakan dalam pengambilan keputusan. KDD tidak hanya mencakup proses penemuan pola, tetapi juga pemahaman dan validasi terhadap pola tersebut agar hasilnya benar-benar relevan dan bernilai dalam konteks bisnis atau penelitian tertentu [3].



**Gambar 2.1 Tahapan Dalam *Knowledge Discovery in Database***

Salah satu tahapan dalam keseluruhan proses *Knowledge Discovery in Database* adalah data mining. Seperti yang ditunjukkan pada proses berikut:

1. *Selection* (Seleksi Data )

Pada bagian ini *Selection* (seleksi) ataupun pemilihan data dari beberapa sekumpulan data dari operasional yang perlu dilakukan sebelum tahapan penggalian sebuah informasi yang ada di dalam *Knowledge discovery database* dimulai dari awal. Adapun Data dari hasil sebuah seleksi yang akan dipakai atau digunakan untuk sebuah proses dari data mining tersebut biasanya disimpan di dalam sesuatu berkas yang besar dan biasanya dapat terpisah dari beberapa basis data operasional yang dilakukan.

2. *Preprocessing* (Pemilihan Data)

Pada bagian ini Proses *Preprocessing* hanya mencakup di antara lain adalah dengan cara membuang duplikasi data yang ada, lalu memeriksa data yang lebih inkonsisten yang ada, dan tentunya dengan memperbaiki kesalahan yang ada pada data tersebut, contohnya seperti kesalahan pada cetak atau disebut tipografi.

### 3. *Transformation* ( Transformasi)

Pada bagian fase saat ini yang dapat dilakukan adalah dengan mentransformasi dari bentuk data yang tentunya belum memiliki sebuah entitas yang sudah sangat jelas kedalam bentuk data yang sangat valid atau bisa di bilang siap untuk pengerjaan yang dilakukan pada proses *Data mining* yang sedang berjalan.

### 4. *Data Mining*

Pada bagian fase saat ini hal yang utama yang dapat dilakukan adalah dengan cara menerapkan sebuah algoritma ataupun metode pencarian pengetahuan yang akan dilakukan pada proses pencarian data.

### 5. *Interpratation/Evaluation* ( Interpretasi/Evaluasi )

Pada bagian fase yang terakhir ini yang dapat dilakukan adalah dengan menaruh beberapa cara dari suatu proses yang akan melalui suatu pembentukan dari keluaran yang sangat mudah dimengerti dan biasanya yang akan bersumber daripada proses yang terjadi pada *Data mining* yang akan di lakukan pada suatu pola informasi yang dicari.

## 2.2 *Data Mining*

*Data mining* adalah suatu proses untuk mengekstraksi atau menggali informasi yang tersembunyi dalam *data base* yang sangat besar, yang bermanfaat bagi banyak pihak yang berkepentingan. Proses ini membantu pemegang data dalam menganalisis dan menemukan hubungan-hubungan yang tidak terduga antara berbagai data yang ada, sehingga memungkinkan untuk memperoleh wawasan baru yang dapat mendukung pengambilan keputusan yang lebih baik.

Secara umum, *data mining*, yang juga sering disebut sebagai penambangan data, dapat didefinisikan sebagai proses eksplorasi, seleksi, dan pemodelan sejumlah besar data untuk menemukan pola atau kecenderungan yang sebelumnya tidak disadari. Tujuan utama dari *data mining* adalah untuk menemukan pola yang sebelumnya tidak diketahui keberadaannya. Setelah pola ditemukan, langkah selanjutnya adalah menggunakan pola tersebut untuk menghasilkan keputusan yang relevan untuk keperluan tertentu yang berkepentingan dengan data tersebut.

*Data mining* adalah proses untuk menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa sejumlah besar data yang tersimpan dalam sistem penyimpanan, dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika [4]. *Data Mining* sendiri adalah kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran besar sehingga menghasilkan informasi baru, informasi inilah yang nantinya sangat berguna untuk pengembangan atau pendukung keputusan[5]. *Data mining* adalah suatu kegiatan mengekstraksi atau menambang pengetahuan dari data yang berukuran besar, informasi inilah yang nantinya sangat berguna untuk pengembangan[6].

*Data mining* adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika untuk menghasilkan sebuah Informasi yang digunakan untuk kebutuhan yang sangat penting bagi manusia, untuk menunjang segala aktivitas yang dilakukan. Terkadang, kebutuhan yang ada terkadang memiliki nilai suatu informasi yang sangat tinggi tidak diimbangi dengan pemberian informasi yang memadai.

Akibatnya informasi tersebut harus kembali digali dalam data yang besar. Salah satu metode yang biasa atau seringkali digunakan dalam mencari informasi dalam data yang besar yaitu *data mining*. Dalam *data mining* terdapat beberapa teknik salah satunya yaitu *association rule*[7]. *Data mining* merupakan proses menemukan suatu informasi dari suatu data yang tersimpan dalam suatu *database* atau *datasheet*. Pembuatan pada model ini biasanya dilakukan dengan proses menggunakan algoritma atau rumus tertentu yang ada pada *data mining*. Proses yang terdapat pada *data mining* ini biasanya sering kali menggunakan berbagai teknik tertentu seperti teknik dalam proses statistik, matematika, dan *machine learning* yang digunakan dalam melakukan identifikasi dan mengolah berbagai data menjadi informasi yang bermanfaat [8].

### **2.3 Metode Regresi Linier**

Regresi linier adalah salah satu metode prediksi. regresi linier merupakan teknik kuantitatif yang digunakan dalam analisis time series yang menggunakan

waktu sebagai dasar prediksi. Regresi linier dibedakan menjadi dua jenis yaitu regresi linier sederhana dan regresi linier berganda. Jika regresi linier sederhana, digunakan ketika ingin menggambarkan hubungan antara satu variabel independent dan satu variabel dependen. Berikut adalah persamaannya[9].

$$Y = \alpha + bX. \quad (1)$$

$$\alpha = \frac{(\sum X^2)(\sum Y) - (\sum XY)(\sum X)}{n(\sum X^2) - (\sum X)^2} \quad (2)$$

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2} \quad (3)$$

Keterangan :

Y = Variabel terkait

X = Variabel bebas

a = Intercept

b = Koefisien variabel X

n = Jumlah data

Sedangkan, regresi linier berganda digunakan ketika ingin mempelajari hubungan antara satu variabel dependen dan dua atau lebih variabel independent. Berikut adalah persamaannya :

$$Y = \alpha + b_1X_1 + b_2X_2 + \dots + b_nX_n \quad (4)$$

Keterangan :

Y = Variabel terikat

X = Variabel bebas

$\alpha$  = Intercept

b = Koefisien variabel X

Metode regresi linier adalah salah satu teknik statistik yang digunakan untuk memodelkan hubungan antara variabel independen dan variabel dependen. Metode regresi linier digunakan untuk memprediksi atau mengestimasi nilai variabel dependen berdasarkan variabel independen yang sudah diketahui [10]. Regresi linear adalah teknik yang digunakan untuk memperkirakan atau memprediksi keterkaitan antara dua variabel dalam penelitian kualitatif. Ini merupakan bagian dari *Machine Learning* yang termasuk dalam *Supervised Learning*, di mana komputer diprogram untuk mengenali pola antara data input dan output label.

*Machine learning* juga digunakan untuk menemukan hubungan yang mendasari antara data input dan output label. Regresi linear digunakan dalam aplikasi untuk membuat prediksi berdasarkan data yang ada, dengan asumsi bahwa hubungan antara variabel-variabel tersebut bisa dijelaskan dengan persamaan garis lurus. Model yang mencoba memodelkan hubungan antar variabel dalam data disebut sebagai model regresi linear [11]. Analisis regresi linier sederhana didasarkan pada hubungan kausal atau fungsional satu variabel independen dengan variabel dependen. Koefisien regresi bertujuan untuk memastikan apakah variabel independen yang terdapat dalam persamaan regresi tersebut secara individu berpengaruh terhadap nilai variabel dependen. Analisis regresi linear sederhana berfungsi untuk menguji hubungan sebab akibat antara variabel faktor penyebab terhadap variabel akibatnya.

Penelitian regresi linear sederhana digunakan untuk menguji hubungan variabel kualitas layanan terhadap kepuasan nasabah. Analisis regresi sederhana adalah sebuah metode pendekatan untuk permodelan hubungan antara satu variabel dependen dan satu variabel independen. Dalam model regresi, variabel independen menerangkan variabel dependennya. Dalam analisis regresi sederhana, hubungan antara variabel independen (X) akan diikuti oleh perubahan pada variabel dependen (Y) secara tetap. Sementara hubungan non linier, perubahan variabel X tidak diikuti dengan perubahan variabel Y secara *proposional*[12].

Regresi Linear Sederhana adalah metode peramalan yang meninjau hubungan antara variabel independent terhadap variabel dependent. Analisis regresi linear sederhana bertujuan untuk mengukur kekuatan hubungan antara dua variabel, serta menunjukkan arah hubungan antara variabel independent dengan variabel dependent[13]. Regresi linier adalah suatu bentuk regresi yang menunjukkan hubungan linier, yang pada gilirannya menemukan korelasi antara target atau variabel dependen sesuai dengan prediktor variabel independen. Metode ini bekerja paling baik ketika kumpulan data yang biasanya diberikan adalah kumpulan data linier yang pada gilirannya membentuk hubungan linier pada plot. Namun, dalam proses aplikasi yang ada pada sebenarnya (regresi linear), kumpulan sampel data merupakan langkah yang sangat kritis[14].

## 2.4 Prediksi

Prediksi atau peramalan dapat diartikan sebagai suatu kegiatan untuk memproyeksikan kejadian di masa depan. Proses ini melibatkan pengamatan data atau informasi dari masa lalu atau saat ini dengan menggunakan metode matematika atau statistik. Prediksi bertujuan untuk mengetahui, melihat dan memperkirakan prospek ekonomi atau kegiatan usaha. Prediksi dapat berupa kualitatif (berupa deskripsi tanpa angka) atau kuantitatif (berupa angka). Namun, peramalan kualitatif cenderung sulit menghasilkan hasil yang akurat karena variabelnya relatif [15].

Prediksi adalah proses memproyeksikan suatu yang akan datang. Prediksi merupakan bentuk kombinasi antara seni dengan ilmu dalam memperkirakan kondisi di masa depan, dengan memproyeksikan data masa lalu ke masa depan menggunakan model matematika dan perkiraan subjektif[16]. Prediksi menentukan jumlah kebutuhan untuk bulan depan dalam hal dukungan data historis atau kumpulan waktu atau periode waktu yang dianalisis. Ini memungkinkan Anda menghitung untuk memprediksi jumlah kebutuhan bulan depan. Prediksi dapat digunakan tidak hanya untuk memprediksi deret waktu, tetapi juga untuk klasifikasi. Secara alami dan dapat menghasilkan kelas berdasarkan atribut yang ada [17].

Prediksi adalah tentang sesuatu yang akan terjadi di masa depan. Prediksi adalah salah satu proses memperkirakan secara sistematis tentang sesuatu yang paling mungkin terjadi di masa depan berdasarkan informasi masa lalu dan sekarang yang dimiliki, agar kesalahannya (selisih antara sesuatu yang terjadi dengan hasil perkiraan dapat diperkecil[18].

## 2.5 Kepuasan Pelanggan

Dalam menciptakan kepuasan pelanggan, perusahaan harus dapat meningkatkan nilai pelanggan, semakin baik penilaian pelanggan akan semakin tinggi pula kepuasan pelanggan. Konsumen termotivasi dalam berbelanja karena adanya unsur dorongan dan kebutuhan yang muncul karena gaya hidup. Pada era kompetisi sengit dimana semua perusahaan menjual *core product* yang relatif sama, maka perusahaan harus mampu membangun hubungan jangka panjang yang

saling menguntungkan dengan pelanggan. Persaingan dalam suatu bisnis berkembang pesat yang membuat satu dan perusahaan lain terus bersaing dan bertahan dalam bisnis tersebut.

Perusahaan harus menentukan beberapa strategi pemasaran untuk bertahan dan mencapai tujuan utama[19]. Untuk menjaga para pelanggan agar tidak memilih produk pesaing tentu bukan hal yang mudah. perusahaan harus melakukan beberapa hal yang dapat mengikat hati pelanggan untuk tetap setia pada produk perusahaan. Salah satunya dengan pelayanan terbaik untuk pelanggan. Pelayanan terbaik tentu harus memahami kebutuhan dan keluhan para pelanggan. Sehingga pelayanan yang dilakukan oleh perusahaan tepat sasaran. Selain itu, kepuasan pelanggan ikut menjadi bagian penting untuk menjaga loyalitas pelanggan perusahaan. Kualitas pelayanan yang mendefinisikan sebagai kondisi dinamis yang berhubungan dengan produk, jasa, sumber daya manusia, proses dan lingkungan yang memenuhi atau melebihi harapan. Kepuasan konsumen merupakan salah satu elemen penting dalam peningkatan kinerja pemasaran dalam suatu perusahaan. Kepuasan yang dirasakan oleh pelanggan dapat meningkatkan intensitas membeli dari pelanggan tersebut. Dengan terciptanya tingkat kepuasan pelanggan yang optimal maka mendorong terciptanya loyalitas di benak pelanggan yang merasa puas tadi.

Loyalitas merupakan persentase dari orang yang pernah membeli dalam kerangka waktu tertentu dan melakukan pembelian ulang sejak pembelian yang pertama[20]. Kepuasan konsumen merupakan salah satu faktor penentu keberhasilan perusahaan, selain itu kepuasan konsumen dapat menjadi alat bersaing bagi suatu perusahaan dalam menghadapi pesaingnya. Kepuasan konsumen merupakan salah satu elemen penting dalam peningkatan kinerja pemasaran dalam suatu perusahaan. Kepuasan yang dirasakan oleh pelanggan dapat meningkatkan intensitas membeli dari pelanggan tersebut [21].

Keputusan pembelian adalah keputusan informasi tentang keunggulan suatu produk yang disusun sedemikian rupa sehingga menimbulkan rasa menyenangkan yang akan merubah seseorang untuk melakukan keputusan pembelian. Pengertian mengenai perilaku konsumen oleh perusahaan selaku produsen sangat penting dan perlu diperhatikan lebih lanjut [22].

## 2.6 Evaluasi Model

Evaluasi algoritma digunakan untuk mengukur performa model yang diterapkan, dalam hal ini regresi linier. dua metrik utama yang digunakan untuk mengevaluasi model ini adalah *Mean Absolute Error (MAE)*, *Root Mean Squared Error (RMSE)*, Metrik-metrik ini membantu dalam menilai seberapa baik model memprediksi data yang ada.

### 2.6.1 Mean Absolute Error (MAE)

MAE mengukur rata-rata selisih absolut antara nilai prediksi dan nilai aktual. Metrik ini menunjukkan seberapa besar rata-rata kesalahan prediksi tanpa mempertimbangkan arah kesalahan. Semakin kecil nilai MAE, semakin akurat model. Rumusnya:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Di mana  $y_i$  adalah nilai aktual dan  $\hat{y}_i$  adalah nilai prediksi.

### 2.6.2 Root Mean Squared Error (RMSE)

RMSE mengukur akar kuadrat dari rata-rata kuadrat selisih antara nilai prediksi dan aktual. RMSE sensitif terhadap outlier dan memberikan penalti lebih besar pada kesalahan besar. Rumusnya:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

Di mana  $y_i$  adalah nilai aktual,  $\hat{y}_i$  adalah nilai prediksi dan  $n$  adalah jumlah data RMSE akan memberikan nilai dalam satuan yang sama dengan data asli, yang memudahkan interpretasi kesalahan dalam konteks masalah yang sedang dianalisis.

## 2.7 Alat Bantu Pemrograman untuk Implementasi

Dalam penelitian ini, berbagai perangkat lunak digunakan untuk menganalisis dan membandingkan metode Regresi linier, Berikut adalah beberapa perangkat lunak yang digunakan.

### 1. *Microsoft Excel*

*Microsoft Excel* adalah perangkat lunak yang mudah digunakan dan sangat populer dalam dunia bisnis dan analisis data sederhana. Fitur analisis data bawaan seperti *AnalysisToolPak* memungkinkan pengguna untuk melakukan analisis regresi linier tanpa perlu keahlian teknis mendalam.



**Gambar 2.2 Excel**

**Sumber : Wikipedia.org**

*Excel* sangat populer karena kemampuannya dalam mengelola data dalam bentuk tabel, melakukan perhitungan otomatis menggunakan rumus, serta menyajikan data dalam bentuk grafik dan diagram yang mudah dipahami.

### 2. RapidMiner

RapidMiner adalah platform analisis data yang menyediakan lingkungan visual untuk membangun model prediktif, termasuk regresi linier. RapidMiner memudahkan pengguna untuk melakukan analisis data, pemodelan, dan evaluasi, sehingga cocok untuk penelitian ini.



**Gambar 2.3 RapidMiner**

**Sumber: linkedin.com**

Fitur-fitur yang tersedia di dalam RapidMiner mencakup seluruh tahapan dalam proses *Knowledge Discovery in Database* (KDD), seperti *data preprocessing* (pembersihan data, transformasi, normalisasi), *data mining* (penerapan model prediktif), serta *evaluasi dan visualisasi* hasil model. Platform ini juga menyediakan berbagai *operator* yang dapat dikombinasikan secara fleksibel, sehingga pengguna dapat menyesuaikan proses analitik sesuai dengan kebutuhan penelitian.

## **2.8 Metode Penelitian**

Penelitian ini menggunakan pendekatan kuantitatif dengan metode deskriptif untuk menggambarkan secara rinci proses seleksi data pada tahap awal KDD dan bagaimana seleksi data tersebut mempengaruhi kualitas hasil dari proses data mining, dengan menggunakan data sekunder yang berasal dari basis data operasional yang telah tersedia pada Sambal Sinyar-Nyar, mencakup informasi relevan seperti Harga, Kualitas Rasa, dan Pelayanan, guna memprediksi tingkat kepuasan pelanggan menggunakan metode regresi linier.

### **2.8.1 Variabel Penelitian**

Penelitian ini menganalisis empat variabel utama yang terdiri dari tiga variabel independen dan satu variabel dependen untuk memprediksi kepuasan pelanggan, yaitu:

1. Harga ( $X_1$ ) Menilai persepsi pelanggan terhadap harga produk restoran, yang harus sebanding dengan kualitas dan pelayanan yang diberikan. Harga yang wajar dapat meningkatkan kepuasan dan loyalitas pelanggan, sementara harga yang terlalu tinggi dapat menurunkan kepuasan.
2. Kualitas dan Rasa ( $X_2$ ) Menggabungkan kualitas bahan baku, kebersihan, dan cita rasa makanan. Keduanya mempengaruhi pengalaman keseluruhan pelanggan dan kepuasan mereka.
3. Pelayanan ( $X_3$ ) Mengukur kualitas layanan restoran, termasuk sikap karyawan, kecepatan, dan kesigapan dalam memenuhi kebutuhan pelanggan. Pelayanan yang baik berperan penting dalam meningkatkan kepuasan pelanggan.

4. Kepuasan Pelanggan (Y) Variabel dependen yang diukur berdasarkan sejauh mana harga, kualitas rasa, dan pelayanan memenuhi ekspektasi pelanggan, yang berpengaruh pada loyalitas pelanggan.

### **2.8.2 Metode Pengumpulan Data**

Data yang digunakan dalam penelitian ini diperoleh melalui beberapa teknik pengumpulan data sebagai berikut:

1. Studi Pustaka

Studi pustaka merupakan pendekatan yang mengumpulkan data dengan mengkombinasikan berbagai literatur yang relevan, seperti skripsi, buku, dan jurnal penelitian yang telah ada, untuk mendukung dasar teori dalam penelitian ini.

2. Observasi

Observasi dilakukan sebagai metode pengumpulan data secara sistematis, di mana peneliti memperhatikan dan mencatat interaksi atau fenomena yang terjadi di lingkungan restoran sebagai bagian dari pengalaman pelanggan.

3. Kuesioner

Kuesioner merupakan instrumen pengumpulan data primer yang fleksibel dan mudah digunakan. Kuesioner ini akan disebarakan kepada responden melalui platform seperti Google Formulir dan sosial media (WhatsApp) untuk memperoleh opini, sikap, harapan, dan keinginan responden terkait kepuasan pelanggan di Sambal Sinyar-Nyar.

## 2.9 Kerangka Kerja Penelitian

Kerangka kerja penelitian yang akan dilakukan disajikan pada tabel dibawah ini yaitu sebagai berikut:

**Tabel 2.1 Kerangka kerja penelitian**

| No | Kegiatan                   | Des | Jan | Feb | Mar | Aprl | Mei |
|----|----------------------------|-----|-----|-----|-----|------|-----|
| 1  | Penentuan topik penelitian | ✓   |     |     |     |      |     |
| 2  | Pendefinisian Masalah      | ✓   |     |     |     |      |     |
| 3  | Menganalisa Masalah        | ✓   | ✓   |     |     |      |     |
| 4  | Menentukan Tujuan          | ✓   | ✓   |     |     |      |     |
| 5  | Pengumpulan Data           |     |     | ✓   | ✓   | ✓    |     |
| 6  | <i>Cleaning Data</i>       |     |     | ✓   | ✓   | ✓    | ✓   |
| 7  | Transformasi Data          |     |     |     |     | ✓    | ✓   |
| 8  | Merancang Algoritma        |     |     |     |     |      | ✓   |
| 9  | Pengujian Algoritma        |     |     |     |     |      | ✓   |
| 10 | Evaluasi Akhir             |     |     |     |     | ✓    | ✓   |
| 11 | Pengajuan Seminar Proposal |     |     |     |     |      | ✓   |
| 12 | Seminar Proposal           |     |     |     |     |      | ✓   |

Catatan :

- \* Durasi setiap kegiatan dapat disesuaikan dengan kebutuhan penelitian.
- \* Waktu pelaksanaan bergantung pada kalender akademik dan kesiapan.