

BAB II KERANGKA TEORITIS

A. Landasan Teori

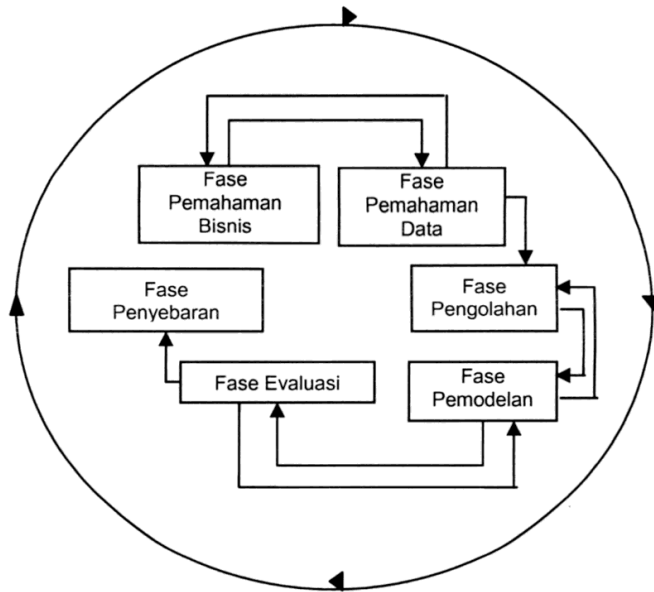
1. Konsep Data Mining

Data *mining* merupakan kegiatan mengekstrasi data dalam jumlah yang besar untuk mendapatkan informasi dan digunakan untuk membuat suatu pengambilan keputusan (Jollyta, Ramdhan, & Zarlis, 2020, p. 46). Data mining memiliki 5 (lima) peranan secara umum yaitu estimasi, prediksi, klasifikasi, *clustering*, dan asosiasi (Suntoro, 2019, p. 3). Menurut (Kusrini & Luthfi, 2009, p. 4) data mining merupakan proses otomatis terhadap data yang sudah ada dalam jumlah besar untuk mendapatkan pola atau hubungan yang mungkin memberikan indikasi yang bermanfaat.

Menurut (Buulolo, 2020, p. 7) menyatakan bahwa data mining dibagi menjadi beberapa kelompok berdasarkan fungsi dan tujuannya, yaitu;

- (a) deskripsi, deskripsi dapat digunakan untuk mengidentifikasi pola yang sering terjadi dan pola tersebut dapat diubah menjadi aturan yang dapat digunakan untuk memperlancar kegiatan;
- (b) klasifikasi, merupakan pengelompokan berdasarkan antara variabel target dan variabel kriteria;
- (c) estimasi, merupakan pengelompokan berdasarkan antara variabel target dan variabel kriteria dengan menggunakan variabel target berbentuk numerik;
- (d) prediksi, merupakan pengelompokan dengan nilai hasil prediksi akan digunakan untuk masa yang akan datang berdasarkan data-data sebelumnya ;
- (e) pengklasteran, merupakan pengelompokan record-record, kelas dan objek yang mempunyai kesamaan antara satu dengan yang lain dan mempunyai perbedaan dengan record-record pada kelompok lain;
- (f) asosiasi, pada data mining asosiasi yaitu menemukan atribut yang muncul pada waktu tertentu.

CRISP-DM merupakan salah satu metode yang menjelaskan tentang tahapan data mining (Ville, 2001, p. 37). CRISP-DM digunakan sebagai standar proses data mining untuk memecahkan suatu permasalahan pada penelitian atau bisnis; Dalam CRISP-DM, siklus hidup sebuah data mining memiliki enam tahapan yang dijelaskan pada gambar 2.1 (Kusrini & Luthfi, 2009, p. 8)



Gambar 2. 1 Tahapan CRISP-DM

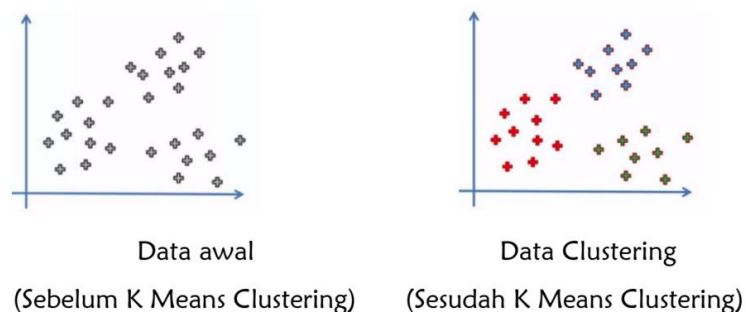
Terdapat enam tahapan CRISP-DM, yaitu (Kusrini & Luthfi, 2009, p. 9):

- (1) Fase Pemahaman Bisnis(Business Understanding Phase), merupakan tahapan dilakukan penentuan tujuan dari sebuah proyek dan kebutuhan secara detail dalam lingkup penelitian atau bisnis secara keseluruhan, serta menerjemahkan tujuan dan batasan menjadi formula dari permasalahan data mining dan menyiapkan strategi awal untuk mencapai tujuan;
- (2) Fase Pemahaman data(Data Understanding phase), merupakan tahapan dilakukan pengumpulan data yang kemudian dianalisis untuk mencari informasi tentang pola data dan kualitas data;
- (3) Fase Pengolahan data(Data Preparation Phase), merupakan tahapan menganalisis dan menyiapkan data dengan menyesuaikan kasus dan variabel yang digunakan untuk tahapan pemodelan;
- (4) Fase Pemodelan(Modeling Phase), pada tahapan ini dilakukan pemilihan dan pengaplikasian teknik pemodelan yang sesuai dengan kasus dan permasalahan;
- (5) Fase Evaluasi(Evaluation Phase), merupakan tahapan mengevaluasi model yang telah dirancang dengan menentukan apakah terdapat permasalahan penting dari penelitian atau bisnis yang tidak terpecahkan dan melihat apakah model yang dirancang telah mencapai tujuan bisnis atau belum;
- (6) Fase Penyebaran(Deployment Phase), merupakan tahapan menggunakan model data mining dengan pembuatan laporan atau penerapan proses data mining secara berulang pada perusahaan.

2. Pengertian Clustering dan Konsep Algoritma K-means

Clustering atau yang sering disebut pengelompokan merupakan salah satu model yang dicapai dengan membagi suatu populasi yang heterogen menjadi sejumlah cluster yang homogen (Jollyta, Ramdhan, & Zarlis, 2020, p. 53). *Clustering* dilakukan dengan mengumpulkan data yang sejenis yang berasal dari data set yang lebih besar. (Jollyta, Ramdhan, & Zarlis, 2020, p. 54). *Clustering* ini digunakan untuk mengidentifikasi kelompok dengan mengelompokkan data yang memiliki atribut yang mirip (Vulandari, 2017, p. 53).

Algoritma k-means merupakan metode membagi data yang ada kedalam bentuk satu atau lebih *cluster* dan merupakan salah satu teknik *clustering non-hierarki* dimana k-means membagi data kedalam *cluster*/kelompok berdasarkan dari kesamaan karakteristik data dan dikelompokkan kedalam cluster yang sama dan sementara itu data yang memiliki karakteristik yang berbeda dikelompokkan kedalam cluster yang lain (Jollyta, Ramdhan, & Zarlis, 2020, p. 119). Algoritma *k-means* beroperasi dengan membagi data ke dalam k buah klaster yang telah ditentukan seperti yang dijelaskan pada gambar 2.2, berikut ilustrasi dari clustering menggunakan k-means (Sudipa, et al., 2023, p. 20).



Gambar 2. 2 Ilustrasi Clustering Menggunakan K-Means

Contoh kasus penerapan algoritma k-means beserta tahapan algoritma k-means (Marisa, et al., 2023, p. 42) :

Dalam tabel 2.1 disajikan data dummy berupa 10 data SME player yang telah dilakukan normalisasi dan transformasi yang akan dikelompokkan berdasarkan persamaan data kriteria dari 4 kriteria yang dimiliki player yaitu :

- C1 SME Scope : 1=Small, 2=Micro, 3=Macro
- C2 Market : 1=Regional, 2=National, 3=International

C3 Product : 1=Fashion, 2=Kriya, 3=Kuliner, 4=Digital Creative, 5=Furniture, 6=Home Stay, 7=Foto and Video, 8=Buku, 9=Other

C4 Market Place : 1=Digital, 2=Direct market, 3= Mix Market.

| SME Player | Criteria | | | |
|---------------|-------------------|----------------|-----------------|----------------------|
| | C1 (SME Scope) | C2 (Market) | C3 (Product) | C4 (Market Place) |
| SME Player-1 | 2 | 1 | 1 | 1 |
| SME Player-2 | 2 | 1 | 3 | 3 |
| SME Player-3 | 1 | 2 | 3 | 3 |
| SME Player-4 | 1 | 1 | 2 | 3 |
| SME Player-5 | 3 | 3 | 3 | 3 |
| SME Player-6 | 1 | 2 | 4 | 2 |
| SME Player-7 | 2 | 2 | 4 | 2 |
| SME Player-8 | 1 | 1 | 5 | 1 |
| SME Player-9 | 1 | 1 | 2 | 1 |
| SME Player-10 | 2 | 2 | 2 | 1 |

Langkah 1, menentukan jumlah kluster iterasi pertama secara acak dan dinotasikan K. Dalam desain model ini kluster ditentukan sejumlah 3(K=3) dengan memilih secara acak dari data di atas dengan detail data pada data di bawah ini. Maka dapat dinotasikan $C_1(2,1,1,1)$, $C_2=(3,3,3,3)$, $C_3=(2,2,4,2)$.

| SME Player (S) | Cr1 | Cr2 | Cr3 | Cr4 | Centroids |
|----------------|-----|-----|-----|-----|-----------|
| SME Player-1 | 2 | 1 | 1 | 1 | C1 |
| SME Player-2 | 2 | 1 | 3 | 3 | |
| SME Player-3 | 1 | 2 | 3 | 3 | |
| SME Player-4 | 1 | 1 | 2 | 3 | |
| SME Player-5 | 3 | 3 | 3 | 3 | C2 |
| SME Player-6 | 1 | 2 | 4 | 2 | |
| SME Player-7 | 2 | 2 | 4 | 2 | |
| SME Player-8 | 1 | 1 | 5 | 1 | C3 |
| SME Player-9 | 1 | 1 | 2 | 1 | |
| SME Player-10 | 2 | 2 | 2 | 1 | |

Langkah 2, menghitung nilai jarak data ke centroid, dengan menggunakan rumus Euclidean distance dengan rumus:

$$D(a, b) = \sqrt{\sum_{k=1}^n (a_k - b_k)^2}$$

Berdasarkan data di atas, jarak data dari centroid masing-masing kriteria adalah sebagai berikut:

Untuk mendapatkan (S_n, C_1) :

$$D(S_1, C_1) = \sqrt{(S_{1a} - C_{1a})^2 + (S_{1b} - C_{1b})^2 + (S_{1c} - C_{1c})^2 + (S_{1d} - C_{1d})^2}$$

$$D(S_1, C_1) = \sqrt{(2 - 2)^2 + (1 - 1)^2 + (1 - 1)^2 + (1 - 1)^2} = 0$$

$$D(S_2, C_1) = \sqrt{(S_{2a} - C_{1a})^2 + (S_{2b} - C_{1b})^2 + (S_{2c} - C_{1c})^2 + (S_{2d} - C_{1d})^2}$$

$$D(S_2, C_1) = \sqrt{(2-2)^2 + (1-1)^2 + (3-1)^2 + (1-3)^2} = 0,828$$

Selanjutnya data $D(S_n, C_1)$ menggunakan cara yang sama dengan hasil perhitungan pada data di atas kolom Cr1.

Untuk mendapatkan (S_n, C_2) :

$$D(S_1, C_2) = \sqrt{(S_{1a} - C_{2a})^2 + (S_{1b} - C_{2b})^2 + (S_{1c} - C_{2c})^2 + (S_{1d} - C_{2d})^2}$$

$$D(S_1, C_2) = \sqrt{(2-3)^2 + (1-3)^2 + (1-3)^2 + (1-3)^2} = 3,606$$

$$D(S_2, C_2) = \sqrt{(S_{2a} - C_{2a})^2 + (S_{2b} - C_{2b})^2 + (S_{2c} - C_{2c})^2 + (S_{2d} - C_{2d})^2}$$

$$D(S_2, C_2) = \sqrt{(2-3)^2 + (1-3)^2 + (3-3)^2 + (3-3)^2} = 2,236$$

Selanjutnya data $D(S_n, C_2)$ menggunakan cara yang sama dengan hasil perhitungan pada data di atas kolom Cr2.

Untuk mendapatkan (S_n, C_3) :

$$D(S_1, C_3) = \sqrt{(S_{1a} - C_{3a})^2 + (S_{1b} - C_{3b})^2 + (S_{1c} - C_{3c})^2 + (S_{1d} - C_{3d})^2}$$

$$D(S_1, C_3) = \sqrt{(2-2)^2 + (1-2)^2 + (1-4)^2 + (1-2)^2} = 3,317$$

$$D(S_2, C_3) = \sqrt{(S_{2a} - C_{3a})^2 + (S_{2b} - C_{3b})^2 + (S_{2c} - C_{3c})^2 + (S_{2d} - C_{3d})^2}$$

$$D(S_2, C_3) = \sqrt{(2-2)^2 + (1-2)^2 + (3-4)^2 + (3-2)^2} = 1,732$$

Selanjutnya data $D(S_n, C_3)$ menggunakan cara yang sama dengan hasil perhitungan pada data di bawah ini kolom Cr3.

| SME Player | Cr1 | Cr2 | Cr3 | Cr4 | $D(S_n, C_1)$ | $D(S_n, C_2)$ | $D(S_n, C_3)$ |
|---------------|-----|-----|-----|-----|---------------|---------------|---------------|
| SME Player-1 | 2 | 1 | 1 | 1 | 0,000 | 3,606 | 3,317 |
| SME Player-2 | 2 | 1 | 3 | 3 | 2,828 | 2,236 | 1,732 |
| SME Player-3 | 1 | 2 | 3 | 3 | 3,162 | 2,236 | 1,732 |
| SME Player-4 | 1 | 1 | 2 | 3 | 2,449 | 3,000 | 2,646 |
| SME Player-5 | 3 | 3 | 3 | 3 | 3,606 | 0,000 | 2,000 |
| SME Player-6 | 1 | 2 | 4 | 2 | 3,464 | 2,646 | 1,000 |
| SME Player-7 | 2 | 2 | 4 | 2 | 3,317 | 2,000 | 0,000 |
| SME Player-8 | 1 | 1 | 5 | 1 | 4,123 | 4,000 | 2,000 |
| SME Player-9 | 1 | 1 | 2 | 1 | 1,414 | 3,606 | 2,646 |
| SME Player-10 | 2 | 2 | 2 | 1 | 1,414 | 2,646 | 2,236 |

Langkah 3, mengelompokkan data sesuai centroid, dengan cara mengelompokkan data sesuai jarak terpendek masing-masing item. Proses ini dapat dihitung dengan mencari nilai terkecil diantara nilai $D(S_n, C_1)$, $D(S_n, C_2)$, $D(S_n, C_3)$, kemudian cluster ditentukan berdasarkan nilai terkecil yang diperoleh salah satu nilai Euclidean distance pada masing-masing itemset. Hasil penentuan cluster dapat terlihat pada data di bawah ini.

| SME Player | $D(S_n, C_1)$ | $D(S_n, C_2)$ | $D(S_n, C_3)$ | Nearest Distance | Cluster |
|--------------|---------------|---------------|---------------|------------------|---------|
| SME Player-1 | 0,000 | 3,606 | 3,317 | 0,000 | C1 |

| SME Player | $D(S_n, C_1)$ | $D(S_n, C_2)$ | $D(S_n, C_3)$ | Nearest Distance | Cluster |
|---------------|---------------|---------------|---------------|------------------|---------|
| SME Player-2 | 2,828 | 2,236 | 1,732 | 1,732 | C3 |
| SME Player-3 | 3,162 | 2,236 | 1,732 | 1,732 | C3 |
| SME Player-4 | 2,449 | 3,000 | 2,646 | 2,449 | C1 |
| SME Player-5 | 3,606 | 0,000 | 2,000 | 0,000 | C2 |
| SME Player-6 | 3,464 | 2,646 | 1,000 | 1,000 | C3 |
| SME Player-7 | 3,317 | 2,000 | 0,000 | 0,000 | C3 |
| SME Player-8 | 4,123 | 4,000 | 2,000 | 2,000 | C3 |
| SME Player-9 | 1,414 | 3,606 | 2,646 | 1,414 | C1 |
| SME Player-10 | 1,414 | 2,646 | 2,236 | 1,414 | C1 |

Langkah 4, menentukan centroid untuk iterasi 2, dengan menghitung rata-rata dari hasil penjumlahan data masing-masing kelompok cluster yang terlihat pada data di bawah ini.

| SME | C1 | C2 | C3 | C4 |
|---------------|-----|------|------|-----|
| Cluster1 | | | | |
| SME Player-1 | 2 | 1 | 1 | 1 |
| SME Player-4 | 1 | 1 | 2 | 3 |
| SME Player-9 | 1 | 1 | 2 | 1 |
| SME Player-10 | 2 | 3 | 2 | 1 |
| Centroid-1 | 1,5 | 1,25 | 1,75 | 1,5 |
| Cluster2 | | | | |
| SME Player-5 | 3 | 3 | 3 | 3 |
| Centroid-2 | 3 | 3 | 3 | 3 |
| Cluster3 | | | | |
| SME Player-2 | 2 | 1 | 3 | 3 |
| SME Player-3 | 1 | 2 | 3 | 3 |
| SME Player-6 | 1 | 2 | 4 | 2 |
| SME Player-7 | 2 | 2 | 4 | 2 |
| SME Player-8 | 1 | 1 | 5 | 1 |
| Centroid-3 | 1,4 | 1,6 | 3,8 | 2,2 |

Dengan hasil perhitungan nilai rata-rata data tiap kelompok cluster pada data diatas didapat nilai centroid dengan detail data centroid yang dapat dinotasikan $C_1(1.5,1.25,1.75,1.5)$, $C_2(3,3,3,3)$, $C_3(1.4,1.6,3.8,2.2)$.

Langkah 5, mengulang proses seperti iterasi sebelumnya dengan data centroid yang berbeda yaitu menghitung jarak data ke centroid, dengan menggunakan rumus Euclidean distance (persamaan 1).

Berdasarkan pada data di atas maka jarak data dari centroid masing-masing kriteria adalah sebagai berikut:

Untuk mendapatkan (S_n, C_1) :

$$D(S_1, C_1) = \sqrt{(S_{1a} - C_{1a})^2 + (S_{1b} - C_{1b})^2 + (S_{1c} - C_{1c})^2 + (S_{1d} - C_{1d})^2}$$

$$D(S_1, C_1) = \sqrt{(2 - 1,5)^2 + (1 - 1,25)^2 + (1 - 1,75)^2 + (1 - 1,5)^2} = 1,061$$

$$D(S_2, C_1) = \sqrt{(S_{2a} - C_{1a})^2 + (S_{2b} - C_{1b})^2 + (S_{2c} - C_{1c})^2 + (S_{2d} - C_{1d})^2}$$

$$D(S_2, C_1) = \sqrt{(2 - 1,5)^2 + (1 - 1,27)^2 + (3 - 1,75)^2 + (1 - 1,5)^2} = 2,031$$

Selanjutnya data $D(S_n, C_1)$ menggunakan cara yang sama dengan hasil perhitungan pada data di atas kolom Cr1.

Untuk mendapatkan (S_n, C_2) :

$$D(S_1, C_2) = \sqrt{(S_{1a} - C_{2a})^2 + (S_{1b} - C_{2b})^2 + (S_{1c} - C_{2c})^2 + (S_{1d} - C_{2d})^2}$$

$$D(S_1, C_2) = \sqrt{(2 - 3)^2 + (1 - 3)^2 + (1 - 3)^2 + (1 - 3)^2} = 3,606$$

$$D(S_2, C_2) = \sqrt{(S_{2a} - C_{2a})^2 + (S_{2b} - C_{2b})^2 + (S_{2c} - C_{2c})^2 + (S_{2d} - C_{2d})^2}$$

$$D(S_2, C_2) = \sqrt{(2 - 3)^2 + (1 - 3)^2 + (3 - 3)^2 + (3 - 3)^2} = 2,236$$

Selanjutnya data $D(S_n, C_2)$ menggunakan cara yang sama dengan hasil perhitungan pada data di atas kolom Cr2.

Untuk mendapatkan (S_n, C_3) :

$$D(S_1, C_3) = \sqrt{(S_{1a} - C_{3a})^2 + (S_{1b} - C_{3b})^2 + (S_{1c} - C_{3c})^2 + (S_{1d} - C_{3d})^2}$$

$$D(S_1, C_3) = \sqrt{(2 - 1,4)^2 + (1 - 1,6)^2 + (1 - 3,8)^2 + (1 - 2,2)^2} = 3,162$$

$$D(S_2, C_3) = \sqrt{(S_{2a} - C_{3a})^2 + (S_{2b} - C_{3b})^2 + (S_{2c} - C_{3c})^2 + (S_{2d} - C_{3d})^2}$$

$$D(S_2, C_3) = \sqrt{(2 - 1,4)^2 + (1 - 1,6)^2 + (3 - 3,8)^2 + (3 - 2,2)^2} = 1,414$$

Selanjutnya data $D(S_n, C_3)$ menggunakan cara yang sama dengan hasil perhitungan pada data di bawah kolom Cr3.

| SME Player | Cr1 | Cr2 | Cr3 | Cr4 | $D(S_n, C_1)$ | $D(S_n, C_2)$ | $D(S_n, C_3)$ |
|---------------|-----|-----|-----|-----|---------------|---------------|---------------|
| SME Player-1 | 2 | 1 | 1 | 1 | 1,061 | 3,606 | 3,162 |
| SME Player-2 | 2 | 1 | 3 | 3 | 2,031 | 2,236 | 1,414 |
| SME Player-3 | 1 | 2 | 3 | 3 | 2,151 | 2,236 | 1,265 |
| SME Player-4 | 1 | 1 | 2 | 3 | 1,620 | 3,000 | 2,098 |
| SME Player-5 | 3 | 3 | 3 | 3 | 3,021 | 0,000 | 2,408 |
| SME Player-6 | 1 | 2 | 4 | 2 | 2,475 | 2,646 | 0,632 |
| SME Player-7 | 2 | 2 | 4 | 2 | 2,475 | 2,000 | 0,775 |
| SME Player-8 | 1 | 1 | 5 | 1 | 3,335 | 4,000 | 1,844 |
| SME Player-9 | 1 | 1 | 2 | 1 | 0,791 | 3,606 | 2,280 |
| SME Player-10 | 2 | 2 | 2 | 1 | 1,061 | 2,646 | 2,280 |

Langkah 6, mengelompokkan data sesuai centroid, dengan cara mengelompokkan data sesuai jarak terpendek masing-masing item. Proses ini dapat dihitung dengan mencari nilai terkecil diantara nilai (S_n, C_1) , $D(S_n, C_2)$, $D(S_n, C_3)$, kemudian cluster ditentukan berdasarkan nilai terkeci yang diperoleh salah satu nilai Euclidean distance pada masing-masing itemset. Hasil penentuan cluster dapat terlihat pada data di bawah ini.

| SME Player | $D(S_n, C_1)$ | $D(S_n, C_2)$ | $D(S_n, C_3)$ | Nearest Distance | Cluster |
|---------------|---------------|---------------|---------------|------------------|---------|
| SME Player-1 | 1,061 | 3,606 | 3,162 | 1,061 | C1 |
| SME Player-2 | 2,031 | 2,236 | 1,414 | 1,414 | C3 |
| SME Player-3 | 2,151 | 2,236 | 1,265 | 1,265 | C3 |
| SME Player-4 | 1,620 | 3,000 | 2,098 | 1,620 | C1 |
| SME Player-5 | 3,021 | 0,000 | 2,408 | 0,000 | C2 |
| SME Player-6 | 2,475 | 2,646 | 0,632 | 0,632 | C3 |
| SME Player-7 | 2,475 | 2,000 | 0,775 | 0,775 | C3 |
| SME Player-8 | 3,335 | 4,000 | 1,844 | 1,844 | C3 |
| SME Player-9 | 0,791 | 3,606 | 2,280 | 0,791 | C1 |
| SME Player | $D(S_n, C_1)$ | $D(S_n, C_2)$ | $D(S_n, C_3)$ | Nearest Distance | Cluster |
| SME Player-10 | 1,061 | 2,646 | 2,280 | 1,414 | C1 |

Dari hasil iterasi kedua tidak ada perubahan posisi kluster, sehingga proses iterasi berhenti sampai pada iterasi kedua, dan cluster yang dihasilkan adalah yang disajikan pada data di bawah ini.

| SME | C1 | C2 | C3 | C4 |
|---------------|----|----|----|----|
| Cluster1 | | | | |
| SME Player-1 | 2 | 1 | 1 | 1 |
| SME Player-4 | 1 | 1 | 2 | 3 |
| SME Player-9 | 1 | 1 | 2 | 1 |
| SME Player-10 | 2 | 3 | 2 | 1 |
| Cluster2 | | | | |
| SME Player-5 | 3 | 3 | 3 | 3 |
| Cluster3 | | | | |
| SME Player-2 | 2 | 1 | 3 | 3 |
| SME Player-3 | 1 | 2 | 3 | 3 |
| SME Player-6 | 1 | 2 | 4 | 2 |
| SME Player-7 | 2 | 2 | 4 | 2 |
| SME Player-8 | 1 | 1 | 5 | 1 |

Data di bawah ini meletakkan dataset ke dalam hasil kelompok kluster yang terbentuk. Menghasilkan 3 kluster dengan rincian C1 memiliki 4 anggota, C2 memiliki 1 anggota, dan C3 memiliki 5 anggota.

| SME | C1 | C2 | C3 | C4 |
|---------------|-------|---------------|---------|------------|
| Cluster1 | | | | |
| SME Player-1 | Micro | Regional | Fashion | Digital |
| SME Player-4 | Small | Regional | Kriya | Mix Market |
| SME Player-9 | Small | Regional | Kriya | Digital |
| SME Player-10 | Micro | National | Kriya | Digital |
| Cluster2 | | | | |
| SME Player-5 | Macro | International | Kuliner | Mix Market |

| SME | C1 | C2 | C3 | C4 |
|--------------|-------|----------|------------------|---------------|
| Cluster3 | | | | |
| SME Player-2 | Micro | Regional | Kuliner | Mix Market |
| SME Player-3 | Small | National | Kuliner | Mix Market |
| SME Player-6 | Small | National | Digital Creative | Direct Market |
| SME Player-7 | Micro | National | Digital Creative | Direct Market |
| SME Player-8 | Small | Regional | Furniture | Digital |

3. Algoritma Silhouette Coefficient

Silhouette coefficient merupakan metode evaluasi yang dihitung untuk menguji seberapa optimal atau seberapa tepat sebuah cluster yang telah terbentuk dari proses *Clustering* pada persamaan berikut (Irwansyah & Faisal, 2015, p. 17):

- (1) Hitung rata-rata jarak dari suatu data, menggunakan Persamaan 1 maka didapatkan rata-rata dengan cara memisalkan i terhadap semua data lain yang berada dalam satu cluster sebagai berikut;

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j)$$

Dimana,

$a(i)$ = perbedaan rata-rata objek (i) ke semua objek lain pada A

$d(i, j)$ = jarak antara data i dengan j

A = *cluster*

- (2) Hitung rata-rata jarak data i tersebut dengan semua data di cluster lain, dan diambil nilai terkecilnya menggunakan Persamaan 2;

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j)$$

Dimana,

$d(i, C)$ = perbedaan rata-rata objek (i) ke semua objek lain pada C

C = *cluster* lain selain *cluster* A atau *cluster* C tidak sama dengan *cluster* A

- (3) Setelah menghitung $d(i, c)$ untuk semua c , maka diambil nilai terkecil dengan menggunakan persamaan 3;

$$b(i) = \min_{C \neq A} d(i, C)$$

Cluster B yang mencapai minimum (yaitu, $d(i, B) = b(i)$) disebut tetangga dari objek (i). ini adalah cluster terbaik kedua untuk objek (i)

- (4) Nilai silhouette coefficient didefinisikan seperti pada persamaan 4.

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

| Nilai Silhouette Coefficient | Keterangan |
|-------------------------------------|-------------------|
| Dengan rentang $0.7 < SC \leq 1$ | Struktur Kuat |
| Dengan rentang $0.5 < SC \leq 0.7$ | Struktur Sedang |
| Dengan rentang $0.25 < SC \leq 0.5$ | Struktur Lemah |
| Dengan rentang $SC \leq 0.25$ | Tidak Terstruktur |

4. Pengertian Bahasa Pemrograman

Bahasa pemrograman atau bahasa komputer merupakan instruksi untuk memerintahkan komputer yang berisi sintaks dan makna bahasa yang dipakai untuk mendefinisikan program komputer (Rosad, 2022, p. 1). Bahasa pemrograman bagi pemrogram(pengembang) digunakan sebagai cara untuk berkomunikasi dengan komputer (Wali et al., 2023, p. 1). Pemrograman merupakan kegiatan untuk membuat program yang meliputi (Ayub, 2020, p. 2) ;

- (a) Mempelajari persoalan dan melakukan analisis persoalan;
- (b) Menuliskan spesifikasi kebutuhan program;
- (c) Membuat desain algoritma berdasarkan spesifikasi program;
- (d) Menulis program berdasarkan desain algoritma;
- (e) Melakukan debugging dan uji coba program.

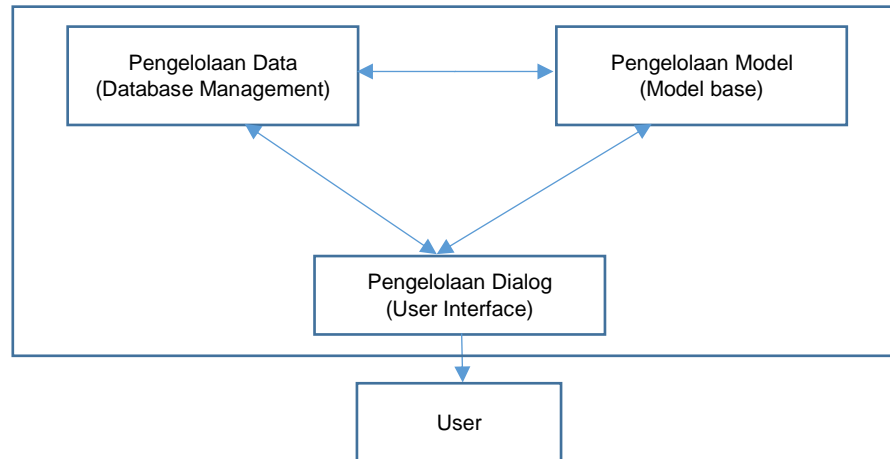
5. Pengertian Data

Data merupakan sekumpulan fakta yang digunakan sebagai bahan dalam menganalisis atau menjadi referensi yang bisa digunakan untuk mendukung suatu penelitian atau pendapat (Arhami & Nasir, 2020, p. 15). Data merupakan fakta yang dapat digunakan untuk menggambarkan suatu keadaan dan peristiwa yang berbentuk kategori(baik, tinggi, buruk, , rendah, dan sebagainya) (Putra & Hanggara, 2022, p. 3). Penyajian data dapat berbentuk histogram, grafik, tabel, polygon dan ogiv (Putra & Hanggara, 2022, p. 4).

6. Pengertian Sistem Pendukung Keputusan / Decision Support System(DSS)

Sistem pendukung keputusan merupakan sistem yang interaktif yang menyediakan pemodelan data, informasi, serta manipulasi data dan digunakan untuk membantu dalam pengambilan keputusan dalam kondisi semi terstruktur dan terstruktur (Kusrini, 2007, p. 15). DSS biasanya digunakan untuk mendukung penyelesaian dalam suatu masalah (Kusrini, 2007, p. 16). Decision Support System merupakan sistem yang mampu memberikan kemampuan pengkomunikasian untuk memberikan usulan dalam pengambilan sebuah keputusan yang biasa digunakan oleh manajer (Sudipa, et al., 2023, p. 2).

Sistem Pendukung keputusan secara umum dibangun oleh tiga komponen besar yaitu database management, Model base dan Software System/User Interface; Gambar 2.3 menggambarkan tentang komponen SPK;



Gambar 2. 3 Komponen SPK

Tiga komponen dari SPK yaitu :

- (a) Database Management, merupakan data yang terorganisasi dalam suatu basis data dan diperlukan data yang relevan dengan permasalahan yang akan dipecahkan untuk keperluan SPK;
- (b) Model Base, merupakan suatu model yang merepresentasikan permasalahan kedalam format kuantitatif. Dengan mengembangkan dan membandingkan solusi alternative, model base ini memungkinkan pengambil keputusan menganalisa secara utuh;
- (c) User Interface/Pengelolaan Dialog, gabungan antara komponen database management dan model base merupakan komponen user interface. User interface menampilkan tampilan sistem yang dimengerti oleh pengguna.

7. Pengertian SDLC(System Development Life Cycle)

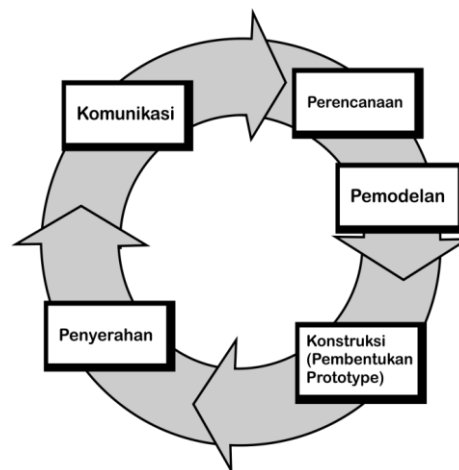
SDLC merupakan metodologi yang digunakan untuk mengembangkan sistem mulai dari pengembangan, pemeliharaan, penggunaan sistem informasi (Kadir, 2014, p. 344).SDLC memiliki tahapan sebagai berikut (Zein, et al., 2023, p. 18);

- (1) Perancangan(planning), tahapan yang dimulai dengan identifikasi masalah hingga identifikasi kebutuhan;
- (2) Analisis(analysis), tahapan mengidentifikasi kebutuhan dan persyaratan dari pengguna;

- (3) Desain(design), tahapan ini dilakukan perancangan arsitektur perangkat lunak, desain database, desain antarmuka, dan rancangan algoritma yang dilakukan;
- (4) Implementasi(implementation), tahapan ini dilakukan proses pengkodean dan integrasi modul-modul perangkat lunak serta memastikan perangkat lunak bisa berfungsi sesuai dengan kebutuhan pengguna;
- (5) Pengujian(testing), tahapan ini dilakukan pengujian fungsional, kinerja, dan keamanan dari perangkat lunak untuk memastikan perangkat lunak bekerja sesuai kebutuhan pengguna;
- (6) Pemeliharaan(maintenance), tahapan ini pengguna harus memastikan perangkat lunak tetap dapat berfungsi dengan baik dan melakukan perubahan sesuai dengan kebutuhan pengguna.

8. Pengertian Prototyping

Prototyping merupakan proses pembuatan ide oleh pengembang bagai mana membuat sistem yang dapat berfungsi setelah pembuatan sistem selesai (McLeod & P.Schell, 2008, p. 201). Model prototype dapat digunakan untuk menghubungkan ketidakpahaman pelanggan mengenai hal teknis dan menjelaskan kebutuhan pengguna secara detail kepada pengembang; Model prototipe dimulai dari mengumpulkan kebutuhan pelanggan, hingga pembuatan program prototype agar pengguna bisa membayangkan seperti apa yang diinginkan oleh pengguna (A.S & Shalahuddin, 2013, p. 31).



Gambar 2. 4 Langkah-Langkah Prototyping

Sumber : (Pressman & Maxim, 2019)

Langkah-langkah pada model prototyping adalah :

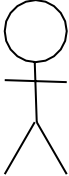
- (1) komunikasi, dilakukan dengan komunikasi antara pengembang dan pengguna tentang tujuan pembuatan sistem dan mengidentifikasi kebutuhan;
- (2) perencanaan, melakukan perencanaan cepat setelah dilakukannya komunikasi;
- (3) pemodelan, yaitu membuat model dan desain yang fokus pada gambaran dari segi software apakah dapat dilihat dan digunakan oleh pengguna;
- (4) konstruksi(pembentukan prototype), yaitu memulai pembuatan prototype;
- (5) penyerahan, yaitu penyerahan prototype kepada pengguna untuk dievaluasi oleh pengguna dan pemberian penilaian tentang kekurangan atau kebutuhan terhadap sistem.

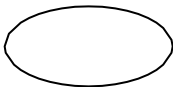

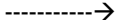
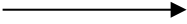

9. Pengertian UML (*Unified Modeling Language*)

Unified Modeling Language merupakan bahasa pemodelan yang digunakan untuk menggambarkan sistem perangkat lunak berdasarkan objek pada sistem. UML hanya memberikan notasi standar yang biasa digunakan untuk pemodelan objek, UML tidak menentukan metode yang digunakan saat menggambarkan sistem (A.S & Shalahuddin, 2014, p.130). UML terdiri dari beberapa diagram, yaitu :

a. Use Case Diagram

Menurut (A.S and Shalahuddin, 2014, p. 155) Use Case Diagram merupakan pemodelan yang digunakan untuk menggambarkan perilaku dari sistem informasi yang akan dibuat. Use Case Diagram ini mendeskripsikan bagaimana sebuah interaksi antara satu atau lebih aktor dengan sistem informasi yang akan dibuat. Use Case Diagram merupakan salah satu bagian dari jenis UML(*Unified Modeling Language*) yang menggambarkan hubungan dari interaksi antara actor dengan sistem. Use Case Diagram ini dapat mendeskripsikan jenis interaksi antara pengguna sistem dengan sistem.


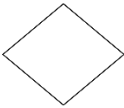



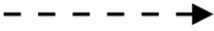

| SIMBOL | NAMA | KETERANGAN |
|---|-------|---|
|  | ACTOR | Proses, orang, atau sistem lain yang berinteraksi dengan sistem informasi yang dibuat berada di luar sistem informasi itu sendiri. Walaupun notasi seorang aktor adalah gambaran seseorang, biasanya simbol tersebut diwakili |

| SIMBOL | NAMA | KETERANGAN |
|---|--------------------------------------|--|
|  | <i>USECASE</i> | oleh kata benda di awal kalimat dengan nama aktor tersebut. Fungsionalitas yang disediakan oleh sistem sebagai entitas atau entitas yang melakukan pertukaran data dalam jumlah besar antar aktor biasanya dinyatakan dengan kata kerja di awal frasa dalam nama use case. |
|  | <i>ASOSIASI/ ASSOCIATION</i> | Komunikasi antar aktor dan use case yang berpartisipasi dalam use case tersebut berinteraksi dengan aktor. Hubungan antara use case tambahan dan use case yang memungkinkan use case yang ditambahkan menjadi independen, meskipun use case yang ditambahkan tidak memiliki nama yang sama dengan use case yang ditambahkan . |
| <i><<extend>></i>  | <i>EKSTENSI/ EXTEND</i> | Hubungan generalisasi-spesialisasi antara dua use case dimana satu fitur lebih umum dibandingkan yang lainnya (umum-spesifik). |
|  | <i>GENERALISASI / GENERALIZATION</i> | Hubungan usecase tambahan ke sebuah usecase untuk menjalankan fungsional |
| <i><<include>></i>  | <i>MENGGUNAKAN INCLUDE</i> | |

b. Class Diagram



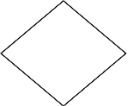
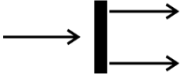


Diagram class menunjukkan alur database program. Setiap laporan sistem harus memiliki class diagram. Menurut para ahli, class diagram memiliki arti yang berbeda: itu mendefinisikan kelas yang akan dibangun pada sistem dan menggambarkan struktur sistem (Sukamto & Shalahudin, 2018, p. 141). Selain itu, class diagram juga berfungsi sebagai alat untuk menggambarkan dan memvisualisasikan struktur kelas yang terdapat dalam suatu sistem

(Satzinger, 2011, p. 28). Class diagram juga menunjukkan bagaimana kelompok objek bersama property, operasi, dan relasi yang sama.

| SIMBOL | NAMA | KETERANGAN |
|---|------------------|--|
|  | Generalization | Hubungan di mana objek anak (descendent) berbagi perilaku dan struktur data dengan objek induk (ancestor) di atasnya |
|  | Navy Association | Upaya untuk menghindari hubungan dengan lebih dari dua hal. |
|  | Class | Grup dari objek yang memiliki karakteristik dan operasi yang sama. |
|  | Collaboration | Deskripsi dari urutan tindakan yang ditampilkan sistem yang menghasilkan hasil yang dapat diukur untuk seorang aktor |
|  | Realization | Suatu objek melakukan operasi yang sebenarnya. |
|  | Dependency | Hubungan di mana elemen yang bergantung pada elemen yang tidak mandiri dipengaruhi oleh perubahan pada elemen mandiri. |
|  | Association | Yang menghubungkan objek satu sama lain. |



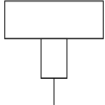
c. Activity Diagram

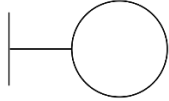
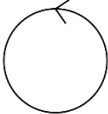
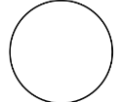


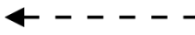
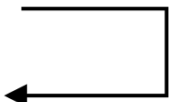
Menurut (Sugiarti, 2018, p. 135) Activity diagram menunjukkan aktivitas dalam proses atau sistem bisnis. Diagram aktivitas ini hanya menunjukkan aktivitas yang dapat dilakukan oleh sistem, bukan aktivitas yang dapat dilakukan oleh individu. Pada activity diagram, alur aktivitas adalah urutan proses bisnis atau menu yang ada pada sistem yang akan dibuat. Untuk membuat pemodelan yang baik dari alur kerja sistem, aktivitas diagram biasanya digunakan bersama dengan teknik pemodelan lainnya, seperti Diagram Kasus dan Diagram Status. Activity diagram juga berguna untuk menganalisis diagram use case dengan memberikan deskripsi aktor.

| SIMBOL | NAMA | KETERANGAN |
|---|---------------------|--|
|  | Status Awal/Initial | Status awal aktivitas sistem |
|  | Activity | Kebanyakan aktivitas yang dilakukan sistem dimulai dengan kata kerja. |
|  | Decision | Asosiasi yang menggabungkan lebih dari satu bisnis menjadi satu |
|  | Join | Asosiasi penggabungan adalah ketika lebih dari satu aktivitas dilakukan bersamaan. |
|  | Status Akhir/Final | Status akhir sistem. |
|  | Swimline | Memisahkan organisasi bisnis yang bertanggung jawab atas operasi tersebut. |

d. Sequence Diagram

Untuk menggambarkan sequence diagram, Anda harus tahu apa yang ada dalam sebuah use case dan metode apa yang dimiliki classnya (Sugiarti, 2018, p. 120). Sequence diagram menunjukkan interaksi bagaimana operasi sistem dilakukan. Mereka juga menunjukkan interaksi antar kelas untuk bertukar pesan atau data;

| SIMBOL | NAMA | KETERANGAN |
|---|----------|--|
|  | Actor | Menggambarkan entitas yang tidak termasuk dalam sistem dan berinteraksi dengan mereka. |
|  | Lifeline | Membuat objek terhubung selama urutan (pesan dikirim atau diterima). |
|  | General | Merepresentasikan entitas tunggal dalam sequence |

| SIMBOL | NAMA | KETERANGAN |
|---|-----------------|---|
|  | Boundary | Berfungsi sebagai bagian sistem, seperti antarmuka pengguna dan alat yang berinteraksi satu sama lain. |
|  | Control | Elemen mengatur cara informasi mengalir dalam sebuah scenario. |
|  | Entity | Elemen yang bertanggung jawab menyimpan informasi |
|  | Activation | Suatu titik di mana sebuah objek mulai terlibat dalam sebuah rangkaian yang menunjukkan atau mengirim atau menerima sebuah objek. |
|  | Message Entry | Berfungsi untuk menunjukkan pesan atau hubungan antar objek dan menunjukkan urutan kejadian yang terjadi. |
|  | Message to Self | Simbol ini menggambarkan Pesan atau hubungan objek itu sendiri, yang menunjukkan urutan peristiwa yang terjadi, diwakili oleh simbol ini. |
|  | Message Return | Menggambarkan hasil pengiriman pesan dari kanan ke kiri. |

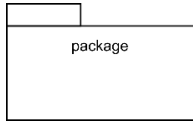
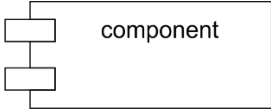



e. Component Diagram

Diagram yang menggambarkan kumpulan dan hubungan antar komponen merupakan diagram yang disebut dengan component diagram (A.S & Shalahuddin, 2014, p.148). Diagram component merupakan diagram yang digunakan untuk memperlihatkan keterkaitan dan ketergantungan antar komponen dalam sebuah sistem. Pengertian component diagram menurut UML yaitu :

- (a.) Komponen data yang digunakan untuk memanipulasi data;

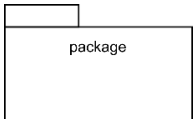
- (b.) Komponen business processing yang digunakan untuk menangani proses bisnis;
- (c.) Komponen security yang digunakan dalam menangani keamanan;
- (d.) Komponen user interface yang digunakan dalam menangani tampilan pada sebuah sistem;

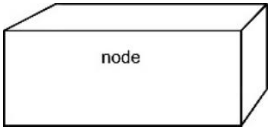


Berikut merupakan simbol-simbol yang dimiliki oleh component diagram :

| SIMBOL | NAMA | KETERANGAN |
|---|------------|--|
|  | Package | Package merupakan sebuah bungkusan dari satu atau lebih komponen |
|  | Component | Komponen sistem |
|  | Dependency | Arah panah mengarah pada komponen yang digunakan karena bergantung antar komponen. |
|  | Interface | Sebagai antarmuka komponen agar tidak dapat mengakses komponen secara langsung |
|  | Link | Relasi antar komponen |

f. Deployment Diagram

Deployment diagram merupakan diagram yang menunjukkan konfigurasi komponen dalam proses eksekusi aplikasi (Sukamto & Shalahudin, 2018, p. 154); Deployment diagram digunakan untuk memodelkan hal-hal seperti sistem tambahan dan sistem client/server; Salah satu fungsi dari deployment diagram yaitu untuk menggambarkan dan menspesifikasikan proses. Berikut symbol-simbol yang digunakan untuk menggambarkan deployment diagram:

| SIMBOL | NAMA | KETERANGAN |
|---|---------|--|
|  | Package | Package merupakan sebuah bungkusan dari satu atau lebih node |

| SIMBOL | NAMA | KETERANGAN |
|---|------------|---|
|  | Node | Node biasanya mengacu pada perangkat keras (hardware) atau perangkat lunak yang tidak dibuat sendiri (software). Agar rancangan tetap konsisten, komponen yang dimasukkan ke dalam node harus sesuai dengan komponen yang telah didefinisikan sebelumnya pada diagram komponen. |
|  | Dependency | Kebergantungan antar <i>node</i> , arah panah mengarah pada <i>node</i> yang dipakai. |
|  | Link | Relasi antar <i>node</i> . |

B. Tinjauan Pustaka

Penelitian rujukan merupakan acuan yang dibutuhkan seorang peneliti untuk melakukan penelitian. Penelitian rujukan pada penelitian ini diambil berdasarkan kesamaan metode yang digunakan yaitu Algoritma *K-means*. Banyak penelitian yang menggunakan metode ini dalam berbagai kasus. Antara lain :

- (a) **Implementasi Metode K-Means Pada Sistem Informasi Geografis Pemetaan Kerusakan Jalan di Dinas Pekerjaan Umum dan Penataan Ruang Kabupaten Jepara** (Nuriyah Ramadhani et al., 2022) Jalan adalah sarana transportasi darat yang sangat penting untuk komunikasi antara kota dan desa. Kabupaten Jepara memiliki banyak sarana transportasi darat, terutama jalan. Namun, karena kurangnya laporan dan informasi dari warga, pemerintah Kabupaten Jepara masih memiliki banyak jalan yang rusak dan belum diperbaiki. Akibatnya, pemerintah tidak dapat melakukan penanganan yang cepat dan tepat. Membangun sarana dan prasarana dalam bidang infrastruktur dan penataan ruang adalah tanggung jawab Dinas Pekerjaan Umum dan Penataan Ruang (DPUPR). Tugas ini bertanggung jawab kepada bupati melalui sekretaris daerah. Untuk mengatasi masalah ini, Sistem Informasi Geografis Pemetaan Kerusakan Jalan (SIPEKA), yang dibangun melalui web dan menggunakan metode *k-means*, memungkinkan pemerintah, melalui DPUPR Kabupaten Jepara, untuk melanjutkan perbaikan jalan. Sistem dibangun menggunakan framework

CodeIgniter, dan pemetaan kerusakan jalan dibuat menggunakan OpenStreetMap. Pengujian sistem dilakukan menggunakan metode Black box untuk memastikan bahwa sistem bekerja dengan baik sesuai dengan rencana yang dibuat.

- (b) **Implementasi K-means Dalam Prioritas Perbaikan Jembatan Dan Saluran Air Di Kecamatan Ngronggot** (Isag & Irawan, 2021) Pembangunan nasional mencakup perbaikan infrastruktur. Pada dasarnya, pembangunan adalah proses yang dilakukan untuk mencapai tujuan tertentu, yaitu meningkatkan kesejahteraan masyarakat. Pemerintah kecamatan Ngronggot sangat memperhatikan kerusakan saluran air dan jembatan, yang menghambat sebagian besar aktivitas masyarakat. Karena masih menggunakan metode manual untuk mencatat hasil survei, dalam praktiknya telah dibuat keputusan yang tepat tentang prioritas lokasi perbaikan. Jadi, dengan menggunakan metode K-Means, sistem ini akan membantu kecamatan Ngronggot dalam menentukan lokasi perbaikan mana yang paling penting. Sistem ini akan mengikuti lima (lima) kriteria pembobotan yaitu tingkat kerusakan, fasilitas, anggaran, konstruksi, dan tahun pembuatan. Algoritma K-Means digunakan untuk meningkatkan keakuratan sistem penentuan prioritas lokasi perbaikan. Pada hasil perhitungan menggunakan metode K-Means, hasil akhir dibagi menjadi tiga kelompok. Kelompok pertama adalah perbaikan. Perhitungan berdasarkan kriteria ini menunjukkan tingkat kerusakan yang tinggi dan oleh karena itu harus diprioritaskan. Prioritas kedua mempunyai nilai prioritas remediasi sedang, dan prioritas ketiga mempunyai nilai prioritas remediasi ringan. Prioritas lokasi perbaikan jembatan dan saluran air dipengaruhi oleh bobot masing-masing kriteria dan hasil penerapan algoritma K-means sehingga diperoleh hasil prioritas untuk pelaksanaan perbaikan selanjutnya.
- (c) **Penerapan Algoritma K-Means Clustering untuk Pengelompokan Pembangunan Jalan pada Dinas Pekerjaan Umum dan Penataan Ruang** (Kurniadi et al., 2023). Di Dinas Pekerjaan Umum dan Penataan Ruang (PUPR) Kabupaten Garut, proses penentuan prioritas pembangunan jalan kurang efektif. Ini terkait dengan keabsahan pencapaian tujuan dan lama rekap. Ini adalah hasil dari pengelompokan yang buruk. Untuk mengelompokkan data pembangunan jalan berdasarkan jenis, panjang, dan nominal pembangunan, penelitian ini menggunakan metode data mining standar industri lintas industri dan algoritma k-means. Hasil penelitian menunjukkan bahwa kelompok yang dipilih memiliki lima klaster dengan nilai indeks Davies-Bouldin terkecil, yaitu 0,1617. Kelompok 1, 2, 3, 4, dan 5 memiliki karakteristik pembangunan jalan yang berbeda,

bergantung pada jenis pembangunan. Dengan mengetahui jenis kelompok dan ciri-cirinya, Dinas PUPR Kabupaten Garut dapat memprioritaskan pembangunan jalan kecil dibandingkan jalan sedang atau besar. Dengan melakukan ini, mereka dapat memprioritaskan pengajuan berdasarkan kelompoknya.

- (d) **Cacah Klaster pada Klasterisasi dengan Algoritma K-Means Menggunakan Silhouette Coeficient dan Elbow Method** (Guntara & Lutfi, 2023) Sangat penting untuk mengambil tindakan segera atas masalah undur diri siswa (DO). Karena banyaknya mahasiswa yang meninggalkan sekolah (DO), kualitas kinerja perguruan tinggi akan menurun. Ini berdampak pada kepercayaan masyarakat terhadap sistem pendidikan di sekolah. Algoritma K-Means akan digunakan untuk klasterisasi untuk mengetahui terjadinya mahasiswa DO. Untuk mengetahui seberapa konsisten hasil penentuan cacah klaster, dua metode *Elbow* dan *Silhouette Coefficient* digunakan bersamaan untuk menentukan cacah klaster yang paling baik (optimal). Jumlah NIM, jumlah semester tidak aktif, jumlah SKS, IPK, angkatan, dan jumlah rekor siswa (mahasiswa) akan dikumpulkan untuk kelas berjenis pembelajaran yang tidak diawasi. Berdasarkan data pelatihan, metode Elbow dan Silhouette digunakan untuk menemukan cacah klaster optimal yang sama, yaitu $K=2$. Berdasarkan nilai k ini, proses klasterisasi dilakukan menggunakan algoritma K-Means, yang menghasilkan kelas Cluster_0 dan Cluster_1. Jumlah klaster untuk Cluster_0 adalah 85,1%, yang terdiri dari 813 siswa, dan untuk Cluster_1, terdapat 14,9%, yang terdiri dari 142 siswa. Karena IPK mereka $<2,00$, jumlah sks yang diambil kurang dari 90 sks, dan jumlah semester tidak aktif (cuti) lebih dari 4 semester, cluster_0 ini didominasi oleh mahasiswa DO.
- (e) **Penerapan Algoritma K-Means Untuk Pemetaan Potensi Calon Mahasiswa Baru** (Triyono & Irmayansyah, 2022) Pengelompokan calon mahasiswa berdasarkan berbagai kriteria, yang kemudian dikelompokkan berdasarkan potensinya yang rendah atau tinggi, adalah proses pemetaan potensi calon mahasiswa baru. Proses ini digunakan untuk membantu bagian promosi mendapatkan data dan informasi yang diperlukan untuk menentukan rencana promosi yang lebih lanjut. Dengan menggunakan Algoritma K-Means Clustering, penelitian ini dapat memberikan pemetaan potensi mahasiswa baru dengan menganalisis kelompok data awal, mengubah data awal, dan melakukan perhitungan pengelompokan. Kemudian, hasil perhitungan pengelompokan dapat dianalisis kembali untuk mengetahui dari mana sekolah, lokasi sekolah, dan informasi dari mana masing-masing anggota di masing-masing kelompok berasal. Ini dilakukan untuk mengidentifikasi prospek siswa baru dan membantu

bagian promosi dengan data dan informasi untuk menentukan rencana promosi untuk periode berikutnya. Dengan menggunakan rasio silhouette terhadap algoritma K-Means, uji validitas cluster telah dilakukan. Nilai yang didapat sebesar 0.6113 menunjukkan bahwa cluster yang dibuat termasuk dalam kategori "struktur menengah", yang didasarkan pada tabel kategori silhouette yang dibuat oleh Kauffman dan Rousseuw.

- (f) Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni** (Septianingsih, 2022) Definisi nasional dan global menguraikan berbagai indikator kriteria untuk perumahan yang layak, yang meliputi akses ke sanitasi yang memadai, ruang hidup yang cukup, ketahanan bangunan, ketersediaan air minum, dan keamanan tempat tinggal. Selain itu, pemerintah mempertimbangkan variabel status kepemilikan rumah. Penelitian ini bertujuan untuk mengkategorikan provinsi-provinsi di Indonesia menurut sembilan variabel yang terkait dengan perumahan yang layak: akses ke perumahan yang layak, air minum yang aman, penerangan listrik, sanitasi yang memadai, luas area hunian kurang dari 7,2 m², luas lantai terbesar, dinding bambu terluas, atap serat palem terluas, dan apakah rumah tersebut disewa atau dikontrak. Dengan memanfaatkan analisis K-Means Clustering bersama dengan metode validasi lebar siluet koefisien, penelitian ini menilai tingkat validasi cluster. Analisis mengidentifikasi empat cluster, dengan cluster 1 yang berisi 10 anggota, cluster 2 terdiri dari 19 anggota, cluster 3 memiliki 3 anggota, dan cluster 4 terdiri dari 2 anggota. Klaster 1 menunjukkan persentase terendah di antara klaster lainnya untuk variabel yang berkenaan dengan area perumahan di bawah 7,2 m², atap jerami terluas, dan kepemilikan rumah sewa atau kontrak. Sebaliknya, Klaster 2 membanggakan persentase tertinggi untuk akses ke perumahan layak dan air minum bersih jika dibandingkan dengan klaster lainnya. Klaster 3 menunjukkan persentase terendah relatif terhadap klaster lainnya untuk akses ke perumahan layak, air minum bersih, penerangan listrik, dan sanitasi yang layak, sementara itu memiliki persentase tertinggi untuk area perumahan di bawah 7,2 m², lantai tanah terluas, dan ukuran dinding. Terakhir, Klaster 4 mendapat peringkat tertinggi di antara variabel untuk penerangan listrik, sanitasi yang layak, atap jerami terluas, dan kepemilikan rumah sewa atau kontrak, tetapi memiliki persentase terendah untuk lantai tanah terluas dan dinding bambu terluas dibandingkan dengan klaster lainnya.
- (g) Penerapan K-Means Dan Rank Order Centroid Pada Proporsi Individu Dengan Keterampilan Teknologi Informasi Dan Komputer** (Nurfitriana & Voutama, 2023) Sumber daya manusia yang berkualitas tinggi diperlukan untuk

mendukung zaman yang kian berkembang karena kemajuan teknologi yang begitu cepat. Penelitian ini bertujuan untuk mengelompokkan orang-orang di Indonesia dengan keterampilan teknologi informasi dan komputer berdasarkan wilayah. Untuk menilai akurasi, penelitian ini menggunakan clustering K-Means dan metode Rank Order Centroid; metode evaluasi clustering Davies-Bouldin Index juga digunakan. Clustering K-means adalah algoritma sederhana yang tidak membutuhkan target kelas. Metode ROC digunakan karena proses K-Means gagal menemukan centroid awal. Berdasarkan data yang diakses dari situs Badan Pusat Statistik Nasional, proporsi orang berusia 15-59 tahun yang memiliki keterampilan TIK menurut provinsi selama periode 2017–2021 dibagi menjadi 3 cluster. 8 provinsi termasuk dalam cluster tingkat tinggi, 22 provinsi termasuk dalam cluster tingkat sedang, dan 4 provinsi termasuk dalam cluster tingkat rendah. Nilai evaluasi DBI adalah 0,163625, yang hampir 0. Ini menunjukkan bahwa kualitas akurasi dari hasil c Hasil penelitian ini dapat digunakan oleh pemerintah untuk meningkatkan kualitas sumber daya manusia di daerah yang memiliki keterampilan IT dan komputer yang kurang. Menggunakan algoritma clustering tambahan dan ROC sebagai perbandingan, saran untuk penelitian lebih lanjut.

- (h) **Clustering Pemetaan Tingkat Kemiskinan Di Provinsi Jawa Barat Menggunakan Algoritma K-Means** (Nugraha, 2023) Kemiskinan adalah salah satu masalah besar yang masih dihadapi Indonesia. Pemerintah harus memprioritaskan masalah kemiskinan, yang memiliki banyak segi dalam masyarakat dan merupakan masalah yang kompleks. Pada awal pandemi, tingkat kemiskinan di Indonesia meningkat sebesar 9,78% pada bulan Maret dan 10,19% pada bulan September. Dengan 17.856 kasus, Jawa Barat menempati tingkat kemiskinan ekstrem tertinggi di seluruh Indonesia, menurut radarsukabumi.com. Penelitian ini menggunakan metode clustering dengan algoritma k-means dan QGIS untuk pemetaan daerah rawan kemiskinan. Hasil klasifikasi daerah rawan kemiskinan di Provinsi Jawa Barat dari tahun 2015 hingga 2020 menunjukkan bahwa dua belas kabupaten/kota tidak rentan, empat belas kabupaten/kota rawan, dan satu kabupaten/kota sangat rentan. Hasil evaluasi kelompok dengan koefisien siluet yang sama adalah 0,55. Hasil evaluasi masuk dalam kategori struktur sedang dengan interpretasi penempatan cluster yang masuk akal.
- (i) **Penerapan normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliah Tunggal** (Kusnaldi et al., 2022) Dalam memberikan bantuan

UKT kepada mahasiswa di Universitas Budi Darma, ada kendala. Batuan tersebut kurang sasaran dan labat. Hal ini disebabkan oleh fakta bahwa siswa yang menghadapi kesulitan keuangan adalah yang berhak atas bantuan ini. Oleh karena itu, data siswa harus diklasifikasikan menurut jenjang sosial mereka. Data mahasiswa Universitas Budi Darma dapat digunakan untuk menentukan siswa mana yang layak mendapatkan batuan. Dengan mengumpulkan informasi dari data mahasiswa. Normalisasi data dilakukan untuk mendapatkan data yang lebih akurat agar dapat digunakan terlebih dahulu. Normalisasi data diperlukan jika data mahasiswa dapat dikelompokkan dengan benar. Untuk menyamakan rentang nilai untuk setiap atribut dengan skala tertentu, metode skala desimal adalah transformasi data dengan normalisasi. Menggerakkan nilai desimal dari data ke arah yang diinginkan adalah salah satu metode normalisasi yang paling sering digunakan dalam normalisasi data. Setelah data dinormalisasi, metode selanjutnya adalah data mining untuk menggali informasi tentang data siswa. Ini mengumpulkan kelompok data siswa yang sangat penting untuk mendapatkan bantuan UKT. Algoritme K-Means digunakan untuk mengelompokkan data siswa. Pengujian metode manual, di mana terdapat tiga kelompok (kelompok 0, kelompok 1, dan kelompok 3), sama dengan pengujian aplikasi data mining, yaitu rapidminer. Oleh karena itu, kelompok atau kelompok 0 yang terdiri dari 22 orang adalah yang layak mendapatkan bantuan uang kuliah berdasarkan sampel. Tujuan dari penelitian ini adalah untuk mengevaluasi dampak penerapan normalisasi data pada teknik K-Means untuk mengelompokkan data mahasiswa yang digunakan sebagai rekomendasi saat memilih program bantuan UKT.

- (j) **(Ahmad Harmain, Paiman, Henri Kurniawan, Kusriani, Dina Maulina,2021) Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas** (Harmin et al., 2021) Potensi kebakaran hutan di Indonesia sangat tinggi, terutama selama musim kemarau. Oleh karena itu, diperlukan tindakan bersama untuk mengurangi risiko kebakaran hutan. Untuk mencapai tujuan ini, dibutuhkan teknik teknologi yang lebih canggih dan terbaru untuk mengidentifikasi daerah yang memiliki kemungkinan besar terjadi kebakaran hutan. Untuk menentukan apakah suatu wilayah memiliki potensi kebakaran atau tidak, Anda dapat menggunakan informasi yang diberikan oleh sistem pencitraan dan informasi satelit (MODIS), yang mencakup parameter seperti Latitude, Longitude, Brightness, FRP (Power Radiative Fire), dan Keyakinan. Salah satu metode machine learning yang dapat digunakan untuk mengelompokkan wilayah-wilayah

tersebut adalah K-Means. Untuk menguji akurasi hasil pengelompokan K-Means, metode Davies Bouldin Index (DBI) dan Silhouette Coefficient dapat digunakan.

Tabel 2. 1 Tinjauan Pustaka

| No | Peneliti | Judul | Jurnal | Kontribusi |
|----|--|--|--|---|
| 1 | Dina Nuriyah Ramadhani , Aris Trijaka Harjanta , Aptanang Tyogi , Serli Agnes Ajhara | Implementasi Metode K-Means Pada Sistem Informasi Geografis Pemetaan Kerusakan Jalan Di Dinas Pekerjaan Umum Dan Penataan Ruang Kabupaten Jepara | Science And Engineering National Seminar 7 (Sens 7)/Vol. 7 No. 1 (2022): Sens 7 https://conference.upgris.ac.id/index.php/sens/article/view/3609 | Berkontribusi Dalam Membantu Menentukan Jumlah Cluster Yang Ingin Digunakan Pada Penelitian Ini |
| 2 | Arya Bagaskara Isag , Rony Heri Irawan | Implementasi K-Means Dalam Prioritas Perbaikan Jembatan Dan Saluran Air Di Kecamatan Ngronggot | Seminar Nasional Inovasi Teknologi/Vol. 5 No. 2 (2021) https://proceeding.unpkediri.ac.id/index.php/inotek/article/view/1146 | Berkontribusi Dalam Penerapan K-Means Dalam Prioritas Perbaikan |
| 3 | Dede Kurniadi , Yoga Handoko Agustin , Hari Ilham Nur Akbar, Ida Farida | Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Pembangunan Jalan Pada Dinas Pekerjaan Umum Dan Penataan Ruang | Jurnal Teknologi Informasi, Volume 20 No. 1, 64-77 Issn 1693-8348 E-Issn 2615-7128 https://ejournal.uksw.edu/Aiti/Article/View/7100 | Berkontribusi Dalam Membantu Memberi Informasi Tentang Jalan Pada Dpupr |
| 4 | Mohammad Guntara , Nafisatul Lutfi | Cacah Klaster Pada Klasterisasi Dengan Algoritma K-Means Menggunakan Silhouette Coefficient Dan Elbow Method | Jurnal Teknologi Informasi, Vol.2, E-Issn: 2962-4118, https://ejournal.akakom.ac.id/index.php/juti/article/view/944 | Berkontribusi Sebagai Acuan Dalam Perhitungan Silhouette Coefficient |
| 5 | Safi'l Eko Triyono , Irmayansyah | Penerapan Algoritma K-Means Untuk Pemetaan Potensi Calon Mahasiswa Baru | Jurnal Ilmiah Teknologi- Informasi & Sains, Volume 12 http://teknis.stikombinaria.ga.ac.id/index.php/JBS | Berkontribusi sebagai acuan dalam perhitungan silhouette coefficient |
| 6 | Amin Septianingsih | Analisis K-Means Clustering Pada Pemetaan Provinsi Indonesia Berdasarkan Indikator Rumah Layak Huni | Jurnal Lebesgue: Jurnal Ilmiah Pendidikan Matematika, Matematika Dan Statistika http://lebesgue.lppmbina.bangsa.id/index.php/home | Berkontribusi Sebagai Acuan Dalam Perhitungan Silhouette Coefficient |
| 7 | Diana Nurfitriana, Apriade Voutama | Penerapan K-Means Dan Rank Order Centroid Pada Proporsi Individu Dengan Keterampilan Teknologi Informasi Dan Komputer | Jurnal Teknologi Terpadu Vol. 9 No. 2 2023, 70-78 https://journal.nurulfikri.ac.id/index.php/jtt | Berkontribusi Sebagai Acuan Penentuan Jumlah Cluster |
| 8 | I Wayan Setya Adi Nugraha | Clustering Pemetaan Tingkat Kemiskinan Di Provinsi Jawa Barat Menggunakan Algoritma K-Means | Jurnal Ilmiah Wahana Pendidikan, Januari 2023, 9 (2), 234-244 https://jurnal.peneliti.net/index.php/JIWP | Berkontribusi Dalam Penggunaan Metode Data Mining CRISP-DM |
| 9 | Muhammad Rafli Kusnaldi, Timotius Gulo, Soeb Aripin | Penerapan normalisasi Data Dalam Mengelompokkan Data Mahasiswa Dengan Menggunakan Metode K-Means Untuk Menentukan Prioritas Bantuan Uang Kuliahtunggal | Journal Of Computer System And Informatics (Josyc) https://ejournal.seminar-id.com/index.php/josyc/article/view/2112 | Berkontribusi Sebagai Acuan Dalam Perhitungan Normalisasi Data |

| No | Peneliti | Judul | Jurnal | Kontribusi |
|----|--|--|--|--|
| 10 | Ahmad Harmain, Paiman, Henri Kurniawan, Kusriani, Dina Maulina | Normalisasi Data Untuk Efisiensi K-Means Pada Pengelompokan Wilayah Berpotensi Kebakaran Hutan Dan Lahan Berdasarkan Sebaran Titik Panas | Jurnal TEKNIMEDIA - Volume 2, Nomor 2, Desember 2021: 83 – 89 https://jurnal.stmikswnw.ac.id/Index.Php/Teknimedia/Article/Download/49/43 | Berkontribusi Sebagai Acuan Dalam Perhitungan Normalisasi Data |

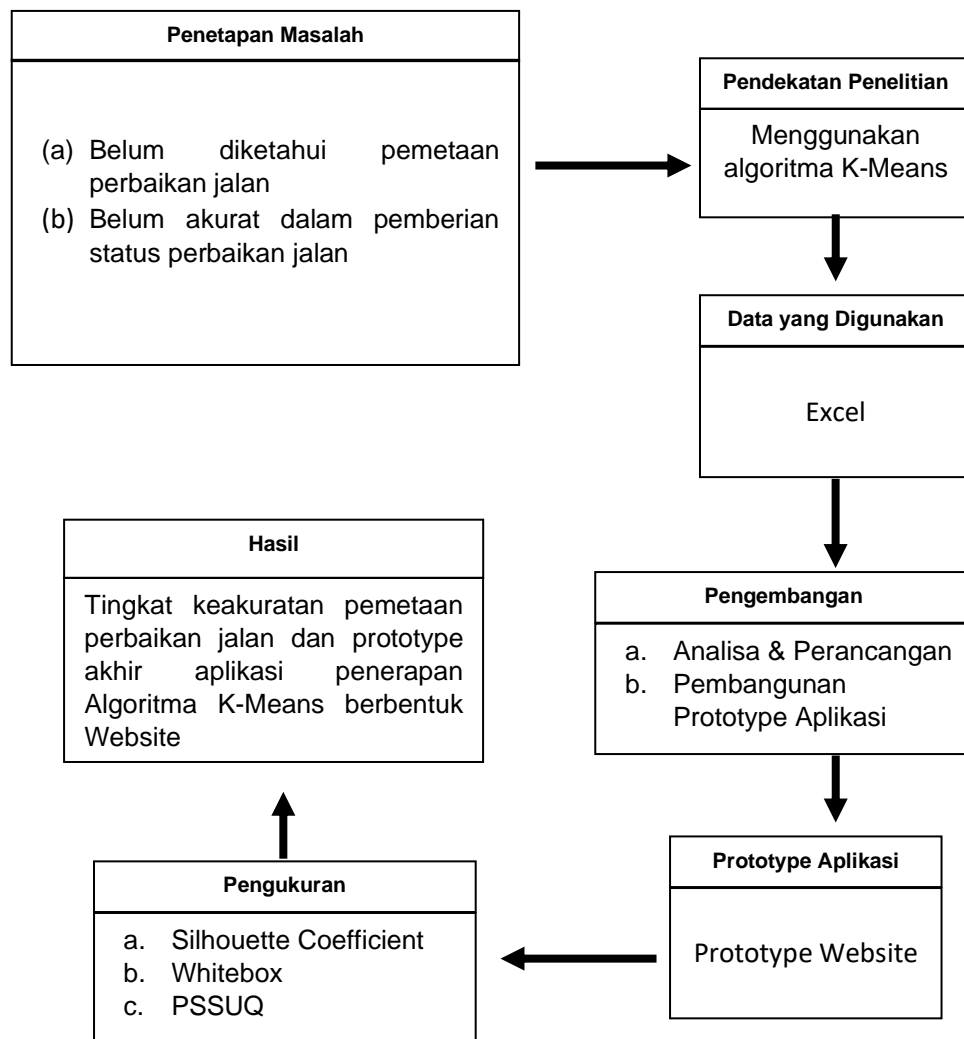
Berdasarkan tinjauan pustaka yang sudah dipaparkan pada tabel 2.1 didapatkan pengetahuan yang dijadikan rujukan dalam pelaksanaan penelitian ini. Kontribusi dan rujukan penelitian ini memberikan pengetahuan ilmu mengenai permasalahan pada pemetaan perbaikan jalan dengan proses perhitungan K-means. Penelitian dilakukan karena ada pengembangan dari penelitian sebelumnya dengan menerapkan metode K-means dan pengujian hasil menggunakan algoritma *Silhouette Coefficient* pada pemetaan perbaikan jalan dengan menggunakan variabel kondisi, fungsi jalan, dan perkerasan jalan.

C. Kerangka Berfikir

Berikut kerangka pemikiran yang disusun pada penelitian ini dibuat untuk mewakili konsep pemecahan masalah penelitian yang meliputi masalah yang akan diteliti, metode penelitian, dan hasil dari penelitian, berikut dijelaskan pada gambar sebagai berikut dan keterangannya:

Keterangan kerangka pemikiran pada gambar 2.5 dapat dijelaskan sebagai berikut :

1. Pada penelitian yang dilakukan ini diawali dengan penetapan permasalahan mengenai objek dengan melakukan identifikasi masalah sebagai berikut :
 - (a) Belum diketahui pemetaan perbaikan jalan ;
 - (b) Belum akurat dalam pemberian status perbaikan jalan
2. Pada penelitian ini pendekatan yang dilakukan yaitu menggunakan algoritma K-Means;
3. Penelitian ini menggunakan data berupa data excel;
4. Pengembangan yang dilakukan pada penelitian ini dengan melakukan analisa data dan perancangan serta pembangunan prototype aplikasi;
5. Prototype aplikasi yang akan dibangun berupa prototype website;
6. Setelah itu dilakukan pengukuran dengan menggunakan algoritma *Silhouette Coefficient* sebagai uji keakuratan hasil penerapan algoritma k-means, pengujian *Whitebox* sebagai uji sistem, dan PSSUQ untuk uji kebergunaan;
7. Hasil dari penelitian ini adalah tingkat keakuratan pemetaan perbaikan jalan dan prototype aplikasi penerapan algoritma K-Means.



Gambar 2. 5 Kerangka Berfikir

D. Hipotesis

Metode k-means merupakan metode membagi data kedalam cluster berdasarkan kesamaan jenis atau karakteristik data. Menurut beberapa penelitian sebelumnya, metode k-means ini mampu menyelesaikan masalah dalam ketepatan pembagian data kedalam classnya secara akurat. Metode k-means yang bisa membagi data dengan akurat kedalam classnya maka hal ini berkaitan dengan penelitian ini yaitu mengelompokkan data yang diduga dapat diketahui pemetaan perbaikan jalan agar dapat membantu dalam pengambilan keputusan yang tepat dalam perbaikan jalan. Berdasarkan pemahaman diatas dan penelitian yang telah dilakukan oleh Nuriyah Ramadhani et al., Implementasi Metode K-Means Pada Sistem Informasi Geografis Pemetaan Kerusakan Jalan di Dinas Pekerjaan Umum dan Penataan Ruang Kabupaten Jepara, 2022, dengan hasil penelitian mendapatkan pengelompokan kerusakan jalan kedalam tingkatannya yang tepat, maka dapat ditetapkan hipotesisnya yaitu metode k-means diduga dapat mengetahui pemetaan perbaikan jalan dan hasil yang akurat.