

BAB II KERANGKA TEORITIS

A. LANDASAN TEORI

1. Data Mining

Data Mining merupakan salah satu cabang ilmu pada bidang statistika dan ilmu komputer yang berfokus pada proses sistematis untuk mengekstraksi pengetahuan dari kumpulan data dalam jumlah besar. Proses ini dilakukan dengan memanfaatkan metode dan algoritma cerdas yang terstruktur secara komprehensif. Data yang telah diolah melalui teknik data mining tidak hanya menjadi sumber informasi baru, tetapi juga dapat dimanfaatkan sebagai dasar dalam menganalisis perkembangan teknologi maupun sebagai bahan pertimbangan dalam pengambilan keputusan di perusahaan maupun instansi lainnya (Suhirman, 2023, p. 20). Bagian ini merupakan tahapan di mana data yang telah dikumpulkan kemudian dimasukkan ke dalam aplikasi *data mining* untuk dianalisis lebih lanjut, sehingga dapat ditemukan pola-pola tersembunyi dari kumpulan data yang besar. Menurut Suhirman (2023), pemilihan metode atau teknik dalam *data mining* harus disesuaikan dengan tujuan analisis yang ingin dicapai, karena setiap teknik memiliki karakteristik dan kegunaan yang berbeda sesuai kebutuhan penelitian maupun penerapannya.



Gambar 2. 1 Proses Data Mining

Sumber: (Suhirman, 2023, p. 28)

Proses *data mining* dilakukan dengan memanfaatkan beragam teknik statistik, matematis, dan kecerdasan buatan untuk menganalisis data secara sistematis dan otomatis. Melalui proses ini, informasi yang diperoleh dapat dimanfaatkan untuk mendukung pengambilan keputusan, mengenali pola atau tren pasar, meningkatkan efektivitas operasional, hingga merumuskan strategi bisnis

yang lebih tepat. Tujuan utama *data mining* adalah menggali pengetahuan atau wawasan yang tidak dapat diperoleh hanya melalui pengamatan langsung terhadap data mentah (Rahayu *et al.*, 2018, p. 2).

Berikut adalah rangkaian proses data mining dapat dilihat dalam Gambar 2.2;



Gambar 2. 2 Proses Tahapan Data Mining

Sumber: (Suhirman, 2023, p. 45)

Data preprocessing merupakan rangkaian tahapan yang diterapkan pada data mentah (*raw data*) sebelum digunakan dalam proses analisis lebih lanjut. Tahap ini bertujuan untuk meningkatkan kualitas data sehingga hasil analisis yang diperoleh lebih akurat, sekaligus meminimalkan adanya ketidaksesuaian, kesalahan, atau kekurangan yang terdapat pada data mentah. (Rahayu *et al.*, 2018, p. 46).

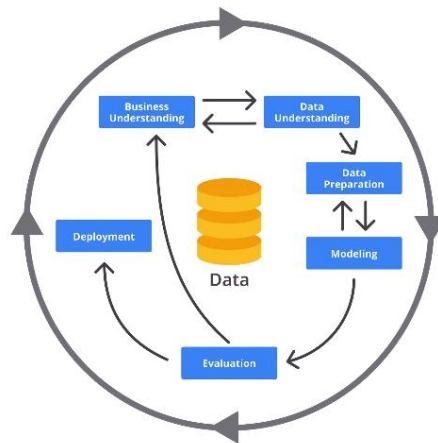
Transformasi data merupakan proses yang digunakan dalam analisis data untuk mengubah nilai variabel agar sesuai dengan asumsi atau kebutuhan tertentu dalam pemodelan. Proses ini berperan dalam memperbaiki distribusi data, mengurangi potensi bias, serta mempermudah interpretasi hasil analisis. Beberapa teknik transformasi yang umum digunakan antara lain normalisasi, transformasi logaritmik, dan metode sejenis lainnya. (Rahayu *et al.*, 2018, p. 52).

Normalisasi data adalah proses mengubah nilai suatu variabel ke dalam skala yang seragam agar setiap variabel memiliki rentang nilai yang sebanding. Tujuan dari normalisasi adalah mencegah dominasi variabel dengan nilai yang lebih besar terhadap hasil perhitungan atau analisis, sehingga setiap atribut memiliki kontribusi yang proporsional dalam proses pengolahan data (Rahayu *et al.*, 2018, p. 52).

2. CRISP-DM (Cross-Industry Standard Process for Data Mining)

Metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) merupakan sebuah kerangka kerja atau metodologi yang banyak digunakan dalam bidang data mining dan analisis data. Kerangka ini memberikan panduan sistematis bagi para profesional untuk menjalankan proyek data mining mulai dari tahap perencanaan hingga implementasi. CRISP-DM dikenal sebagai metodologi

yang terstruktur dan komprehensif, sehingga mampu meningkatkan konsistensi serta efektivitas dalam proses analisis data lintas industri (SLN, 2023, p. 35).



Gambar 2. 3 Fase CRISP-DM dan proses rekursif pada tahap Data Mining
Sumber: (SLN, 2023, p. 35)

CRISP-DM terdiri dari enam tahap utama, yaitu: (SLN, 2023, p. 36)

- (a) **Pemahaman Bisnis (*Business Understanding*)**: Tahap ini berfokus pada pemahaman yang mendalam mengenai permasalahan yang akan diselesaikan melalui proses data mining. Pada tahap ini dilakukan interaksi dengan pemangku kepentingan guna merumuskan tujuan bisnis, menentukan target yang ingin dicapai, serta mengidentifikasi berbagai kendala yang mungkin memengaruhi jalannya proses analisis.
- (b) **Pemahaman Data (*Data Understanding*)**: Tahap ini menekankan pada proses pengumpulan serta pemahaman data yang relevan untuk mendukung proyek data mining. Kegiatan yang dilakukan mencakup identifikasi dan eksploitasi berbagai sumber data yang tersedia, eksplorasi awal untuk mengetahui kondisi data, serta analisis karakteristik utama dari data tersebut. Melalui tahap ini, peneliti dapat memperoleh gambaran awal mengenai kualitas, kelengkapan, serta potensi permasalahan yang terdapat pada data sebelum masuk ke tahap pengolahan lebih lanjut.
- (c) **Persiapan Data (*Data Preparation*)**: Tahap ini merupakan bagian penting dalam metodologi CRISP-DM yang dikenal sebagai data preparation. Aktivitas pada tahap ini mencakup pembersihan data untuk menghilangkan noise atau data yang tidak relevan, transformasi data agar sesuai dengan format analisis yang diinginkan, serta penggabungan data dari berbagai sumber jika diperlukan. Selain itu, dilakukan pula proses pelabelan atau pengkodean data agar dapat diinterpretasikan dengan baik oleh algoritma

data mining. Dengan adanya tahap ini, kualitas data yang digunakan dalam pemodelan dapat terjamin sehingga hasil analisis yang diperoleh menjadi lebih akurat dan reliabel.

- (d) **Modeling**: Pada tahap ini dilakukan pemilihan metode atau algoritma data mining yang paling sesuai dengan tujuan analisis. Proses ini mencakup pembangunan dan pengembangan model analisis data menggunakan teknik tertentu, seperti regresi, klasifikasi, klasterisasi, maupun analisis asosiasi. Pemilihan teknik sangat bergantung pada jenis masalah yang dihadapi serta tujuan akhir dari proyek. Tahap ini juga dapat melibatkan pengaturan parameter dan pengujian awal model agar diperoleh performa yang optimal.
- (e) **Evaluasi (Evaluation)**: Setelah model dikembangkan, langkah berikutnya adalah melakukan pengujian dan evaluasi untuk memastikan bahwa model tersebut sesuai dengan tujuan bisnis serta memiliki kinerja yang optimal. Proses evaluasi dilakukan dengan menggunakan teknik statistik maupun metode validasi, seperti pengukuran akurasi, presisi, recall, atau koefisien evaluasi lainnya. Tahap evaluasi ini bertujuan untuk menilai sejauh mana model mampu menjawab permasalahan yang telah dirumuskan pada tahap awal, sekaligus menentukan apakah model tersebut layak untuk dilanjutkan ke tahap implementasi atau masih memerlukan penyempurnaan lebih lanjut.
- (f) **Implementasi (Deployment)**: Tahap terakhir adalah deployment, yaitu mengimplementasikan model ke dalam lingkungan nyata sebagai bagian dari aktivitas organisasi atau perusahaan. Pada tahap ini, model yang telah diuji dan dievaluasi diterapkan agar dapat digunakan secara operasional. Proses implementasi dapat mencakup pengkodean model ke dalam aplikasi, integrasi dengan sistem yang sudah ada, serta pelatihan personel yang akan menggunakan atau mengelola model tersebut. Tujuan utama dari tahap ini adalah memastikan bahwa hasil data mining dapat memberikan nilai nyata bagi organisasi dan mendukung pengambilan keputusan secara efektif.

3. Sistem Pendukung Keputusan

Konsep Sistem Pendukung Keputusan (SPK) pertama kali diperkenalkan pada tahun 1970 oleh Michael S. Scott-Morton dengan istilah *Management Decision System*. Sejak saat itu, sejumlah perusahaan dan lembaga akademik mulai melakukan penelitian serta mengembangkan konsep SPK tersebut. Pada dasarnya, SPK dirancang untuk membantu proses pengambilan keputusan pada berbagai tingkatan, baik dalam mengidentifikasi permasalahan, memilih data yang relevan, maupun menentukan pendekatan yang tepat dalam proses pengambilan keputusan (Setiyaningsih, 2015, p. 4).

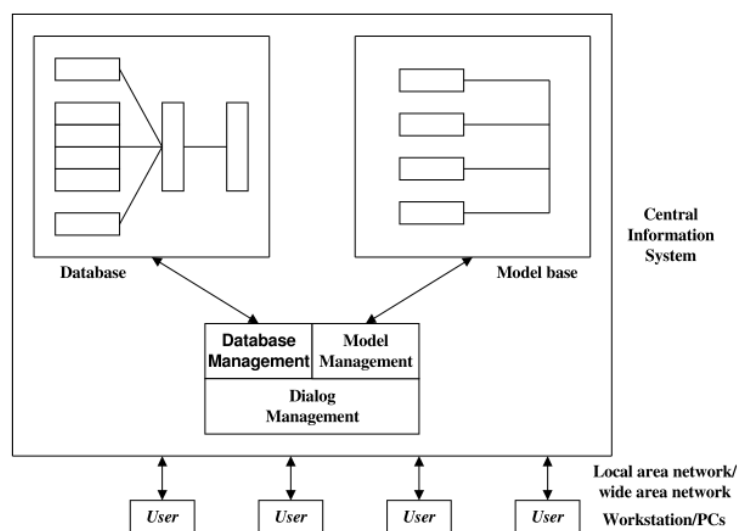
Menurut setyaningsih pada bukunya di halaman 5 pada tahun 2015 ada beberapa pendapat mengenai SPK, antara lain disebutkan di bawah ini:

- (a) Menurut Scott, Sistem Pendukung Keputusan (SPK) didefinisikan sebagai sebuah sistem interaktif berbasis komputer yang dirancang untuk membantu pengambil keputusan dengan memanfaatkan data serta model-model keputusan. Sistem ini berfungsi dalam penyelesaian masalah yang bersifat semi-terstruktur maupun tidak terstruktur, dengan tujuan utama meningkatkan efektivitas dalam proses pengambilan keputusan.
- (b) Menurut Alavi dan Napier, Sistem Pendukung Keputusan (SPK) adalah sekumpulan prosedur pemrosesan data dan informasi yang berbasis pada penggunaan model, dengan tujuan menghasilkan berbagai alternatif jawaban yang dapat mendukung manajemen dalam proses pengambilan keputusan. Sistem ini idealnya bersifat sederhana, mudah digunakan, serta adaptif terhadap kebutuhan pengguna.
- (c) Menurut Little, Sistem Pendukung Keputusan (SPK) merupakan sistem informasi berbasis komputer yang dirancang untuk menghasilkan berbagai alternatif keputusan. Sistem ini berfungsi membantu manajemen dalam menyelesaikan permasalahan yang bersifat semi terstruktur maupun tidak terstruktur dengan memanfaatkan data serta model yang tersedia.
- (d) Menurut Sprague dan Carlson, Sistem Pendukung Keputusan (SPK) merupakan sistem komputer yang berfungsi sebagai pendukung, bukan pengganti, dalam proses pengambilan keputusan. Sistem ini ditujukan untuk membantu menyelesaikan permasalahan yang bersifat semi terstruktur maupun tidak terstruktur melalui pemanfaatan data dan model.
- (e) Menurut Al-Hamdany, Sistem Pendukung Keputusan (SPK) merupakan sistem informasi interaktif yang berfungsi membantu proses pengambilan keputusan dengan cara menyajikan informasi yang disusun secara khusus sesuai dengan pendekatan penyelesaian masalah dan kebutuhan aplikasi para pengambil keputusan. Sistem ini tidak bertindak sebagai pengambil keputusan, melainkan hanya menyediakan dukungan informasi yang relevan.

Dengan demikian, Sistem Pendukung Keputusan (SPK) dapat didefinisikan sebagai sistem berbasis komputer yang terdiri dari berbagai komponen yang saling berinteraksi, termasuk sistem bahasa, sistem pengetahuan, dan sistem penanganan masalah. SPK berfungsi untuk membantu pengambil keputusan dengan mengolah dan menyajikan informasi dari data yang tersedia, sehingga

proses penyelesaian masalah dapat dilakukan dengan lebih cepat, tepat, dan akurat (Yuswardi *et al.*, 2022, p. 3).

Secara umum, Sistem Pendukung Keputusan (SPK) dibangun dari tiga komponen utama, yaitu model dasar, model sistem manajemen basis data (MBMS), dan subsistem dialog atau antarmuka pengguna (User System Interface). Ketiga komponen ini bekerja secara terpadu untuk mendukung pengambilan keputusan yang efektif dengan menyediakan pengolahan data, manajemen model, serta interaksi yang intuitif bagi pengguna. Gambar 2.4 menunjukkan komponen SPK.



Gambar 2. 4 Komponen SPK
Sumber: (Setiyaningsih, 2015, p. 13)

- Model dasar mencakup rutinitas, analisis statistik, keuangan, peramalan, ilmu manajemen, serta model kuantitatif lainnya yang menyediakan kemampuan analitik bagi Sistem Pendukung Keputusan. Komponen ini memungkinkan pengambilan keputusan berbasis data dengan pendekatan kuantitatif yang sistematis.
- Model Sistem Manajemen Basis Data (MBMS) dirancang untuk membangun model menggunakan bahasa pemrograman, subrutin SPK, dan modul pengembangan lainnya. MBMS berperan dalam manipulasi data model serta memungkinkan integrasi model dengan database yang sesuai, sehingga hubungan antara data dan model dapat terkelola dengan baik.
- Subsistem dialog atau antarmuka pengguna (User System Interface) berfungsi sebagai sarana interaksi antara pengguna dan SPK. Melalui sistem ini, pengguna dapat berkomunikasi secara interaktif dengan sistem,

memungkinkan penerapan model dan pengambilan keputusan dilakukan dengan lebih efisien dan responsif.

4. *K-Means Clustering*

a. Clustering

Clustering merupakan salah satu metode penting dalam analisis data yang bertujuan mengelompokkan sekumpulan objek ke dalam klaster berdasarkan kesamaan karakteristiknya. Dalam proses ini, objek-objek yang berada dalam satu klaster memiliki tingkat kesamaan yang lebih tinggi dibandingkan dengan objek di klaster lain. Prinsip dasar dari *clustering* mencakup homogenitas internal (keseragaman di dalam klaster) dan heterogenitas eksternal (perbedaan antara klaster yang satu dengan yang lain) (Rahayu *et al.*, 2018, p. 111).

Clustering adalah teknik fundamental dalam data mining dan machine learning yang digunakan untuk mengelompokkan data berdasarkan kemiripan karakteristik. Algoritma clustering berperan penting dalam mengidentifikasi pola atau struktur yang ada pada data yang belum memiliki label, sehingga memudahkan analisis dan pengambilan keputusan berbasis data (Rahayu *et al.*, 2018, p. 117).

Secara umum, teknik pengelompokan dapat diterapkan pada berbagai bidang, antara lain dalam analisis geospasial untuk segmentasi pelanggan, pengolahan citra untuk pengelompokan objek gambar, serta alokasi wilayah berdasarkan karakteristik tertentu. Dalam konteks penelitian ini, metode clustering digunakan untuk Pemetaan Wilayah Penyebaran Stasiun Pengisian Kendaraan Listrik Umum (SPKLU) berdasarkan parameter seperti frekuensi penggunaan dan data aktual, sehingga dapat mendukung pengambilan keputusan yang lebih tepat dalam pengembangan infrastruktur kendaraan listrik.

b. Algoritma K-Means

Algoritma K-Means adalah metode pengelompokan yang mengelompokkan data berdasarkan kedekatan dengan pusat klaster (*centroid*). Dalam algoritma ini, setiap klaster merupakan sekumpulan objek data yang memiliki karakteristik serupa. Objek-objek dalam satu klaster memiliki tingkat kemiripan yang tinggi satu sama lain, sementara perbedaan dengan objek di klaster lain relatif lebih besar. Dengan demikian, K-Means menekankan prinsip homogenitas internal dan heterogenitas eksternal pada setiap klaster (Remawati, Aji Putra and Irawati, 2021, p. 39).

Algoritma K-Means bekerja dengan membagi data ke dalam beberapa kluster berdasarkan kedekatan atau kesamaan karakteristik. Algoritma ini menggunakan pendekatan iteratif untuk menentukan posisi pusat kluster (*centroid*) yang optimal. Tujuan dari proses ini adalah meminimalkan variasi di dalam setiap kluster sekaligus memaksimalkan perbedaan antar kluster, sehingga pembagian data menjadi lebih representatif dan bermakna.

Langkah-langkah utama dalam algoritma K-Means adalah sebagai berikut:

- (1) Menentukan jumlah kluster K yang diinginkan.
- (2) Memilih K titik sebagai centroid awal secara acak.

$$C_i = \frac{1}{M} \sum_{j=1}^M x_j$$

- (3) Menghitung jarak setiap data ke centroid terdekat menggunakan metrik seperti *Euclidean Distance*:

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

- (4) Mengelompokkan setiap data ke kluster dengan centroid terdekat.
- (5) Menghitung ulang posisi centroid berdasarkan rata-rata data dalam masing-masing kluster:

$$C_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i$$

di mana C_j adalah centroid baru kluster j , dan S_j adalah himpunan data dalam kluster j .

- (6) Mengulangi langkah 3-5 hingga konvergensi tercapai (centroid tidak berubah secara signifikan atau jumlah iterasi maksimum tercapai (Rahayu *et al.*, 2018, p. 126).

Berikut adalah contoh kasus perhitungan K-Means: (Rahayu *et al.*, 2018, pp. 124–130)

Sebagai ilustrasi, sebuah toko dapat melakukan pengelompokan terhadap pelanggan berdasarkan perilaku pembelian mereka ke dalam tiga kategori, yaitu: Banyak, Sedang, dan Sedikit. Pengelompokan ini didasarkan pada variabel Total Belanja dan Frekuensi Pembelian, sebagaimana ditunjukkan pada Tabel 2.1 berikut.

Tabel 2. 1 Total belanja dalam ribuan

No	Total Belanja	Frekuensi Pembelian
1	8270	12
2	1860	7
3	6390	4
4	6191	9
5	6734	3
6	7265	5
7	1466	3
8	5426	7
9	6578	5
10	9322	9
11	2685	7
12	1769	2
13	7949	4
14	3433	9
15	6311	12
16	6051	14
17	7420	2
18	2184	10
19	5555	9
20	4385	10
21	7396	5
22	9666	2
23	3558	4
24	8849	12
25	3047	12
26	3747	7
27	1189	12
28	3734	13
29	4005	8
30	5658	3

Sumber: (Rahayu *et al.*, 2018, p. 124)

Untuk melakukan perhitungan manual algoritma K-Means dengan jumlah klaster sebanyak 3 pada dataset yang berisi 30 pelanggan, langkah-langkah berikut perlu diikuti secara sistematis.

1. Inisialisasi Centroid;

Langkah pertama adalah memilih tiga titik awal sebagai centroid. Pemilihan ini dapat dilakukan secara acak atau berdasarkan pertimbangan tertentu dari dataset untuk mewakili distribusi awal klaster.

C1 (1189; 2)

C2 (5606.5; 7)

C3 (9666; 14)

2. Penugasan Klaster;

Selanjutnya, untuk setiap titik data, hitung jaraknya terhadap masing-masing centroid yang telah ditentukan. Kemudian, tetapkan setiap titik data ke klaster yang memiliki centroid terdekat berdasarkan hasil perhitungan jarak tersebut.

Jarak data pelanggan pertama dengan pusat *Cluster* pertama

$$d_{11} = \sqrt{(8270 - 1189)^2 + (12 - 2)^2} = 7081.01$$

Jarak data pelanggan pertama dengan pusat *Cluster* kedua

$$d_{12} = \sqrt{(8270 - 5606.5)^2 + (12 - 7)^2} = 2663.50$$

Jarak data pelanggan pertama dengan pusat *Cluster* ketiga

$$d_{13} = \sqrt{(8270 - 9666)^2 + (12 - 14)^2} = 1396.00$$

Hasil perhitungan secara lengkap dan posisi *cluster* pada iterasi pertama seperti pada tabel 2.2 berikut.

Tabel 2. 2 Iterasi Pertama

No	Total Belanja	Frekuensi Pembelian	C1	C2	C3	Cluster
1	8270	12	7081,01	2663,5	1396	C3
2	1860	7	671,02	3746,5	7906	C1
3	6390	4	5201	783,51	3276,02	C2
4	6191	9	5002	584,5	3475	C2
5	6734	3	5545	1127,51	2932,02	C2
6	7265	5	6076	1658,5	2401,02	C2
7	1466	3	277	4140,5	8200,01	C1
8	5426	7	4237	180,5	4240,01	C2
9	6578	5	5389	971,5	3068,01	C2
10	9322	9	8133	3715,5	344,04	C3
11	2685	7	1496,01	2921,5	6981	C1
12	1769	2	580	3837,5	7817,01	C1
13	7949	4	6760	2342,5	1717,03	C3
14	3433	9	2244,01	2173,5	6233	C2
15	6311	12	5122,01	704,52	3355	C2

No	Total Belanja	Frekuensi Pembelian	C1	C2	C3	Cluster
16	6051	14	4862,01	444,56	3615	C2
17	7420	2	6231	1813,51	2246,03	C2
18	2184	10	995,03	3422,5	7482	C1
19	5555	9	4366,01	51,54	4111	C2
20	4385	10	3196,01	1221,5	5281	C2
21	7396	5	6207	1789,5	2270,02	C2
22	9666	2	8477	4059,5	12	C3
23	3558	4	2369	2048,5	6108,01	C2
24	8849	12	7660,01	3242,5	817	C3
25	3047	12	1858,03	2559,5	6619	C1
26	3747	7	2558	1859,5	5919	C2
27	1189	12	10	4417,5	8477	C1
28	3734	13	2545,02	1872,51	5932	C2
29	4005	8	2816,01	1601,5	5661	C2
30	5658	3	4469	51,66	4008,02	C2

Perbarui Centroid;

Langkah berikutnya adalah menghitung ulang posisi centroid untuk setiap klaster dengan menggunakan rata-rata dari semua titik data yang telah ditetapkan pada klaster tersebut. Perhitungan ini memastikan bahwa centroid mencerminkan posisi tengah dari anggota klaster secara akurat.

Pada cluster pertama terdapat 7 data, yaitu data ke-2, 7, 11, 12, 18, 25, dan 27. Sehingga $C11 = (1860 + 1466 + 2685 + 1769 + 2184 + 3047 + 1189)/7 = 2028.57$

$$C12 = (7 + 3 + 7 + 2 + 10 + 12 + 12)/7 = 7.57$$

Pada cluster kedua terdapat 18 data, yaitu data ke-3, 4, 5, 6, 8, 9, 14, 15, 16, 17, 19, 20, 21, 23, 26, 28, 29 dan 30. Sehingga $C21 = (6390 + 6191 + 6734 + 7265 + 5426 + 6578 + 3433 + 6311 + 6051 + 7420 + 5555 + 4385 + 7396 + 3558 + 3747 + 3734 + 4005 + 5628)/18 = 5546.50$

$$C22 = (4 + 9 + 3 + 5 + 7 + 5 + 9 + 12 + 14 + 2 + 9 + 10 + 5 + 4 + 7 + 13 + 8 + 3)/18 = 7.17$$

Pada cluster ketiga terdapat 5 data, yaitu data ke-1, 10, 13, 22, dan 24. Sehingga $C31 = (8270 + 9322 + 7949 + 9666 + 8849)/5 = 8811.20$

$$C32 = (12 + 9 + 4 + 2 + 12)/5 = 7.80$$

3. Iterasi;

Proses iteratif dilakukan dengan mengulangi langkah penghitungan jarak dan pembaruan centroid hingga tidak terjadi perubahan lagi pada penugasan klaster. Pada studi kasus ini, konvergensi dicapai pada iterasi keempat, di mana seluruh titik data telah stabil pada klasternya masing-masing. Hasil perhitungan lengkap beserta posisi centroid pada iterasi keempat disajikan pada Tabel 2.3 berikut.

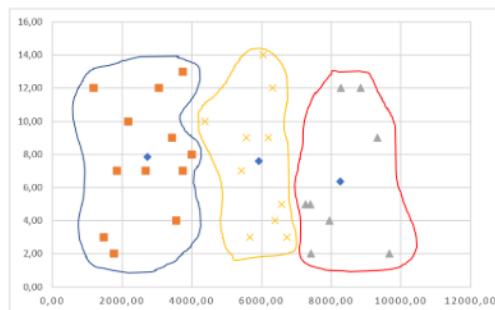
Tabel 2. 3 Hasil Iterasi ke Empat

No	Total Belanja	Frekuensi Pembelian	C1	C2	C3	Cluster
1	8270	12	5546,92	2342,1	6,32	C3
2	1860	7	863,08	4067,9	5407,13	C1
3	6390	4	3666,92	462,11	1877,13	C2
4	6191	9	3467,92	263,1	2076,13	C2
5	6734	3	4010,92	806,11	1533,13	C2
6	7265	5	4541,92	1337,1	1002,13	C3
7	1466	3	1257,09	4461,9	6801,13	C1
8	5426	7	2702,92	501,9	2841,13	C2
9	6578	5	3854,92	650,11	1689,13	C2
10	9322	9	6598,92	3394,1	1054,88	C1
11	2685	7	38,09	3242,9	5581,13	C1
12	1769	2	954,1	4158,9	6498,13	C1
13	7949	4	5225,92	2021,1	318,13	C3
14	3433	9	709,92	2494,9	4834,13	C1
15	6311	12	3587,92	383,13	1956,13	C2
16	6051	14	3327,92	123,27	2216,14	C2
17	7420	2	4696,92	1492,11	847,14	C3
18	2184	10	539,09	3743,9	6083,13	C1
19	5555	9	2831,92	372,9	2712,13	C2
20	4385	10	1661,92	1542,9	3682,13	C2
21	7396	5	4672,92	1468,1	871,13	C3
22	9666	2	6942,92	3738,1	1398,88	C3
23	3558	4	834,93	2369,9	4709,13	C1
24	8849	12	6125,92	2921,1	581,9	C3
25	3047	12	323,94	2800,9	5220,13	C1
26	3747	7	1023,92	2180,9	4520,13	C1
27	1189	12	154,09	4738,9	7047,13	C1
28	3734	13	1010,93	2193,91	4533,13	C1
29	4005	8	1281,92	1922,9	4262,13	C1
30	5658	3	2934,92	269,94	2609,13	C2

Dari hasil cluster diperoleh kelompok sebagai berikut.

- (a) Klaster pertama memiliki pusat (*centroid*) pada koordinat (2723,08; 7,83). Klaster ini merepresentasikan kelompok pelanggan dengan tingkat belanja rendah, dan terdiri dari 12 data.
- (b) Klaster kedua memiliki pusat (*centroid*) pada koordinat (5927,90; 7,60), yang menggambarkan kelompok pelanggan dengan tingkat belanja sedang, terdiri dari 10 data.
- (c) Klaster ketiga memiliki pusat (*centroid*) pada koordinat (8267,13; 6,38), yang menunjukkan kelompok pelanggan dengan tingkat belanja tinggi, terdiri dari 8 data.

Untuk mempermudah pemahaman mengenai distribusi dan karakteristik masing-masing klaster, visualisasi kelompok data pada setiap klaster disajikan pada Gambar 2.5 berikut.



Gambar 2. 5 Hasil Cluster Tiap Kelompok
Sumber: (Rahayu *et al.*, 2018, p. 130)

5. Metode *Silhouette coefficient*

Silhouette Coefficient merupakan salah satu metrik evaluasi internal yang banyak digunakan dalam analisis clustering. Metrik ini mengukur tingkat keserupaan setiap objek dengan klaster tempat objek tersebut berada dibandingkan dengan klaster lainnya. Nilai *Silhouette* berkisar antara -1 hingga 1, di mana nilai yang lebih tinggi menunjukkan kualitas klaster yang lebih baik dan pemisahan antar klaster yang lebih jelas (Rahayu *et al.*, 2018, p. 121). Dengan demikian, *Silhouette Coefficient* merupakan salah satu metode evaluasi yang digunakan untuk menilai kualitas dan kekuatan suatu klaster. Perhitungannya melibatkan dua komponen utama: jarak rata-rata antara suatu objek dengan seluruh objek lain dalam klaster yang sama (*a*) dan jarak rata-rata antara objek tersebut dengan objek di klaster terdekat lainnya (*b*). Nilai-nilai ini kemudian digunakan dalam sebuah formula untuk menghasilkan skor *Silhouette* yang merepresentasikan kesesuaian objek dalam klasternya: (Rahayu *et al.*, 2018, p. 121)

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

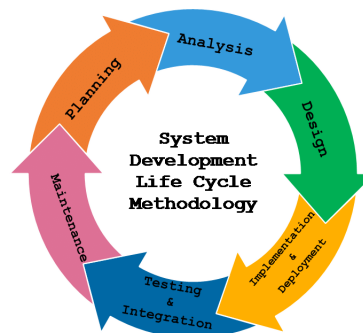
Tabel 2. 4 Tabel Nilai Silhouette Coefficient

Nilai Silhouette Coefficient	Keterangan
Dengan rentang $0.71 < SC \leq 1$	Struktur Kuat
Dengan rentang $0.51 < SC \leq 0.7$	Struktur Baik
Dengan rentang $0.25 < SC \leq 0.5$	Struktur Lemah
Dengan rentang $SC \leq 0.25$	Tidak Terstruktur

Sumber : (Rousseeuw, 1987, p. 55)

6. Systems Development Life Cycle

SDLC (*Software Development Life Cycle*) adalah suatu proses sistematis dalam pengembangan atau modifikasi perangkat lunak, yang memanfaatkan berbagai model dan metodologi untuk merancang, membangun, dan memelihara sistem perangkat lunak secara terstruktur. Proses ini bertujuan untuk memastikan perangkat lunak yang dikembangkan memenuhi kebutuhan pengguna dan berfungsi secara efektif (Hasanah and Untari, 2020, p. 20). Selain itu, SDLC juga dapat dipahami sebagai kerangka kerja atau pola yang digunakan dalam pengembangan sistem perangkat lunak, yang terdiri dari serangkaian tahap seperti yang ditunjukkan pada Gambar 2.6 berikut:



Gambar 2. 6 Konsep SDLC

Sumber: (Niqotaini *et al.*, 2023, p. 8)

- (1) Perencanaan merupakan tahap awal dalam SDLC, di mana dilakukan penyusunan rencana untuk pembangunan atau pengembangan sistem. Tahap ini mencakup identifikasi kebutuhan, tujuan proyek, serta strategi yang akan digunakan untuk memastikan sistem dapat dibangun secara efektif;
- (2) Analisis merupakan tahap di mana kebutuhan yang diperlukan untuk pembangunan sistem dievaluasi secara menyeluruh, termasuk identifikasi

permasalahan dan spesifikasi fungsional sistem;

- (3) Rancangan adalah tahap ketiga, di mana struktur dan arsitektur sistem digambarkan secara rinci, mencakup desain antarmuka, alur proses, dan spesifikasi teknis yang diperlukan;
- (4) Penerapan mencakup semua aktivitas dari tahap perencanaan hingga rancangan yang kemudian direalisasikan dalam pembuatan sistem, termasuk pengkodean dan integrasi komponen;
- (5) Pengujian melibatkan evaluasi sistem yang telah dikembangkan untuk memastikan semua fungsi berjalan sesuai spesifikasi dan untuk mendeteksi adanya kesalahan atau bug;
- (6) Implementasi adalah tahap terakhir, yang dilakukan setelah pengujian selesai dan sistem dianggap bebas dari bug. Pada tahap ini, sistem diterapkan ke lingkungan operasional dan siap digunakan oleh pengguna.

7. Model Prototype

Prototyping adalah suatu proses perancangan di mana dibuat sebuah *prototype* yang merepresentasikan model awal produk. Prototype ini bukanlah tiruan produk final, namun memuat fitur-fitur utama yang akan diuji sebelum produk akhir dikembangkan sesuai fungsinya. Metode *prototyping* memungkinkan pengembang dan pengguna bekerja secara kolaboratif selama proses pembuatan produk, sehingga dapat melakukan evaluasi dan penyempurnaan sebelum tahap produksi final (Hasanah and Untari, 2020, p. 23).

Dalam pengembangan perangkat lunak, *prototyping* merupakan salah satu metode siklus hidup sistem yang didasarkan pada konsep *working model*. Tujuan utamanya adalah mengubah model awal tersebut menjadi sistem akhir yang siap digunakan. Pendekatan ini diharapkan dapat mempercepat proses pembangunan sistem sekaligus menekan biaya pengembangan secara relatif lebih rendah dibandingkan metode konvensional.



Gambar 2. 7 Ilustrasi Model Prototype menurut Roger S. Pressman
Sumber: (Hasanah and Untari, 2020, p. 24)

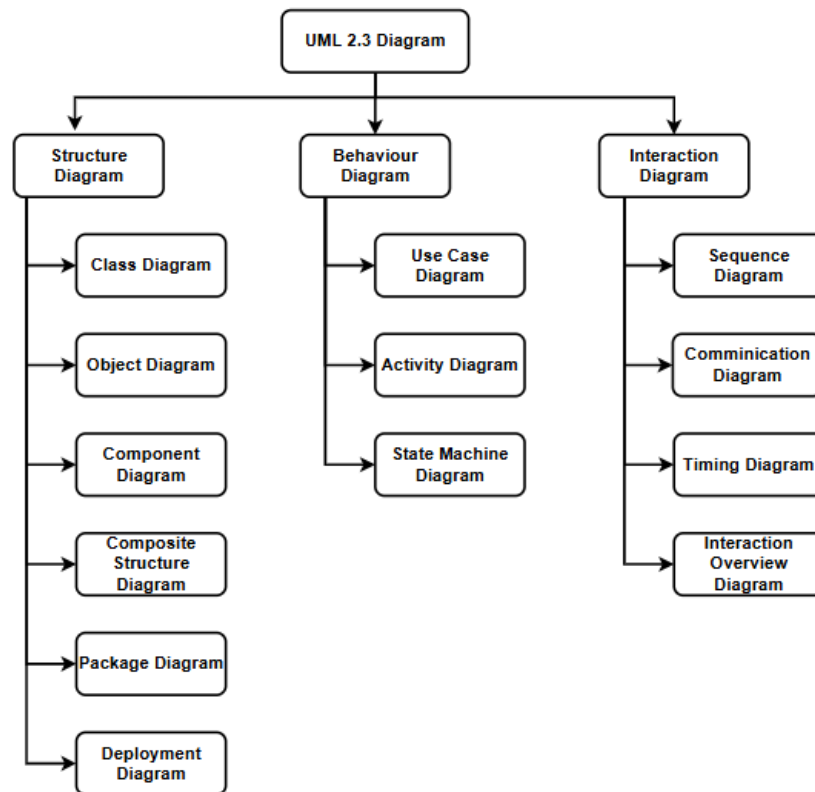
Menurut (Hasanah and Untari, 2020, p. 24) berikut adalah tahapan pengembangan model prototype yaitu:

- (1) Mendengarkan pelanggan, Pada tahap ini, dilakukan pengumpulan kebutuhan sistem dengan memperhatikan masukan dan keluhan dari pengguna. Untuk membangun sistem yang sesuai, penting memahami terlebih dahulu bagaimana sistem yang sedang berjalan serta mengidentifikasi permasalahan yang ada.
- (2) Merancang dan Membuat Prototype, tahap ini mencakup proses merancang dan membangun *prototype* sistem berdasarkan kebutuhan yang telah dikumpulkan dari masukan atau keluhan pengguna. Prototype yang dihasilkan merepresentasikan fitur-fitur utama yang diharapkan ada pada sistem akhir.
- (3) Uji coba Pada tahap ini, *Prototype* yang telah dibuat diuji coba oleh pengguna atau pelanggan untuk menilai kesesuaian dengan kebutuhan mereka. Berdasarkan hasil uji coba, dilakukan evaluasi terhadap kekurangan atau kebutuhan tambahan, kemudian kembali ke tahap pengumpulan masukan pengguna untuk memperbaiki dan menyempurnakan *prototype* yang ada.

8. UML (*Unified Modeling Language*)

Menurut (Hasanah and Untari, 2020, p. 64) UML (*Unified Modeling Language*) merupakan salah satu standar bahasa yang banyak digunakan dalam industri untuk mendefinisikan *requirement* dalam pemrograman berorientasi objek. UML berfungsi untuk mendukung proses analisis dan perancangan, sekaligus menggambarkan arsitektur sistem secara visual. Kehadiran UML muncul untuk memenuhi kebutuhan pemodelan visual yang memungkinkan spesifikasi, deskripsi, pembangunan, dan dokumentasi suatu sistem perangkat lunak. Pendekatan analisis dan perancangan berbasis *object-oriented* mulai diperkenalkan pada era 1970-an hingga pertengahan 1980-an, seiring dengan meningkatnya kompleksitas aplikasi perangkat lunak.

Menurut (Hasanah and Untari, 2020, p. 65) Pada versi 2.3, UML (*Unified Modeling Language*) mencakup 13 jenis diagram yang dikelompokkan ke dalam tiga kategori utama. Rincian diagram tersebut dapat dilihat pada Gambar 2.8 berikut:



Gambar 2. 8 Bagan UML versi 2.3
Sumber : (Hasanah and Untari, 2020, p. 66)

Berdasarkan Gambar 2.8, berikut disajikan penjelasan singkat mengenai pembagian kategori diagram UML tersebut:

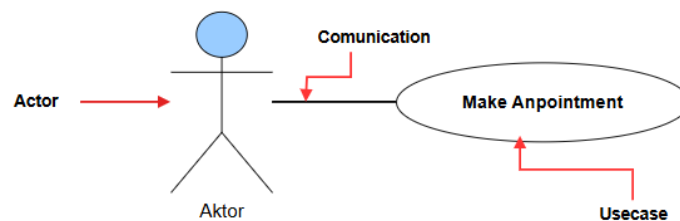
- (a) Structure Diagram merupakan sekumpulan diagram yang digunakan untuk memodelkan struktur statis dari sistem yang sedang dikembangkan.
- (b) Behavior Diagram adalah kumpulan diagram yang digunakan untuk memvisualisasikan perilaku sistem atau rangkaian perubahan yang terjadi dalam sistem tersebut.
- (c) Interaction Diagram merupakan kumpulan diagram yang menggambarkan interaksi antara sistem dengan sistem lain atau interaksi antar subsistem di dalam suatu sistem.

Secara keseluruhan, UML (Unified Modeling Language) menyediakan berbagai diagram yang dapat digunakan dalam perancangan aplikasi berorientasi objek:

a. Use Case Diagram

Menurut (Hasanah and Untari, 2020, p. 71) mengemukakan bahwa Use Case adalah teknik yang digunakan untuk mendokumentasikan persyaratan fungsional suatu sistem serta menggambarkan fungsionalitas

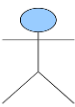
yang diharapkan dari sistem tersebut. Diagram Use Case menekankan apa yang dilakukan oleh sistem, bukan bagaimana cara sistem melakukannya. Setiap *use case* merepresentasikan interaksi antara aktor dan sistem. Manfaat dari Use Case antara lain adalah untuk memfasilitasi komunikasi rancangan sistem dengan klien serta merancang *test case* untuk seluruh fitur yang ada pada sistem. Komponen-komponen Use Case dapat dilihat pada Gambar 2.9 berikut:

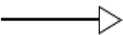


Gambar 2. 9 Komponen Use Case
Sumber : (Hasanah and Untari, 2020, p. 72)

Menurut Fitria Nur Hasanah dan Sri Untari (2020), terdapat beberapa komponen yang membentuk Use Case Diagram. Rincian komponen tersebut disajikan pada Tabel 2.5 berikut;

Tabel 2. 5 Komponen Use Case Diagram

Simbol	Keterangan	Deskripsi
 Aktor	Aktor	Aktor merepresentasikan entitas baik berupa individu, perangkat, atau sistem lain yang berinteraksi dengan sistem yang sedang dimodelkan.
 Nama Use	Use Case	Use Case merupakan representasi fungsional dari suatu sistem yang bertujuan agar pengguna atau pelanggan dapat memahami kegunaan dan fungsi sistem yang akan dikembangkan
 Asosiasi / association	Asosiasi	<i>Association</i> berfungsi untuk menghubungkan atau menciptakan relasi antar elemen dalam Use Case Diagram.

Simbol	Keterangan	Deskripsi
Generalisasi / <i>generalization</i> 	Generalisasi	<i>Generalization</i> , yang juga dikenal sebagai <i>inheritance</i> (pewarisan), adalah konsep di mana sebuah elemen dapat menjadi spesialisasi atau turunan dari elemen lain dalam Use Case Diagram.
Include/uses <<include>> 	<<include>>	<<include>> adalah relasi yang menunjukkan perilaku yang harus dipenuhi agar suatu peristiwa (<i>event</i>) dapat terjadi, di mana dalam kondisi ini sebuah <i>use case</i> menjadi bagian dari <i>use case</i> lainnya.
Ekstensi / <i>extend</i> <<extend>> 	<<extend>>	<<extend>> adalah relasi yang menunjukkan perilaku yang hanya dijalankan dalam kondisi tertentu, contohnya seperti mengaktifkan alarm.

Sumber: (Hasanah and Untari, 2020, pp. 70–76)

b. Activity Diagram

Menurut (Hasanah and Untari, 2020, p. 79) Activity Diagram digunakan untuk memvisualisasikan aliran aktivitas dalam sistem yang sedang dirancang, mulai dari titik awal, keputusan (*decision*) yang mungkin terjadi, hingga titik akhir setiap aliran. Diagram ini juga mampu menggambarkan proses paralel yang berlangsung secara bersamaan pada beberapa eksekusi. Activity Diagram merupakan bentuk khusus dari *state diagram*, di mana sebagian besar *state* merepresentasikan tindakan (*action*) dan transisinya dipicu oleh selesainya *state* sebelumnya (proses internal). Dengan demikian, Activity Diagram tidak menggambarkan perilaku internal sistem atau interaksi antar subsistem secara detail, melainkan memfokuskan pada proses dan jalur aktivitas secara keseluruhan dari level atas.

Activity diagram digunakan untuk memvisualisasikan langkah-langkah atau aktivitas yang terjadi dalam suatu sistem. Setiap *use case*

minimal memiliki satu Activity Diagram yang menggambarkan proses bisnis dan urutan aktivitas di dalamnya. Diagram ini sering digunakan dalam *business modeling* untuk menampilkan alur aktivitas proses bisnis secara jelas. Struktur Activity Diagram menyerupai *flowchart* atau *Data Flow Diagram (DFD)* pada pendekatan perancangan terstruktur (Hasanah and Untari, 2020, p. 80).

Activity Diagram dibuat berdasarkan satu atau beberapa *use case* dalam *Use Case Diagram*. Diagram ini memvisualisasikan aliran aktivitas dalam sistem yang sedang dirancang, termasuk titik awal, kemungkinan keputusan, serta titik akhir setiap aliran, sekaligus menggambarkan proses yang dapat berlangsung secara paralel pada beberapa eksekusi. Activity Diagram tidak menampilkan perilaku internal sistem secara rinci, melainkan lebih menekankan pada proses dan jalur aktivitas secara keseluruhan dari level atas (Hasanah and Untari, 2020, p. 80).

Menurut Fitria Nur Hasanah dan Sri Untari (2020), terdapat beberapa simbol yang digunakan dalam pembuatan Activity Diagram, yang dapat dijelaskan lebih rinci pada Tabel 2.6;

Tabel 2. 6 Diagram Activity

Simbol	Keterangan	Deskripsi
	Status awal	Initial State pada Activity Diagram merupakan titik awal atau kondisi permulaan dari suatu aktivitas dalam sistem.
	Aktivitas	Aktivitas merupakan suatu tindakan atau kegiatan yang dijalankan oleh sistem, yang umumnya dinyatakan menggunakan kata kerja.
	Percabangan / <i>decision</i>	Decision menggambarkan percabangan dalam aliran aktivitas, yaitu saat terdapat lebih dari satu jalur aktivitas yang dapat dipilih.
	Penggabungan	Penggabungan menunjukkan

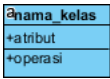
Simbol	Keterangan	Deskripsi
	<i>/ join</i>	Penggabungan dua atau lebih aliran aktivitas menjadi satu jalur tunggal.
	Status akhir	Status akhir pada Activity Diagram menunjukkan titik akhir atau kondisi akhir dari suatu aktivitas yang dijalankan oleh sistem.
 Atau 	Swimlane	Swimlane digunakan untuk memisahkan entitas bisnis atau aktor yang memiliki tanggung jawab atas pelaksanaan aktivitas tertentu dalam sistem.

Sumber : (Hasanah and Untari, 2020, p. 82)

c. Class Diagram

Menurut (Zein *et al.*, 2016, p. 124) Class Diagram digunakan untuk memodelkan struktur objek dalam sistem perangkat lunak. Diagram ini menggambarkan kelas, atribut, metode, serta hubungan yang ada antar kelas. Simbol-simbol yang digunakan beserta deskripsinya dapat dilihat pada tabel 2.7 berikut:

Tabel 2. 7 Class Diagram

Simbol	Keterangan	Deskripsi
	Kelas	Kelas merepresentasikan entitas atau komponen dalam kerangka struktur sistem.
	Antarmuka / <i>Interface</i>	Antarmuka merepresentasikan konsep antarmuka dalam pemrograman berorientasi objek.

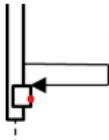
Simbol	Keterangan	Deskripsi
	Asosiasi / <i>Association</i>	Asosiasi menggambarkan hubungan antar kelas yang memiliki makna umum, dan biasanya dilengkapi dengan keterangan <i>multiplicity</i> .
	Asosiasi berarah / <i>Directed Association</i>	Asosiasi berarah menunjukkan hubungan antar kelas di mana satu kelas bergantung pada kelas lain, dan biasanya disertai dengan keterangan <i>multiplicity</i> pada asosiasinya.
	<i>Generalisasi</i>	<i>Generalisasi</i> menggambarkan hubungan antar kelas yang menunjukkan konsep umum (superclass) dan konsep spesifik (subclass), atau hubungan generalisasi-spesialisasi.
	Kebergantungan/ <i>Dependency</i>	Menggambarkan hubungan antar kelas yang menandakan adanya ketergantungan satu kelas terhadap kelas lainnya.
	Agregasi/ <i>Aggregation</i>	Merepresentasikan hubungan antar kelas yang menggambarkan konsep "whole-part" atau hubungan keseluruhan-bagian.

Sumber : (Zein *et al.*, 2016, p. 125)

d. **Sequence Diagram**

Menurut (Zein *et al.*, 2016, p. 126) sequence diagram berfungsi untuk memodelkan urutan interaksi antar objek dalam suatu sistem perangkat lunak. Diagram ini menampilkan cara objek saling berinteraksi dan bertukar pesan. Simbol-simbol yang digunakan dalam Sequence Diagram beserta penjelasannya dapat dilihat pada Tabel 2.8:

Tabel 2. 8 Sequence Diagram

Simbol	Keterangan	Deskripsi
	<i>Entity Class</i>	Merupakan komponen utama dalam sistem yang berfungsi sebagai dasar untuk perancangan basis data.
	<i>Boundary Class</i>	Terdiri dari sekumpulan kelas yang berperan dalam memfasilitasi interaksi antara aktor dan sistem.
	<i>Control Class</i>	Objek kontrol adalah objek yang menyimpan logika aplikasi dan berfungsi mengatur alur proses, namun tidak memiliki tanggung jawab terhadap entitas data.
	<i>Lifeline</i>	Garis yang menghubungkan objek dengan aktivasi yang terjadi sepanjang garis waktu hidup (lifeline) objek tersebut.
	<i>Recursive</i>	Pesan yang dikirim oleh suatu objek kepada dirinya sendiri untuk memicu aksi atau metode internal.
	<i>Activation</i>	Menggambarkan pelaksanaan suatu operasi oleh objek, dimana panjang kotak menunjukkan durasi aktivitas operasi tersebut.

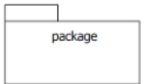
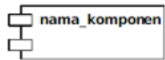
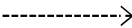
Simbol	Keterangan	Deskripsi
	Message	Jenis pesan di mana objek mengirimkan pesan kepada dirinya sendiri untuk mengeksekusi suatu operasi atau metode internal.

Sumber : (Zein *et al.*, 2016, p. 127)

e. Komponen Diagram

Komponen diagram digunakan untuk memodelkan bagaimana komponen-komponen sistem bekerja bersama untuk mencapai tujuan tertentu. Diagram ini menunjukkan struktur komponen dan interaksi antara komponen-komponen (Zein *et al.*, 2016, p. 131). Simbol-simbol yang digunakan dalam komponen diagram beserta penjelasannya dapat dilihat pada tabel 2.9.

Tabel 2. 9 Komponen Diagram

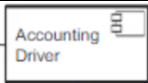
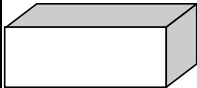
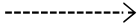
Simbol	Keterangan	Deskripsi
	Package	Package merupakan suatu bungkusan dari satu atau lebih node.
	Komponen	Komponen Sistem.
	Defedency/ Kebergantungan	Kebergantungan antar komponen, arah panah mengarah pada komponen yang dipakai.
	Interface/ Antarmuka	Sebagai antarmuka komponen agar tidak mengakses langsung komponen.
	Link	Relasi antar komponen.

Sumber: (Ibnu Akil, 2015, pp. 56–59)

f. Deployment Diagram

Deployment diagram digunakan untuk memvisualisasikan hubungan antara software dan hardware. Secara spesifik deployment diagram dapat membuat physical model tentang bagaimana komponen perangkat lunak digunakan pada komponen perangkat keras, yang dikenal sebagai node (Zein *et al.*, 2016, p. 132). Simbol-simbol yang digunakan dalam *deployment* diagram beserta penjelasannya dapat dilihat pada tabel 2.10.

Tabel 2. 10 Deployment Diagram

Simbol	Keterangan	Deskripsi
	<i>Package</i>	Merupakan sebuah bungkus dari satu atau lebih node.
	<i>Node</i>	Biasanya mengacu pada perangkat keras (hardware), perangkat lunak yang tidak dibuat sendiri (software), jika didalam node disertakan komponen untuk mengkonsistenkan rancangan maka komponen yang diikuti sertakan harus sesuai dengan komponen yang telah didefinisikan sebelumnya pada diagram komponen.
	Kebergantungan / <i>dependency</i>	Kebergantungan antar node, arah panah mengarah pada node yang dipakai.
	Link	Relasi antar node.

Sumber (Zein *et al.*, 2016, p. 132)

9. Database

Data merupakan fakta atau informasi yang mencerminkan kejadian nyata pada suatu waktu tertentu. Data ini diperoleh dari aktivitas aktual, seperti transaksi penjualan, pembelian, dan sejenisnya. Data berperan sebagai rekaman dari transaksi yang terjadi di suatu organisasi, misalnya melalui tanda terima, faktur, atau formulir. Setelah diproses, data dapat diolah menjadi berbagai bentuk

laporan, seperti laporan penjualan atau dokumen resmi lainnya (Kurniati *et al.*, 2015, p. 3).

Basis data adalah sekumpulan informasi yang tersimpan dalam jangka waktu yang panjang, bahkan bertahun-tahun. Basis data memegang peranan penting bagi setiap organisasi atau bisnis. Saat ini, hampir semua perusahaan, baik besar maupun kecil, menggunakan basis data untuk menyimpan seluruh informasi mereka. Kekuatan basis data terletak pada pengetahuan dan teknologi yang diterapkan melalui perangkat lunak khusus yang dikenal sebagai Sistem Manajemen Basis Data (Database Management System/DBMS), atau dalam dunia IT sering disebut Sistem Basis Data.

10. Mysql

MySQL mulai dikembangkan pada pertengahan 1990an dan kini telah memiliki lebih dari 10 juta instalasi. MySQL adalah sebuah Sistem Manajemen Basis Data (DBMS) yang menggunakan bahasa SQL untuk menghubungkan aplikasi perangkat lunak dengan basis data. MySQL termasuk salah satu DBMS yang populer digunakan karena bersifat open source, mendukung integrasi dengan berbagai bahasa pemrograman, tidak membutuhkan memori RAM yang besar, dan mampu melayani banyak pengguna secara bersamaan (Silalahi, 2022, p. iii).

11. Web Server

Dalam pengembangan sistem berbasis situs web, diperlukan perangkat untuk menjalankan server web. Web server berfungsi sebagai perangkat lunak pada server yang menerima permintaan (request) halaman web melalui protokol HTTP atau HTTPS dari klien, seperti browser web, dan kemudian mengirimkan kembali (response) hasilnya dalam bentuk halaman web, biasanya berupa dokumen HTML (Rerung, 2018, p. 9).

12. Bahasa Pemrograman

Bahasa pemrograman merupakan sistem notasi yang dipakai untuk menulis program komputer. Bahasa ini terbagi menjadi tiga tingkatan, yaitu bahasa mesin, bahasa tingkat rendah, dan bahasa tingkat tinggi (Sohail Aslam, Maqsood Ahmad, 2021, p. 30).

a. Python

Python adalah bahasa pemrograman tingkat tinggi yang dibuat oleh Guido van Rossum pada tahun 1991. Bahasa ini dirancang dengan prinsip keterbacaan kode dan sintaks yang sederhana, sehingga menjadi pilihan populer baik bagi pemula maupun pengembang berpengalaman (Maesaroh *et al.*, 2024, p. 1).

b. Pengertian Hypertext Markup Language (HTML)

Dalam pembuatan sebuah website, HTML memegang peranan penting sebagai bahasa yang umum digunakan untuk membuat halaman web. Hypertext Markup Language (HTML) merupakan bahasa standar untuk membangun website yang dapat diakses melalui internet. Dengan HTML, konten halaman web disusun sedemikian rupa sehingga komputer dapat menerjemahkannya agar dapat ditampilkan dan dipahami oleh pengguna. HTML menjadi standar luas dalam pembuatan website agar tampilan laman dapat terlihat dengan baik di layar komputer (Muhardian, 2016 dalam Rerung, 2018, p. 19).

B. Tinjauan Pustaka

Penelitian terdahulu berperan sebagai landasan bagi peneliti dalam melaksanakan penelitian. Pada studi ini, referensi penelitian dipilih berdasarkan kesamaan metode yang digunakan, yaitu Algoritma K-Means. Metode tersebut telah banyak diterapkan dalam berbagai kasus untuk melakukan pengelompokan data secara efektif. Sebagai berikut:

(a) Penerapan Algoritma K-Means Untuk Pemetaan Penerimaan Bantuan Kesejahteraan Masyarakat di Kota Batam (Azwanti and Putria, 2023). Tahun 2020 menjadi masa yang penuh tantangan bagi seluruh manusia di dunia akibat merebaknya wabah virus Coronavirus Disease 2019 (Covid-19). Di Kota Batam, kasus Covid-19 terus meningkat hingga mencapai lebih dari 2.000 kasus, berdasarkan data dari Gugus Tugas Covid-19 Provinsi Kepulauan Riau. Menanggapi situasi ini, pemerintah melalui Dinas Sosial memberikan bantuan kesejahteraan bagi masyarakat terdampak. Namun, dari hasil pengamatan peneliti, banyak bantuan tersebut belum tepat sasaran sehingga menimbulkan permasalahan di tingkat masyarakat. Penelitian ini menggunakan salah satu teknik data mining, yaitu algoritma K-Means, untuk melakukan pemetaan atau segmentasi bantuan pemerintah bagi masyarakat terdampak Covid-19. Tujuannya adalah mempercepat distribusi daya beli masyarakat serta memberikan informasi yang dapat dimanfaatkan oleh pemuka masyarakat seperti RT dan RW dalam mendata warga yang membutuhkan bantuan. Hasil analisis menunjukkan bahwa dari 30 data yang dianalisis, terdapat segmentasi penerima bantuan yang dibagi ke dalam lima klaster berdasarkan kriteria yang telah ditetapkan.

(b) Penerapan Metode K-Means Clustering Pada Pemetaan Lahan Kopi Di Kabupaten Malang (Alif Fajar Fadhillah, Faisol and Vendyansyah, 2022). Indonesia merupakan negara agraris dengan posisi sebagai produsen kopi

terbesar keempat di dunia. Komoditas kopi memiliki peran strategis, baik sebagai sumber pendapatan bagi petani, sumber devisa negara, bahan baku industri, maupun penyedia lapangan kerja melalui pengolahan, pemasaran, dan perdagangan. Salah satu faktor kunci dalam produksi tanaman pangan adalah ketersediaan lahan. Penelitian ini bertujuan untuk melakukan pemetaan lahan tanaman kopi di Kabupaten Malang. Latar belakang penelitian ini muncul karena pemetaan lahan kopi di Kabupaten Malang selama ini masih dilakukan secara manual berdasarkan tahun, luas lahan, dan produksi tahunan. Metode yang digunakan dalam penelitian ini adalah K-Means Clustering dengan data yang diperoleh dari Dinas Tanaman Pangan, Hortikultura, dan Perkebunan Kabupaten Malang. Pemetaan dilakukan dengan tiga kategori klaster, yaitu rendah, sedang, dan tinggi. Hasil penelitian menunjukkan bahwa masih terdapat banyak wilayah di Kabupaten Malang dengan tingkat produksi tanaman kopi yang rendah. Berdasarkan hasil clustering dari 33 kecamatan selama tiga tahun, diperoleh total 99 data dengan distribusi: C1 (Rendah) = 26 data (2018), 24 data (2019), 24 data (2020); C2 (Sedang) = 3 data (2018), 5 data (2019), 4 data (2020); dan C3 (Tinggi) = 4 data (2018), 4 data (2019), 5 data (2020).

- (c) **Penerapan Metode K-Means Clustering untuk Pemetaan Pengelompokan Lahan Produksi Tandan Buah Segar** (Al Masykur *et al.*, 2023). Kelapa sawit merupakan salah satu tanaman pertanian strategis yang dibudidayakan secara luas di Indonesia, memberikan kontribusi signifikan terhadap pendapatan dan lapangan kerja masyarakat setempat. Salah satu contohnya terdapat pada perkebunan Sei Lukut, Desa Maredan Barat, Kecamatan Tualang, Kabupaten Siak, Provinsi Riau, yang dikelola oleh PT Surya Intisari Raya, sebuah perusahaan swasta. Perkebunan ini terdiri dari empat bagian lahan kelapa sawit yang terbagi menjadi 216 blok dengan total luas sekitar 4.000 hektar. Pemetaan lahan kelapa sawit yang tepat sangat penting untuk mendukung pengelolaan lahan, pembangunan berkelanjutan, serta membantu perusahaan dalam menentukan kebijakan yang dapat meningkatkan akurasi dan produktivitas produksi. Dalam penelitian ini, metode K-Means Clustering digunakan untuk memetakan lahan kelapa sawit di PT Surya Intisari Raya, dengan evaluasi jumlah cluster optimal menggunakan Davies-Bouldin Index dan alat RapidMiner. Hasil penelitian menunjukkan bahwa metode K-Means Clustering efektif dalam pemetaan lahan produksi tandan buah segar, dengan tingkat kecocokan sistem mencapai 99,54% pada RapidMiner. Nilai Davies-Bouldin Index terkecil sebesar 0,921 diperoleh dengan jumlah cluster optimal sebanyak tiga, yaitu: Cluster C1 (Produktivitas Sedang) sebanyak 96 blok lahan, Cluster C2 (Produktivitas

Rendah) sebanyak 41 blok lahan, dan Cluster C3 (Produktivitas Tinggi) sebanyak 79 blok lahan. Penelitian ini menegaskan potensi metode K-Means Clustering dalam pemetaan lahan produksi tandan buah segar serta memberikan informasi penting yang dapat dimanfaatkan perusahaan untuk pengambilan keputusan.

(d) Sistem Informasi Geografis Pemetaan Warga Kurang Mampu Di Kelurahan

Karang Besuki Menggunakan Metode K-Means Clustering (Ali Hasyimi, Faisol and Ariwibisono, 2021). Kemiskinan masih menjadi salah satu tantangan utama di negara berkembang, dan bantuan sosial merupakan salah satu upaya untuk menurunkan tingkat kemiskinan tersebut. Namun, dalam praktiknya, masih ditemukan penerima bantuan sosial yang kurang tepat sasaran akibat pemanfaatan data yang belum optimal serta terbatasnya informasi mengenai kondisi dan lokasi geografis, termasuk di Kelurahan Karangbesuki. Untuk mengatasi masalah tersebut, dibutuhkan sistem informasi geografis (SIG) yang mampu menyajikan informasi terkait pemetaan warga kurang mampu. Sistem ini diharapkan dapat mengelompokkan warga berdasarkan tingkat kebutuhan serta memetakan lokasi mereka secara akurat. Sistem yang dikembangkan berupa SIG pemetaan warga kurang mampu dengan penerapan metode clustering yang dikombinasikan dengan QGIS 2.18. Metode clustering yang digunakan adalah K-Means Clustering dengan kriteria pengelompokan data berupa umur, pekerjaan, pendidikan, dan jumlah tanggungan keluarga. Hasil pengujian sistem dengan 325 sampel data dan pengelompokan menjadi tiga cluster menunjukkan bahwa 178 warga (55%) tergolong tidak layak menerima bantuan, 99 warga (30%) kurang layak menerima bantuan, dan 48 warga (15%) layak menerima bantuan. Evaluasi kinerja sistem dengan membandingkan hasil pengelompokan manual (data lama) dengan pengelompokan menggunakan sistem menunjukkan bahwa sistem ini mampu mengklasifikasikan warga kurang mampu dengan baik, menghasilkan tingkat akurasi sebesar 91%.

(e) Pemetaan Penyebaran Covid-19 Pada Tingkat Kabupaten/Kota Di Pulau

Jawa Menggunakan Algoritma K-Means Clustering (Gayatri and Hendry, 2021). Coronavirus atau Covid-19 merupakan virus yang pertama kali ditemukan di Wuhan, China pada Desember 2019 dan memiliki kemampuan penyebaran yang sangat cepat ke berbagai negara, termasuk Indonesia. Kasus pertama Covid-19 di Indonesia tercatat pada 2 Maret 2020, dan jumlah kasus terus meningkat setiap harinya. Pada 27 Maret 2021, jumlah kasus Covid-19 di Indonesia mencapai 1.492.002 kasus. Peningkatan kasus yang signifikan ini menekankan pentingnya pemetaan tingkat kerawanan penyebaran Covid-19, khususnya di Pulau Jawa, dengan memanfaatkan data resmi pemerintah pada

tingkat provinsi. Pemetaan dilakukan menggunakan tiga parameter utama, yaitu jumlah kasus dirawat, jumlah pasien sembuh, dan jumlah kematian. Untuk menentukan jumlah cluster yang optimal, penelitian ini menerapkan algoritma K-Means Clustering dengan evaluasi menggunakan metode Davies-Bouldin Index (DBI).

- (f) **Metode K-Means Untuk Pemetaan Persebaran Usaha Mikro Kecil Dan Menengah** (Remawati, Aji Putra and Irawati, 2021). Perkembangan di era globalisasi saat ini sangat dipengaruhi oleh sektor ekonomi, yang menjadi indikator keberhasilan kebijakan pemerintah. Masyarakat memiliki peran penting dalam pembangunan nasional, khususnya melalui Usaha Mikro, Kecil, dan Menengah (UMKM). Untuk meningkatkan kontribusi UMKM sebagai tolok ukur keberhasilan ekonomi, diperlukan dukungan pemerintah, misalnya melalui pemberian bantuan bagi pelaku usaha yang mengalami keterbatasan modal. Penelitian ini bertujuan untuk mengidentifikasi kelompok usaha masyarakat sebagai indikator tingkat usaha, sehingga memudahkan pemerintah dalam menyalurkan bantuan. Metode K-Means Clustering diterapkan untuk mengelompokkan tingkat usaha berdasarkan data pendapatan masyarakat saat ini. Hasil penelitian berupa sistem pengelompokan tingkat usaha berbasis website yang digunakan oleh Dinas Koperasi dan UKM, dengan kategori usaha mikro, kecil, dan menengah berdasarkan pendapatan atau asset.
- (g) **Analisis Pemetaan Tingkat Kriminalitas di Kabupaten Karawang menggunakan Algoritma K-Means** (Fahmi, Jajuli and Sulistiyowati, 2021). Kriminalitas merupakan salah satu masalah sosial yang perlu mendapatkan perhatian serius karena menimbulkan kerugian dan dampak negatif bagi masyarakat. Berdasarkan laporan dari jabar.tribunnews.com, Kabupaten Karawang menempati peringkat pertama sebagai daerah dengan tingkat kriminalitas tertinggi di Jawa Barat pada awal pandemi. Hal ini menjadi tantangan bagi pemerintah daerah dan Polres Karawang untuk menangani serta menekan angka kriminalitas. Penelitian ini menerapkan metode clustering menggunakan algoritma K-Means untuk melakukan pemetaan wilayah rawan kriminalitas dengan bantuan perangkat lunak QGIS.
- (h) **Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering** (Hidayati *et al.*, 2021). Clustering merupakan proses pengelompokan sekumpulan data ke dalam beberapa klaster berdasarkan tingkat kemiripannya, yang biasanya diukur melalui perhitungan jarak antar data. Dalam penelitian ini, performa berbagai metode perhitungan jarak diuji menggunakan enam set data dengan jumlah atribut yang berbeda, yaitu 2, 3, 4, dan 6 atribut. Hasil evaluasi

menggunakan Silhouette Coefficient menunjukkan bahwa jarak Chebyshev memiliki performa yang stabil, baik pada data dengan atribut sedikit maupun banyak. Sementara itu, Average distance memberikan nilai Silhouette Coefficient tertinggi dibandingkan metode jarak lain pada data yang mengandung outlier, khususnya pada data dengan tiga atribut. Metode Mean Character Difference hanya menunjukkan hasil yang baik pada data dengan atribut sedikit, sedangkan Euclidean distance, Manhattan distance, dan Minkowski distance cenderung memberikan nilai baik pada data dengan atribut sedikit, tetapi nilai Silhouette Coefficient menurun mendekati 0,5 ketika diaplikasikan pada data dengan banyak atribut.

- (i) **Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering** (Hutagalung, 2022). Penelitian ini bertujuan untuk mengelompokkan data siswa kelas unggulan agar proses pembelajaran di sekolah dapat lebih mudah difasilitasi sesuai dengan kemampuan masing-masing siswa. Pengolahan data dalam jumlah besar akan menjadi tidak efektif dan efisien jika tidak menggunakan sistem berbasis komputer, apalagi dalam mengelompokkan siswa dengan prestasi mumpuni di bidang tertentu, yang memerlukan metode khusus dalam data mining. Setiap tahun jumlah siswa terus bertambah, sehingga volume data meningkat dan proses pengolahan menjadi memakan waktu lama serta sulit dioptimalkan. Pengambilan keputusan secara konvensional tidak mampu menangani data yang sangat besar tersebut, sehingga penerapan data mining diperlukan untuk mengidentifikasi pola-pola dari penumpukan data. Dalam penelitian ini, algoritma K-Means diterapkan dengan sampel sebanyak 120 data nilai rapor siswa SMK Raksana 2 Medan, dan penerapannya berbasis web untuk memudahkan pengolahan data, sehingga menjadi alat bantu yang efektif bagi pendidik dalam pengambilan keputusan.
- (j) **Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat** (Pasaribu *et al.*, 2021). Berdasarkan data hasil produksi buah sawit di PTPN IV Marihat, terdapat beberapa lokasi dengan jumlah panen yang bervariasi. Untuk itu, diperlukan pengelompokan lokasi potensial penghasil buah agar dapat mengetahui area mana saja yang menghasilkan buah sawit dalam jumlah banyak maupun sedikit. Pembagian hasil produksi biasanya dilakukan berdasarkan lokasi atau blok panen buah kelapa sawit, sehingga dibutuhkan metode yang dapat mempermudah pengelompokan lokasi penghasil buah. Dengan pendekatan pengklasteran menggunakan algoritma K-Means, pembagian kelompok lokasi dapat dilakukan berdasarkan luas panen (Ha), realisasi produksi (kg), dan tahun panen. Penelitian ini

menerapkan K-Means untuk mengklaster daerah potensial penghasil buah dengan tujuan mempermudah pengelompokan blok berdasarkan tingkat produksi, baik yang tinggi maupun rendah.

Tabel 2. 11 Tinjauan Pustaka

No	Peneliti	Judul	Jurnal	Kontribusi
1	Nurul Azwantia, Narti Eka Putriab	Penerapan Algoritma K-Means Untuk Pemetaan Penerimaan Bantuan kesejahteraan Masyarakat di Kota Batam	Prosiding Seminar Nasional Ilmu Sosial dan Teknologi (SNISTEK) (2023) https://ejournal.upbatam.ac.id/index.php/prosiding/article/view/8092	Berkontribusi Sebagai Acuan Dalam Penerapan K-Means Dalam Pemetaan
2	Dimas Alif Fajar Fadhillah, Ahmad Faisol, Nurlaily Vendyansyah	Penerapan Metode K-Means Clustering Pada Pemetaan Lahan Kopi Di Kabupaten Malang	JATI (Jurnal Mahasiswa Teknik Informatika) Vol. 6 No. (2022) https://ejournal.itn.ac.id/index.php/jati/article/view/4617	Berkontribusi Dalam Pembuatan Prototyping Produk
3	Dina Nuriyah Ramadhani, Aris Trijaka Harjanta, Aptanang Tyogi, Serli Agnes Ajhara	Penerapan Metode K-Means Clustering untuk Pemetaan Pengelompokan Lahan Produksi Tandan Buah Segar	JURNAL INFORMATIKA, Vol.10 No.1. (2023) https://ejournal.bsi.ac.id/ejournal/index.php/ji/article/view/15621	Berkontribusi Dalam Pembuatan Prototyping Produk
4	Muhammad Ali Hasymi, Ahmad Faisol, FX. Ariwibisono	Sistem Informasi Geografis Pemetaan Warga Kurang Mampu Di Kelurahan Karang Besuki Menggunakan Metode K-Means Clustering	Jurnal Mahasiswa Teknik Informatika (JATI), Vol. 5 No. 1, 2021 https://ejournal.itn.ac.id/index.php/jati/article/view/3269	Berkontribusi Dalam Pembuatan Prototyping Produk
5	Lidia Gayatri, Hendry	Pemetaan Penyebaran Covid-19 Pada Tingkat Kabupaten/Kota Di Pulau Jawa Menggunakan Algoritma K-Means Clustering Lidia	Sebatik Vol. 25 No.2 (2021) https://jurnal.wicida.ac.id/index.php/sebatik/article/view/1307	Kontribusinya yaitu untuk menghitung Min-Max Scalling untuk Normalisasi Data

No	Peneliti	Judul	Jurnal	Kontribusi
6	Dicky Jordan Aji Putra, Dwi Remawati, Tri Irawati	Metode K-Means Untuk Pemetaan Persebaran Usaha Mikro Kecil Dan Menengah	Jurnal TIKomSiN, Vol. 9, No. 2, (2021) https://p3m.sinus.ac.id/jurnal/index.php/TIKomSiN/article/view/574	Kontribusinya yaitu untuk menghitung Min-Max Scalling untuk Normalisasi Data
7	Resti Noor Fahmi, Mohamad Jajuli, Nina Sulistiyowati	Analisis Pemetaan Tingkat Kriminalitas Di Kabupaten Karawang Menggunakan Algoritma K-Means	Journal of Information Technology and Computer Science (INTECOMS) Volume 4 Nomor 1, (2021) https://journal.ipm2kpe.or.id/index.php/INTECOM/article/view/2413	Berkontribusi Dalam Penggunaan Metode Data Mining CRISP DM
8	Rahmatina Hidayati, Anis Zubair, Aditya Hidayat Pratama, Luthfi Indana	Analisis Silhouette Coefficient pada 6 Perhitungan Jarak K-Means Clustering	Techno.COM, Vol. 20, No. 2, (2021) https://tunasbangsa.ac.id/pkm/index.php/kesatria/article/view/41	Berkontribusi Sebagai Acuan Dalam Proses Perhitungan Silhouette Coefficient
9	Juniar Hutagalung, Yopi Hendro Syahputra, Zohana Pertiwi Tanjung	Pemetaan Siswa Kelas Unggulan Menggunakan Algoritma K-Means Clustering	Jurnal Teknik Informatika dan Sistem Informasi Vol. 9, No. 1, (2022) https://jurnal.mdp.ac.id/index.php/jatisi/article/view/1516	Berkontribusi Sebagai Acuan Dalam Proses Perhitungan Algoritma K-means
10	Deny Franata Pasaribu, Irfan Sudahri Damanik, Eka Irawan, Suhada, Heru Satria Tambunan	Memanfaatkan Algoritma K-Means Dalam Memetakan Potensi Hasil Produksi Kelapa Sawit PTPN IV Marihat	Jurnal Teknologi Informasi dan Rekayasa Komputer Vol. 2, No. 1, (2021) http://bios.sinergis.org/index.php/bios/article/view/17	Berkontribusi Sebagai Acuan Dalam Proses Perhitungan Algoritma K-means

Berdasarkan tinjauan pustaka yang dipaparkan pada Tabel 2.11, diperoleh landasan pengetahuan yang menjadi rujukan dalam pelaksanaan penelitian ini. Kontribusi dan referensi penelitian sebelumnya memberikan wawasan ilmiah terkait permasalahan pemetaan wilayah penyebaran SPKLU dengan penerapan algoritma K-Means. Penelitian ini dilakukan karena belum terdapat pengembangan dari studi sebelumnya yang mengaplikasikan K-Means dengan memanfaatkan variabel data transaksi dan data aktual pada SPKLU, sehingga penelitian ini bertujuan untuk

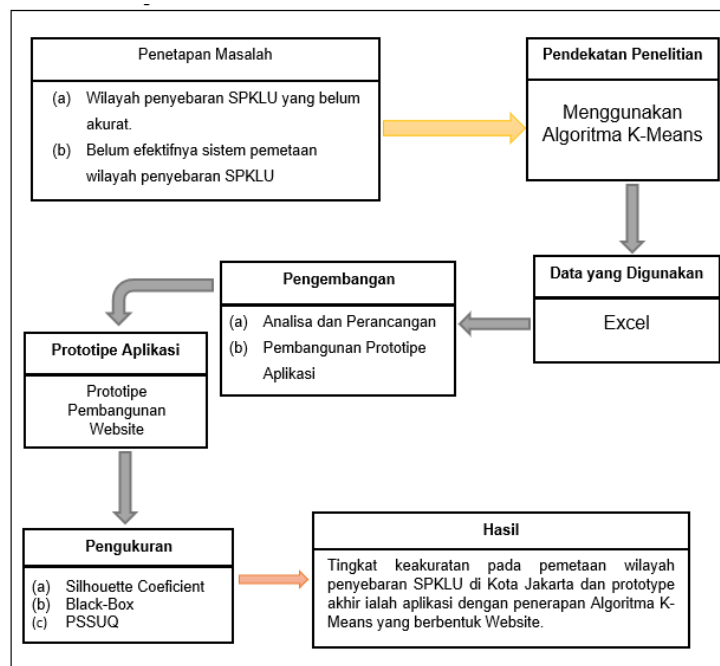
mengisi kekosongan tersebut dan memberikan informasi yang lebih komprehensif mengenai distribusi SPKLU.

C. Kerangka Berfikir

Kerangka berfikir yang dirumuskan dalam penelitian ini disusun untuk merepresentasikan konsep pemecahan masalah secara menyeluruh, yang mencakup permasalahan yang akan diteliti, metode penelitian yang digunakan, serta hasil yang diharapkan.

Penelitian ini diawali dengan penetapan permasalahan terkait objek penelitian melalui identifikasi masalah, yaitu wilayah penyebaran SPKLU yang belum akurat serta sistem pemetaan yang masih belum efektif. Untuk mengatasinya, digunakan pendekatan algoritma K-Means dengan memanfaatkan data dalam bentuk Excel. Selanjutnya, penelitian dikembangkan melalui analisis data, perancangan, dan pembuatan prototipe aplikasi berbasis situs web. Setelah prototipe dibangun, dilakukan pengujian menggunakan Silhouette Coefficient untuk mengukur keakuratan hasil penerapan algoritma K-Means, uji blackbox untuk menguji fungsi sistem, serta PSSUQ untuk menilai tingkat kebergunaan sistem.

Hasil dari penelitian ini diharapkan dapat memberikan nilai akurasi pemetaan wilayah penyebaran SPKLU di Jakarta sekaligus mengetahui tingkat kelayakan prototipe aplikasi yang dikembangkan. Kerangka pemikiran dapat dilihat pada Gambar 2.10.



Gambar 2. 10 Kerangka Berfikir

D. Hipotesis Penelitian

Metode k-means merupakan metode membagi data kedalam cluster berdasarkan kesamaan jenis atau karakteristik data. Metode k-means merupakan metode membagi data kedalam cluster berdasarkan kesamaan jenis atau karakteristik data. merujuk pada penelitian sebelumnya menggunakan algoritma K Means mengenai "Penerapan Algoritma K-Means Untuk Pemetaan Penerimaan Bantuan kesejahteraan Masyarakat di Kota Batam" (Azwanti and Putria, 2023), menunjukkan tingkat keberhasilan penggunaan algoritma K-Means dalam memetakan prioritas penerima bantuan di kota batam . Basis kerja pada algoritma K-Means yaitu melakukan suatu pengelompokan, bahwa pada penelitian ini juga berkaitan dengan pengelompokan data wilayah penyebaran SPKLU di Kota Jakarta. Berdasarkan pemahaman diatas, maka dapat ditetapkan hipotesis penelitian ini adalah bahwa penerapan algoritma K-Means di duga dapat melakukan pemetaan wilayah penyebaran SPKLU di kota jakarta.