

BAB II KERANGKA TEORITIS

A. Landasan Teori

1. Data Mining

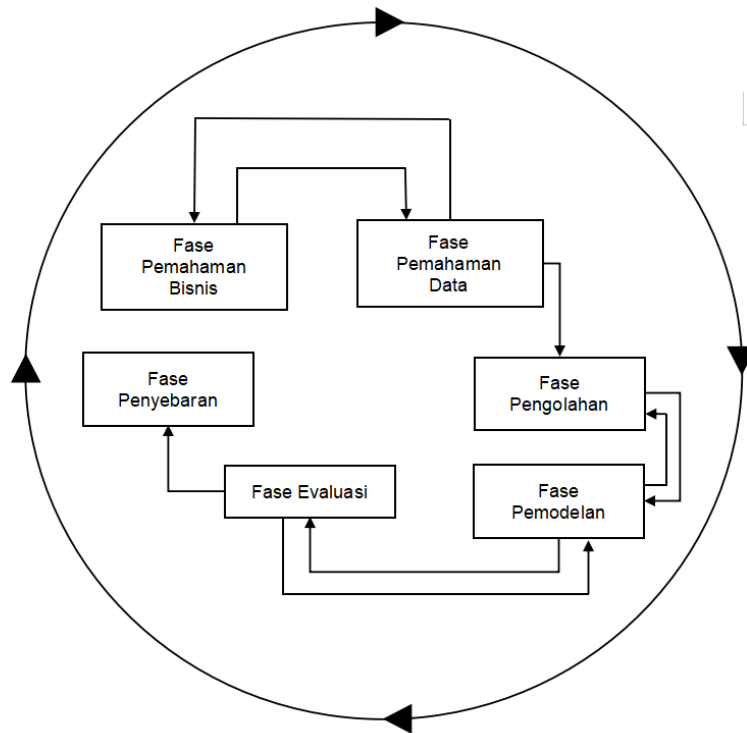
Data Mining adalah suatu proses ekstraksi pengetahuan atau informasi yang berharga dari suatu set data yang besar dan kompleks. Tujuan utama dari Data Mining adalah mengidentifikasi pola, hubungan, atau informasi yang mungkin tidak terlihat secara langsung dalam data, sehingga memberikan wawasan yang lebih dan bernilai (Prastyadi, 2024, p. 2). Data Mining memiliki 5 (lima) peranan secara umum yaitu estimasi, prediksi, klasifikasi, *clustering* dan asosiasi (Suntoro, 2019, p. 3).

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 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.

2. CRISP-DM

CRISP-DM merupakan salah satu metode yang menjelaskan tentang tahapandatamining(Ville,2001,p.37).CRISP-DMdigunakansebagaistandar 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

Sumber : (Kusrini & Luthfi, 2009, p. 8)

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

- (1) Fase Pemahaman Bisnis (*Business Understanding Phase*), merupakan tahapan yang dilakukan untuk 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 yang dilakukan untuk pengumpulan data, kemudian dianalisis untuk mencari informasi tentang pola data dan kualitas data;
- (3) Fase Pengolahan Data (*Data Preparation Phase*), merupakan tahapan untuk menganalisis dan menyiapkan data dengan menyesuaikan kasus dan variabel yang digunakan pada 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 untuk 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 yang menggunakan model data mining dengan pembuatan laporan atau penerapan proses data mining secara berulang pada perusahaan.

Setelah semua proses dilewati, misalkan kita sudah berada pada Tahap *Deployment*, maka proses tidak berhenti dan selesai. Masih ada proses terus berlanjut yang disebut dengan *Continuous Monitoring* dan *Improvement*. Dengan demikian, setiap implementasi model yang dihasilkan nanti terus dimonitor dan dipantau performanya dari waktu ke waktu. Jika performanya turun karena faktor eksternal yang berubah, maka perlu dilakukan proses *Training* (pembelajaran) ulang menggunakan data-data terbaru atau bahkan merombak model dan memulai lagi dari awal (Hapsari dkk, 2021, pp. 137-139).

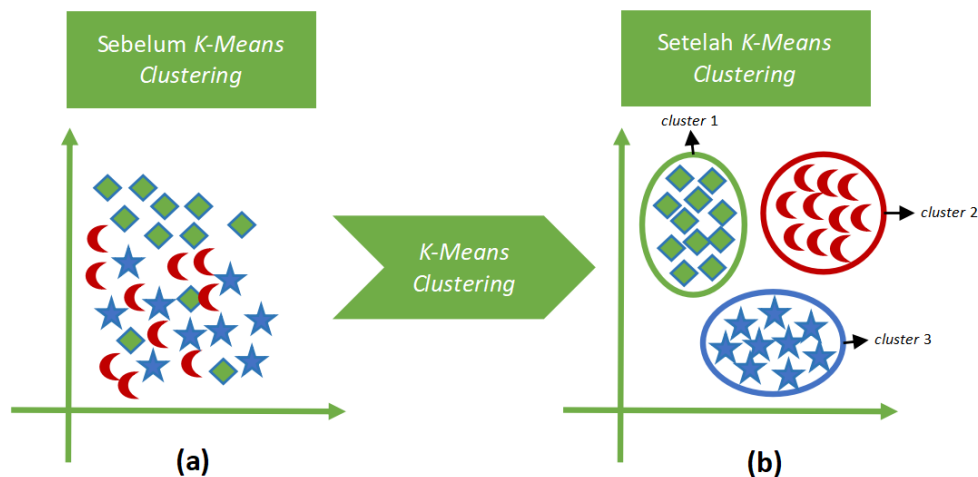
3. Klasterisasi (*Clustering*)

Analisis *Cluster* atau *Clustering* adalah proses memartisi sekumpulan objek data (pengamatan) menjadi himpunan bagian. Setiap subset adalah sebuah *cluster* sehingga objek dalam *cluster* mirip satu sama lain namun berbeda dengan objek di *cluster* lain (Kamber et al, 2012, p. 362). *Clustering* adalah metode yang digunakan untuk membagi rangkaian data menjadi *group* berdasarkan kesamaan yang telah ditentukan sebelumnya. Dengan menggunakan *clustering* kita dapat mengidentifikasi daerah yang padat, menentukan pola-pola distribusi secara keseluruhan dan menemukan keterkaitan yang menarik antara atribut-atribut data (Retno Tri Vlandari, 2017, p. 54).

4. Konsep Algoritma K-Means

Menurut (Ibnu Daqiqil, 2021, p. 217), *K-Means* adalah salah satu “*unsupervised machine learning algorithms*” yang paling sederhana dan populer. Tujuan dari algoritma ini adalah untuk menemukan grup dalam data, dengan jumlah grup yang diwakili oleh variabel *k*. Variabel *k* sendiri adalah jumlah *cluster* yang kita inginkan. Metode *K-Means Clustering* berusaha mengelompokkan data yang ada ke dalam beberapa kelompok, dimana data dalam satu kelompok mempunyai karakteristik yang sama antara satu dengan lainnya dan mempunyai karakteristik yang berbeda dengan data yang ada didalam kelompok yang lain. Karakteristik yang sama itu ditandai dengan jarak atau *distance* yang lebih dekat, mirip seperti KNN. Dengan kata lain, metode *K-Means Clustering* bertujuan untuk meminimalisir *objective function* yang diset dalam proses *clustering* dengan cara meminimalkan variasi antar data didalam suatu *cluster* dan memaksimalkan variasi dengan data yang ada di *cluster* lainnya.

Putra dkk, (2023, p.85) menjelaskan bahwa Algoritma pengelompokan *K-Means* dalam *machine learning* adalah salah satu algoritma *unsupervised learning* (pembelajaran tanpa pengawasan) yang paling sederhana, namun memiliki hasil yang mendekati benar. Algoritma ini memiliki beberapa penerapan aplikasi, termasuk klasifikasi dokumen, segmentasi gambar, sistem rekomendasi, dll.



Gambar 2 . 2 Ilustrasi Data Cluster

Sumber : (Putra dkk, 2023)

Algoritma *K-Means* merupakan algoritma non hirarki (*non-hierarchical*) yang berasal dari metode data *clustering*. Algoritma *K-Means* dapat melakukan partisi data ke dalam kelompok, sehingga data yang memiliki karakteristik yang sama dikelompokkan kedalam set yang sama. Data yang memiliki karakteristik yang berbeda selanjutnya akan dikelompokkan kedalam set yang berbeda (Putra dkk, 2023, p. 85).

Prinsip utama dari algoritma ini adalah menyusun k dari sekumpulan data. Tujuan *K-Means* untuk meminimalkan fungsi objektif pada set kelompok dalam proses pengelompokan. Pada umumnya teknik ini berusaha meminimalkan variasi dalam suatu kelompok dan memaksimalkan variasi antar kelompok lainnya.

Menurut (Daniel dan Eko, 2013, p. 3), langkah-langkah Algoritma *K-Means* adalah sebagai berikut :

- (a) Pilih secara acak k buah data sebagai pusat *cluster*;
- (b) Jarak antara data dan pusat *cluster* dihitung menggunakan *Euclidian distance*. Untuk menghitung jarak semua data ke setiap titik pusat *cluster* dapat menggunakan teori jarak *Euclidian* yang dirumuskan sebagai berikut :

(c)

$$D_{(i,j)} = \sqrt{(X_{1i} - X_{1j})^2 + (X_{2i} - X_{2j})^2 + \dots + (X_{ki} - C_{kj})^2}$$

dimana :

$D_{(i,j)}$ = Jarak data ke i ke pusat *cluster* j

X_{ki} = Data ke i pada atribut data ke k

X_{kj} = Titik pusat ke j pada atribut data ke k

(d) Data ditempatkan dalam *cluster* yang terdekat, dihitung dari tengah *cluster*;

(e) Pusat *cluster* baru akan ditentukan bila semua data telah ditetapkan dalam *cluster* terdekat;

(f) Proses penentuan pusat *cluster* dan penempatan data dalam *cluster* diulang sampai nilai *centroid* tidak berubah lagi.

Contoh kasus penerapan algoritma *K-Means* beserta tahapan algoritma *K-Means* (Marisa, et al., 2023, p. 42) adalah sebagai berikut :

Tabel 2 . 1 Langkah ke-1 contoh kasus penerapan algoritma *K-Means*

SME Player	Criteria			
	Cr1 (SME Scope)	Cr2 (Market)	Cr3 (Product)	Cr4 (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

Sumber : (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 :

- Cr1 SME Scope : 1=Small, 2=Micro, 3=Macro
 Cr2 Market : 1=Regional, 2=National, 3=International
 Cr3 Product : 1=Fashion, 2=Kriya, 3=Kuliner, 4=Digital Creative, 5=Furniture, 6=Home Stay, 7=Foto and Video, 8=Buku, 9=Other
 Cr4 Market Place : 1=Digital, 2=Direct Market, 3= Mix Market

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

Tabel 2 . 2 Langkah ke-2 contoh kasus penerapan algoritma *K-Means*

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	

Sumber : (Marisa, et al., 2023, p. 42)

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 diatas 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 diatas 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 tabel 2.3 kolom Cr3 dibawah ini.

Tabel 2 . 3 Langkah ke-3 contoh kasus penerapan algoritma K-Means

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

Sumber : (Marisa, et al., 2023, p. 42)

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 (S_n, C_1) , (S_n, C_2) , (S_n, C_3) , kemudian *cluster*

ditentukan berdasarkan nilai terkecil yang diperoleh salah satu nilai *Euclidean distance* pada masing-masing item set. Hasil penentuan *cluster* dapat terlihat pada tabel 2.4 di bawah ini.

Tabel 2 . 4 Hasil penentuan *cluster*

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-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

Sumber : (Marisa, et al., 2023, p. 42)

Langkah 4, menentukan *centroid* untuk iterasi 2, dengan menghitung rata-rata dari hasil penjumlahan data masing-masing kelompok *cluster* yang terlihat pada tabel 2.5 di bawah ini.

Tabel 2 . 5 Langkah ke- 4 contoh kasus penerapan algoritma *K-Means*

SME	Cr1	Cr2	Cr3	Cr4
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

SME	Cr1	Cr2	Cr3	Cr4
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

Sumber : (Marisa, et al., 2023, p. 43)

Dengan hasil perhitungan nilai rata-rata data tiap kelompok *cluster* pada data diatas didapatkan 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 diatas 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 diatas 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 diatas 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 tabel 2.6 kolom Cr3 dibawah ini.

Tabel 2 . 6 Langkah ke-6 contoh kasus penerapan algoritma K-Means

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

Sumber : (Marisa, et al., 2023, p. 43)

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) , (S_n, C_2) , (S_n, C_3) , kemudian *cluster* ditentukan berdasarkan nilai terkecil yang diperoleh salah satu nilai *Euclidean distance* pada masing-masing item set. Hasil penentuan *cluster* dapat terlihat pada tabel 2.7 di bawah ini.

Tabel 2 . 7 Hasil penentuan cluster

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	$D(S_n, C_1)$	$D(S_n, C_2)$	$D(S_n, C_3)$	Nearest Distance	Cluster
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-10	1,061	2,646	2,280	1,061	C1

Sumber : (Marisa, et al., 2023, p. 43)

Dari hasil iterasi kedua tidak ada perubahan posisi *cluster*, sehingga posisi iterasi berhenti sampai pada iterasi kedua dan *cluster* yang dihasilkan adalah yang disajikan pada tabel 2.8 dibawah ini.

Tabel 2 . 8 Hasil penentuan *cluster*

SME	Cr1	Cr2	Cr3	Cr4
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

Sumber : (Marisa, et al., 2023, p. 44)

Data dibawah ini meletakkan dataset kedalam hasil kelompok *cluster* yang terbentuk. Menghasilkan 3 *cluster* dengan rincian C1 memiliki 4 anggota, C2 memiliki 1 anggota dan C3 memiliki 5 anggota seperti pada tabel 2.9 dibawah ini.

Tabel 2 . 9 Hasil penentuan *cluster*

SME	Cr1	Cr2	Cr3	Cr4
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	Micro	International	Kuliner	Mix Market
Cluster3				
SME Player-2	Micro	Regional	Kuliner	Mix Market
SME	Cr1	Cr2	Cr3	Cr4
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

Sumber : (Marisa, et al., 2023, p. 44)

5. Sistem Pendukung Keputusan (SPK) atau *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* (Turban, et al., 2005, p.109).

(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) *Software System/User Interface*

Merupakan gabungan antara komponen *database management* dan model *base* yang merupakan komponen *User Interface*. *User Interface* menampilkan tampilan sistem yang dimengerti oleh pengguna.

6. Pelayanan Administrasi Kependudukan

Pelayanan Publik adalah kegiatan atau rangkaian kegiatan dalam rangka pemenuhan kebutuhan pelayanan sesuai dengan peraturan perundang-undangan bagi setiap warga negara dan penduduk atas barang, jasa dan/atau pelayanan administratif yang disediakan oleh penyelenggara pelayanan publik (Undang-undang Republik Indonesia, Nomor 25 Tahun 2009, p. 3).

Administrasi Kependudukan adalah rangkaian kegiatan penataan dan penertiban dalam penerbitan dokumen dan Data Kependudukan melalui pendaftaran penduduk, pencatatan sipil, pengelolaan informasi administrasi kependudukan serta pendayagunaan hasilnya untuk pelayanan publik dan pembangunan sektor lain. Data Kependudukan adalah data perseorangan dan/atau data agregat yang terstruktur sebagai hasil dari kegiatan pendaftaran penduduk dan pencatatan sipil. Dokumen kependudukan adalah dokumen resmi yang diterbitkan oleh Dinas Kependudukan dan Pencatatan Sipil Kabupaten/Kota yang mempunyai kekuatan hukum sebagai alat bukti autentik yang dihasilkan dari pelayanan pendaftaran penduduk dan pencatatan sipil (Peraturan Pemerintah Republik Indonesia, Nomor 40 tahun 2019, p. 3).

Akta Kelahiran adalah salah satu dokumen hasil pencatatan sipil yang meregistrasi setiap kelahiran sebagai peristiwa kependudukan. Setiap kelahiran wajib dilaporkan oleh penduduk kepada pemerintah (instansi pelaksana) di tempat terjadinya peristiwa kelahiran paling lambat 60 (enam puluh) hari sejak kelahiran (Peraturan Bupati Bogor, Nomor 16 Tahun 2013, p. 4).

Persyaratan pencatatan kelahiran adalah sebagai berikut (Peraturan Menteri Dalam Negeri Republik Indonesia, Nomor 9 Tahun 2016, p. 6) :

- (a) Surat keterangan lahir dari dokter/bidan/penolong kelahiran;
- (b) Akta Nikah / Buku Nikah / kutipan akta perkawinan;
- (c) KK dimana penduduk akan didaftarkan sebagai anggota keluarga;
- (d) KTP-el orangtua / wali / pelapor ; atau
- (e) Pasporn bagi WNI bukan penduduk dan orang asing.

Pencatatan kelahiran anak yang tidak diketahui asal-usulnya atau keberadaan orangtuanya dilakukan dengan (Peraturan Menteri Dalam Negeri Republik Indonesia, Nomor 9 Tahun 2016, p. 6) :

- (a) Melampirkan Berita Acara Pemeriksaan (BAP) dari kepolisian; atau
- (b) Menggunakan SPTJM kebenaran data kelahiran yang ditandatangani oleh wali / penanggungjawab.

7. Unified Modeling Language (UML)

UML adalah sebuah bahasa yang berdasar pada grafik/gambar untuk memvisualisasi, menpesifikasikan, membangun dan pendokumentasian dari sebuah sistem pengembangan *software* berbasis O.O (*Object Oriented*). Salah satu dari tahapan SDLC adalah desain. Desain bertujuan agar *software* yang akan dibuat dapat memenuhi kebutuhan *user* baik secara fungsional maupun non fungsional ke dalam model untuk memodelkan sesuai dengan yang diperlukan oleh bahasa pemodelan. Bahasa pemodelan dapat berupa *pseudo-code* yaitu *code*, gambar, diagram atau deskripsi yang menggambarkan sebuah sistem (Sumirat dkk, 2023, pp. 73-74).

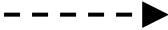
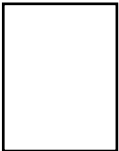
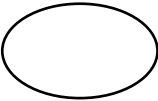
Diagram - diagram yang didefinisikan oleh UML antara lain :

(a) Use Case Diagram

Use Case Diagram adalah komponen yang menggambarkan fungsional dalam sebuah sistem. Sehingga konsumen maupun pembuat saling mengenal dan mengerti alur sistem yang akan dibuat (Sumirat dkk, 2023, p. 82).

Tabel 2 . 10 Simbol Use Case Diagram

NO	GAMBAR	NAMA	KETERANGAN
1		<i>Actor</i>	Menspesifikasikan himpunan peran yang pengguna mainkan ketika berinteraksi dengan <i>use case</i>
2		<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri (<i>independent</i>) akan mempengaruhi elemen yang bergantung padanya yaitu elemen yang tidak mandiri (<i>independent</i>)
3		<i>Generalization</i>	Hubungan dimana objek anak (<i>descendent</i>) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk (<i>ancestor</i>)

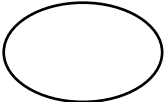
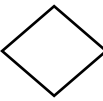
NO	GAMBAR	NAMA	KETERANGAN
4		<i>Include</i>	Menspesifikasikan bahwa <i>use case</i> sumber secara <i>eksplisit</i>
5		<i>Extend</i>	Menspesifikasikan bahwa <i>use case</i> target memperluas perilaku dari <i>use case</i> sumber pada suatu titik yang diberikan
6		<i>Association</i>	Apa yang menghubungkan antara objek satu dengan objek lainnya
7		<i>System</i>	Menspesifikasikan paket yang menampilkan sistem secara terbatas
8		<i>Use Case</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu <i>actor</i>
9		<i>Collaboration</i>	Interaksi aturan-aturan dan elemen lain yang bekerja sama untuk menyediakan perilaku yang lebih besar dari jumlah dan elemen-elemennya (sinergi)
10		<i>Note</i>	Elemen fisik yang eksis saat aplikasi dijalankan dan mencerminkan suatu sumber daya komputasi

(Sumber : Sumirat dkk, 2023, p. 82)

(b) *Activity Diagram*

Activity Diagram menggambarkan berbagai alir aktivitas dalam sistem yang sedang dirancang, bagaimana masing-masing alir berawal, keputusan yang mungkin terjadi dan bagaimana mereka berakhir. *Activity Diagram* juga dapat menggambarkan proses paralel yang mungkin terjadi pada beberapa eksekusi. *Activity Diagram* menggambarkan berbagai alir aktivitas dalam sistem yang sedang dirancang, mulai darimana berasal, kemungkinan yang bisa terjadi dan bagaimana berakhirnya serta menggambarkan proses yang mungkin terjadi pada beberapa eksekusi (Hasanah dkk, 2020, pp. 79-80).

Tabel 2 . 11 Simbol *Activity Diagram*

NO	GAMBAR	NAMA	KETERANGAN
1		<i>Activity</i>	Memperlihatkan bagaimana masing-masing kelas antarmuka saling berinteraksi satu sama lain
2		<i>Action</i>	<i>State</i> dari sistem yang mencerminkan eksekusi dari suatu aksi
3		<i>Initial Node</i>	Bagaimana objek dibentuk atau diawali
4		<i>Activity Final Node</i>	Bagaimana objek dibentuk dan dihancurkan
5		<i>Fork Node</i>	Satu aliran yang pada tahap tertentu berubah menjadi beberapa aliran
6		<i>Decision Node</i>	Mewakili titikpercabangan bersyarat dengan satu <i>input</i> dan beberapa <i>output</i> , fungsinya mengarahkan alur berdasarkan kondisi tertentu

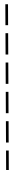
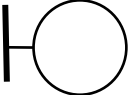
(Sumber : Hasanah dkk, 2020, p. 82)

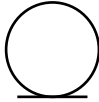
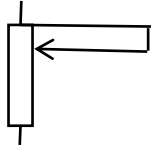
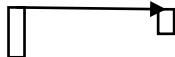
(c) *Sequence Diagram*

Sequence Diagram atau diagram urutan adalah sebuah diagram yang digunakan untuk menjelaskan dan menampilkan interaksi antar objek-objek dalam sebuah sistem secara terperinci. Selain itu *Sequence Diagram* juga akan menampilkan pesan atau perintah yang dikirim beserta waktu pelaksanaannya. Objek-objek

yang berhubungan dengan berjalannya proses operasi biasanya diurutkan dari kiri ke kanan. Tujuan utama pembuatan *Sequence Diagram* adalah untuk mengetahui urutan kejadian yang dapat menghasilkan *output* yang diinginkan. Selain itu, tujuan dari *Sequence Diagram* ini mirip dengan *Activity Diagram* seperti menggambarkan alur kerja dari sebuah aktivitas serta dapat menggambarkan aliran data dengan lebih *detail* termasuk data atau perilaku yang diterima atau dikirimkan. Berikut simbol *Sequence Diagram* (Sumirat dkk, 2023, pp. 86-87) :

Tabel 2 . 12 Simbol *Sequence Diagram*

NO	GAMBAR	NAMA	KETERANGAN
1		<i>Actor</i>	Komponen ini menggambarkan seorang pengguna (<i>user</i>) yang berada di luar sistem dan sedang berinteraksi dengan sistem. Dalam <i>Sequence Diagram</i> , aktor biasanya digambarkan dengan simbol <i>stick figure</i>
2		<i>Activation</i>	Komponen <i>activation box</i> ini merepresentasikan waktu yang dibutuhkan suatu objek untuk menyelesaikan tugasnya. Semakin lama waktu yang diperlukan, maka secara otomatis <i>activation box</i> -nya juga akan menjadi lebih panjang
3		<i>Life Line</i>	Komponen ini digambarkan dengan bentuk garis putus-putus yang terhubung dengan objek, sepanjang <i>Life Line</i> terdapat <i>activation</i>
4		<i>Boundary</i>	Komponen ini berisi kumpulan kelas yang menjadi <i>interface</i> atau interaksi antara satu atau lebih aktor dengan sistem seperti <i>form entry</i> dan <i>form cetak</i>
5		<i>Control</i>	Komponen ini merupakan suatu objek berisi logika aplikasi yang tidak memiliki tanggungjawab kepada entitas, contohnya adalah kalkulasi dan aturan bisnis yang melibatkan berbagai objek

NO	GAMBAR	NAMA	KETERANGAN
6		<i>Entity</i>	Komponen ini merupakan bagian dari sistem yang berisi kumpulan kelas berupa entitas-entitas yang membentuk gambaran awal sistem dan menjadi landasan untuk menyusun basis data
7		<i>Recursive</i>	Komponen ini menggambarkan pengiriman pesan yang dikirim untuk dirinya sendiri
8		<i>Message</i>	Komponen ini untuk menggambarkan komunikasi antar objek. <i>Message</i> biasanya muncul secara berurutan pada <i>Life Line</i> . Komponen <i>message</i> ini direpresentasikan dengan anak panah. Inti dari sebuah diagram urutan terdapat pada komponen <i>Life Line</i> dan <i>Message</i> ini

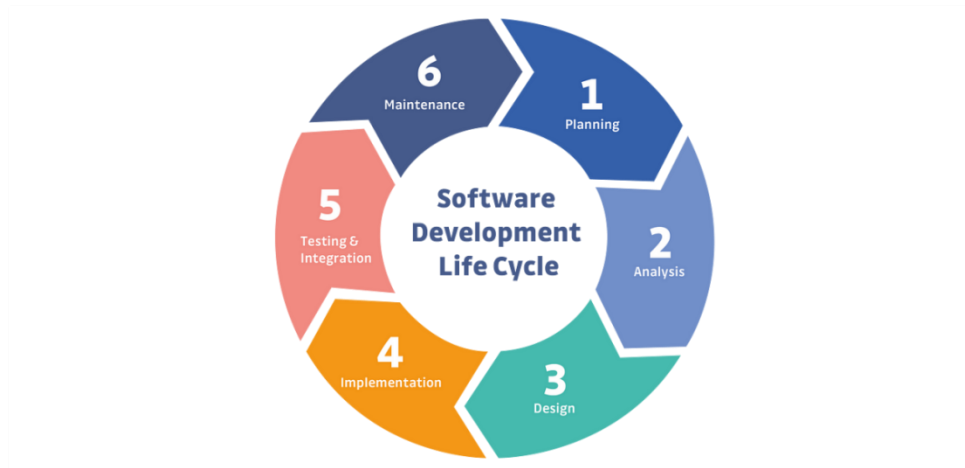
(Sumber : Sumirat dkk, 2023, pp. 87-88)

8. SDLC

SDLC adalah model konseptual dalam proyek manajemen yang menjelaskan tahapan kerja dalam pengembangan sistem informasi. Model ini mencakup studi kelayakan dari tahap awal hingga pemeliharaan aplikasi. Tujuan SDLC adalah menghasilkan sistem berkualitas tinggi yang memenuhi kebutuhan pelanggan dalam hal waktu, biaya, efektivitas dan efisiensi. Akibatnya, sistem ini menyediakan rencana lengkap untuk mengembangkan, memelihara dan menggantikan perangkat lunak tertentu (Wijayanto, 2024, pp. 35-37).

Metode SDLC akan membantu pengembang merancang dan mengembangkan serta mengidentifikasi masalah yang mungkin muncul selama proses pengembangan dan memantau mereka menyelesaikannya. Dengan kata lain, SDLC adalah tahapan kerja yang bermanfaat untuk menghasilkan suatu sistem yang memenuhi persyaratan klien. Ini terdiri dari kerangka yang berisi langkah - langkah yang diambil dalam proses pengembangan *software*. Didalam penerapannya, Model SDLC digunakan untuk merancang sistem yang terstruktur (Perancangan sistem traditional).

Berikut tahapan SDLC seperti Gambar 2.3 di bawah ini :



Gambar 2 . 3 Tahapan SDLC

(Sumber : Wijayanto, 2024, pp. 35-37)

Berikut ini adalah penjelasan dari Gambar 2.3 Tahapan SDLC :

(1) *Planning*

Dalam fase ini, elemen studi yang berkaitan dengan *feasibility study* atau kelayakan pengembangan sistem lebih ditekankan. Menentukan prioritas teknologi, menentukan tujuan dan ruang lingkup pengembangan, menciptakan dan mengkonsolidasikan tim pengembang dan menentukan serta menilai strategi.

(2) *Analysis*

Sebuah analisis sistem akan dilakukan mengenai cara *software* akan dijalankan, Hasilnya dapat mencakup kelebihan dan kekurangan sistem, fungsinya dan perbaikan.

(3) *Design*

Saat ini tugasnya adalah membuat *prototype* dan *output* lainnya termasuk dokumen yang berisi desain, pola dan bagian yang diperlukan untuk proyek.

(4) *Implementation*

Tahapan implementasi sistem dimana anda dapat menerapkan rancangan yang sudah dibuat pada tahapan sebelumnya.

(5) *Testing*

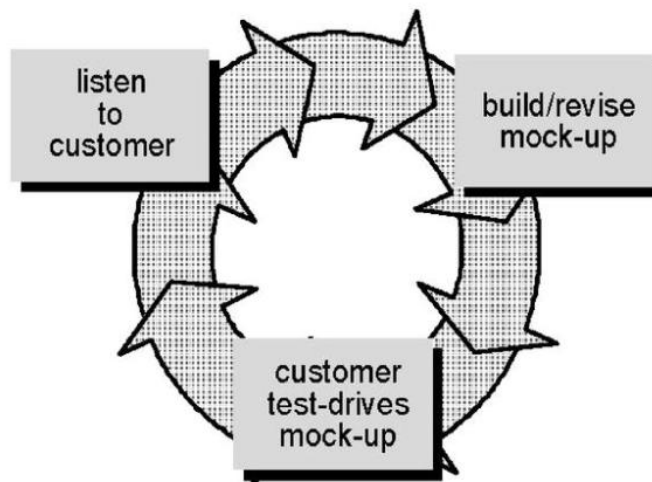
Tahapan ini adalah tahapan untuk dapat melakukan uji coba untuk memastikan bahwa operasinya berjalan lancar.

(6) *Maintenance*

Meskipun *software* sudah jadi, tetap harus dijaga dan diawasi agar *error* dapat diperbaiki secara cepat untuk tujuan perawatan agar data pengguna tetap aman (Wijayanto, 2024, pp. 35-37).

9. Metode *Prototype*

Metode *Prototype* atau sering disebut juga dengan *prototyping* merupakan sebuah metode pengembangan sistem yang didasarkan pada konsep *working model*. Selain itu *Prototype* juga didefinisikan sebagai alat yang memberikan ide bagi pembuat maupun pemakai potensial tentang cara sistem berfungsi dalam bentuk lengkapnya dan proses untuk menghasilkan sebuah *prototype* disebut *prototyping* (Yusniar, 2020, pp. 14-15).



Gambar 2 . 4 Metode *Prototype*

(Sumber : Yusniar, 2020)

Gambar 2.4 menggambarkan siklus pengembangan *prototype* dalam rekayasa perangkat lunak atau pengembangan produk secara iteratif. Terdapat tiga tahapan utama dalam siklus ini yaitu (Pressman, R. S., 2010, p. 7) :

(1) *Listen to costumer*

Tahapan ini merupakan langkah awal di mana pengembang mendengarkan kebutuhan, harapan dan umpan balik dari pelanggan atau pengguna akhir. Informasi ini menjadi dasar untuk membangun atau merevisi *prototype*.

(2) *Build/Revise mock-up*

Berdasarkan informasi dari pelanggan, pengembang kemudian membangun mock-up awal atau merevisi mock-up yang sudah ada. Mock-up adalah representasi awal dari produk yang belum sepenuhnya berfungsi, namun cukup untuk menunjukkan fitur-fitur utama.

(3) *Customer Test-drives Mock-up*

Setelah mock-up selesai, pelanggan melakukan uji coba terhadapnya. Tujuannya adalah untuk mendapatkan umpan balik langsung mengenai apakah mockup tersebut sudah sesuai dengan kebutuhan dan ekspektasi mereka.

Siklus ini kemudian berulang (iteratif), dimana hasil uji coba akan kembali digunakan dalam proses “*Listen to costumer*” untuk revisi lebih lanjut hingga produk akhir memenuhi standar dan keinginan pengguna.

Dalam setiap metode tertentu memiliki kelebihan dan kekurangan termasuk metode *prototype* tersebut. Kekurangan dan kelebihan metode *prototype* akan dijelaskan pada Tabel 2.13 berikut :

Tabel 2 . 13 Kelebihan dan Kekurangan Metode *Prototype*

NO	KELEBIHAN	KEKURANGAN
1	Pengguna (<i>user</i>) berperan aktif dalam pengembangan sistem	Kualitas aplikasi belum teruji dan belum mencantumkan pemeliharaan jangka panjang
2	Waktu yang digunakan lebih efisien	Algoritma dan Bahasa yang digunakan sederhana
3	Adanya komunikasi antar <i>user</i> dengan pengembang	Teknik rancangan tidak baik dilihat dari hubungan pelanggan dan <i>computer</i>
4	Pengembang dapat bekerja lebih baik	

(Sumber : Yusniar, 2020)

10. Database/DBMS

(a) Pengertian *Database*

Database adalah sekumpulan koleksi data yang berhubungan secara logikal dan sebuah deskripsi dari data tersebut. Didesain untuk menemukan keperluan informasi pada sebuah perusahaan (Connolly, p. 15). *Database* merupakan tempat penyimpanan data yang besar yang dapat digunakan secara bersamaan oleh banyak pengguna dan berisi deskripsi dari data itu sendiri selain data operasional milik perusahaan. Menurut (Hoffer, Prescott dan McFadden) *Database* adalah sekumpulan organisasi data yang berelasi, dapat memiliki banyak ukuran dan tingkat kompleksitas.

(b) Pengertian *DBMS (Database Management System)*

Menurut (James A.Hall) *DBMS* adalah sebuah sistem perangkat lunak khusus yang diprogram untuk mengetahui elemen data mana yang bisa diakses (didapatkan otorisasinya) oleh pengguna. *DBMS* merupakan sebuah perangkat lunak yang memungkinkan pengguna untuk mendefinisikan, membuat, mengambil data dan mengontrol akses kepada *database* (Conolly, p. 16).

(a) Komponen *DBMS*

Ada lima komponen utama pada *DBMS* yaitu (Conolly, p. 18) :

(1) *Hardware* (Perangkat Keras)

DBMS dan aplikasi membutuhkan perangkat keras untuk dapat berjalan. Perangkat kerasnya dapat berupa satu *personal computer*, satu *mainframe*, maupun jaringan yang terdiri dari banyak komputer. Perangkat keras yang dibutuhkan tergantung dari permintaan organisasi dan *DBMS* yang digunakan.

(2) *Software* (Perangkat Lunak)

Komponen perangkat lunak terdiri dari perangkat *DBMS* itu sendiri dan program aplikasi bersama dengan sistem aplikasi termasuk perangkat lunak jaringan jika *DBMS* digunakan melalui jaringan.

(3) Data

Data merupakan komponen yang paling penting dari *DBMS* terutama dari sudut pandang pengguna. Data berperan sebagai jembatan antara komponen mesin (*hardware* dan *software*) dan komponen manusia (prosedur dan manusia). *Database* berisikan tentang data maupun *meta data* (data tentang data). Struktur *Database* disebut skema.

(4) Prosedur

Prosedur merujuk pada instruksi atau aturan yang mempengaruhi desain dan penggunaan dari *database*. Para pengguna sistem dan para staf yang mengatur dokumen prosedur *database* yang dibutuhkan dan bagaimana cara menggunakan atau menjalankan sistem. Hal ini mungkin mengandung instruksi yang berisikan :

- (a) Bagaimana cara memasuki *DBMS*.
- (b) Bagaimana menggunakan fasilitas *DBMS* atau program aplikasi tertentu.
- (c) Bagaimana memulai dan mengakhiri *DBMS*.
- (d) Bagaimana membuat salinan dari *database*.
- (e) Bagaimana cara mengatasi kegagalan perangkat keras dan perangkat lunak. Hal ini termasuk prosedur tentang bagaimana mengidentifikasi komponen yang gagal, cara memperbaikinya, mengikuti perbaikan dari kesalahan serta bagaimana memulihkan *database*.
- (f) Bagaimana mengubah struktur pada tabel, mengorganisasi ulang *database* yang terdapat pada lebih dari satu tempat penyimpanan.

(5) Manusia

Orang-orang yang berhubungan dengan sistem antara lain :

(a) *Database designers*

Tugasnya adalah mengidentifikasi data, relasi antar data dan batasan pada data yang akan disimpan di *database*.

(b) *Physical database designer*

Bertugas untuk memutuskan bagaimana desain *logical database* direalisasikan. Hal ini termasuk memetakan desain *logical database* kepada seperangkat *table* dan batasan integritas, memilih struktur penyimpanan yang spesifik dan metode akses data untuk mencapai performa yang baik dan mendesain aturan keamanan yang dibutuhkan oleh data.

(6) *Application Developers*

Ketika *database* diimplementasikan, program aplikasi yang menyediakan fungsi yang dibutuhkan oleh pengguna harus diimplementasikan juga. Ini adalah tanggung jawab dari *Application Developer*. Biasanya, *Application Developer* bekerja dari spesifikasi yang diproduksi oleh *system analyst*. Setiap program mengandung kalimat yang meminta *DBMS* untuk melakukan beberapa operasi pada *database*. Hal ini termasuk mengambil data, memasukkan, mengubah dan menghapus data.

(7) *End-user*

Para pengguna dapat diklasifikasikan berdasarkan bagaimana mereka menggunakan sistem yaitu :

(a) Pengguna dasar

Pengguna dasar adalah pengguna yang terlatih menggunakan *DBMS* secara awam. Mereka mengakses *database* melalui program aplikasi tertulis khusus yang diusahakan untuk membuat operasi sesederhana mungkin. Mereka menggunakan operasi *database* dengan memasukkan perintah sederhana atau memilih pilihan dari *menu*. Hal ini berarti mereka tidak perlu mendalami topik khusus mengenai *database* atau *DBMS*.

(b) Pengguna berpengalaman

Pengguna berpengalaman biasanya sudah mengenal struktur dari *database* dan fasilitas yang ditawarkan oleh *DBMS*. Pengguna berpengalaman mungkin menggunakan bahasa *query* dengan tingkat tinggi seperti SQL untuk melakukan operasi yang dibutuhkan.

(c) MySQL

MySQL merupakan salah satu sistem manajemen basis data relasional (*Relational Database Management System / RDBMS*) yang banyak digunakan dalam pengembangan perangkat lunak, baik pada skala kecil maupun besar. Sistem ini dirancang untuk mampu menangani penyimpanan serta pengelolaan data dalam jumlah besar, serta mendukung akses secara bersamaan oleh banyak pengguna dengan tingkat efisiensi yang tinggi. Raharjo, 2011, p.21) menyatakan bahwa MySQL merupakan RDBMS (atau server database) yang mengelola database dengan cepat, menampung dalam jumlah sangat besar dan dapat diakses oleh banyak user.

11. Bahasa Pemrograman

Bahasa Pemrograman adalah sekumpulan instruksi standar yang digunakan untuk memerintahkan komputer agar melakukan tugas tertentu. Bahasa ini memungkinkan kita untuk “berkomunikasi” dengan komputer dan memberitahunya apa yang harus dikerjakan, langkah demi langkah.

(a) JavaScript

JavaScript merupakan bahasa *pemrograman* yang dirancang untuk membuat halaman web menjadi interaktif dan dinamis, dengan prinsip dasar sebagai bahasa *scripting* yang mudah dipahamidan dijalankan di browser (Flanagan, 2011, p.4). Konsep JavaScript sendiri merupakan penggabungan antara fungsi-fungsi pemrograman berorientasi objek dengan kemampuan manipulasi dokumen HTML secara real-time (Crockford, 2008, p.12). Resig (2013, p.7) menjelaskan JavaScript sebagai bahasa pemrograman sisi-klien yang memungkinkan pengembang menambahkan logika, validasi serta interaksi langsung pada antarmuka pengguna. Lebih lanjut, hakekat penggunaan JavaScript terletak pada fleksibilitas dan kemampuannya dalam mendukung berbagai framework modern, sehingga mempercepat proses pengembangan aplikasi web (Duckett, 2014, p.21).

Seiring perkembangan teknologi, JavaScript tidak hanya terbatas pada sisi-klien, tetapi juga dapat dijalankan di sisi-server melalui lingkungan runtime seperti Node.js. Hal ini memperluas perannya hingga ke pengembangan aplikasi server, API, bahkan aplikasi desktop dan mobile hybrid (Tilkov & Vinoski, 2010, p. 28). Standarisasi melalui ECMAScript turut memperkaya fitur bahasa ini, seperti arrow function, async/await, class dan modular programming, yang mendukung gaya penulisan kode lebih efisien dan terstruktur (ECMA International, 2015, p.9). Selain itu, ekosistem Java Script yang luas melalui package manager (NPM), framework dan

library populer seperti React, Angular, Vue.js, serta Express.js menjadikannya fondasi utama dalam pengembangan aplikasi web berskala besar (Meijer, 2018, p.15).

(b) Hypertext Markup Language (HTML)

Hypertext Markup Language (HTML) adalah bahasa markup standar yang digunakan untuk menyusun struktur dan konten halaman web secara deklaratif. Berbeda dengan bahasa pemrograman, HTML tidak mengandung logika pemrosesan atau perintah komputasional, melainkan terdiri atas serangkaian tag yang berfungsi untuk mendefinisikan elemen-elemen dasar seperti teks, gambar, tautan, tabel, maupun formulir. Hakikat HTML adalah sebagai fondasi utama dari sebuah halaman web yang memberikan kerangka bagi konten agar dapat dikenali, diinterpretasikan dan ditampilkan secara visual oleh web browser (Duckett, 2014, p.5).

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 K-Means Clustering pada Toko Eidelweis**(oleh : Reno Supardi dkk., 2020).Penelitian ini menerapkan algoritma K-Means ($k=2$, 18 data, 2 atribut: jumlah barang dan terjual) pada data penjualan seragam aparat Toko Eidelweis. Sistem yang dikembangkan dengan Visual Basic .NET dan SQL Server ini bertujuan mengoptimalkan pemanfaatan data penjualan untuk efisiensi operasional, dengan fitur login, input, output, laporan, dan menu utama.
- (b) **Analisis Pemetaan Wilayah Menggunakan Metode Algoritma K-Means**(oleh : Nova Nur Dzuhriati dkk., 2021). Penelitian ini berhasil mengelompokkan potensi desa wisata di Kabupaten Cirebon. Jumlah *cluster* terbaik/optimal adalah 10, dengan nilai Davies Bouldin Index sebesar 0,061. *Cluster* 3 dengan 70 anggota, diidentifikasi sebagai *cluster* yang memiliki potensi wisata terbesar. Disarankan untuk pengembangan penelitian selanjutnya agar dapat menggunakan Algoritma *Machine Learning* lainnya.
- (c) **Penerapan Metode K-Means Clustering Untuk Manajemen Stok Gudang** (oleh : Catrine Wahana dkk., 2024). Penelitian ini menggunakan K-Means untuk mengelompokkan 214 produk Toko Eidelweis berdasarkan frekuensi penjualan dan perputaran barang, menghasilkan 25 produk laris (C1) dan 189 produk tidak laris (C2) dengan SSE 6281. Aplikasi web yang dikembangkan bertujuan meningkatkan

efisiensi pencatatan stok, pemantauan, dan prediksi kebutuhan, membantu staf gudang dalam pengambilan keputusan serta mengoptimalkan manajemen persediaan.

- (d) **Implementasi Algoritme *K-Means++* Untuk *Clustering* Penjualan Bahan Bangunan** (oleh : Mohammad Ferdiansyah dkk., 2024). Penelitian ini mengimplementasikan *K-Means++* untuk mengkluster data penjualan bahan bangunan. Hasil evaluasi menggunakan Davies-Bouldin Index (DBI) dan Silhouette Coefficient menunjukkan *K-Means++* lebih unggul dari *K-Means*, dengan DBI lebih rendah (pemisahan kluster lebih baik) dan Silhouette Coefficient lebih tinggi (kluster lebih padat dan terpisah).
- (e) **Studi Perbandingan Algoritma Klustering Dalam Pengelompokan Persediaan Produk (Studi Kasus : Subdirektorat Perencanaan Sarana Prasarana Dan Logistik PTN X)** (oleh : Choirun Nisa dkk., 2021). Penelitian ini membandingkan kinerja *K-Means*, *K-Medoids*, dan Agglomerative Hierarchical Clustering untuk pengelompokan optimal data persediaan produk. Menggunakan Silhouette Coefficient untuk validasi, *K-Means Clustering* terbukti paling optimal dengan nilai 0.52, menunjukkan kluster yang padat dan terpisah baik, serta kemiripan data yang tinggi dalam setiap kluster.
- (f) **Klasterisasi Dokumen Tugas Akhir Menggunakan *K-Means Clustering* sebagai Analisa Penerapan Sistem Temu Kembali** (oleh : Very Kurnia Bakti dkk., 2017). Penelitian ini mengatasi masalah pencarian dokumen TA yang tidak terkelompok temanya dengan menerapkan algoritma *K-Means* untuk klasterisasi abstrak TA berbahasa Indonesia. Hasilnya menunjukkan kluster yang baik, dibuktikan dengan nilai Davies-Bouldin Index (DBI) sebesar 0,001, yang mengindikasikan kluster padat dan terpisah sangat baik, sehingga menjadi rekomendasi untuk sistem temu kembali yang lebih efektif.
- (g) **Clustering Artikel pada Portal Berita Online Menggunakan Metode *K-Means*** (oleh : Dwiki Krisnanda Wardydkk., 2022). Penelitian ini bertujuan mengotomatisasi kategorisasi artikel berita bulanan di portal berita menggunakan *K-Means Clustering*, dikombinasikan dengan TF-IDF, seleksi fitur, dan PCA. Dengan Mean Silhouette 0.35 (menunjukkan kluster cukup baik), serta pertimbangan Accuracy, Precision, Recall, dan F1-score, implementasi ini berpotensi signifikan mengurangi beban kerja editor dan meningkatkan efisiensi manajemen konten.
- (h) **Implementasi Metode *K-Means Clustering* pada Sistem Rekomendasi Dosen Tetap Berdasarkan Penilaian Dosen** (oleh : Yessica Putri Santosodkk., 2018). Penelitian ini memanfaatkan *K-Means Clustering* pada 70 data penilaian dosen untuk merekomendasikan dosen tetap. Hasilnya, 39 dosen (55,71% tingkat

keberhasilan) teridentifikasi "layak", menunjukkan efektivitas K-Means sebagai alat bantu signifikan dalam proses pengambilan keputusan penerimaan dosen.

- (i) **Implementasi Algoritma K-Means Clustering Pada Analisis Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial Kesehatan (BPJS)** (oleh : Parasian D P Silitonga, S.Kom., M.Cs., dkk, 2016). Penelitian ini menggunakan K-Means Clustering untuk memahami pola penyebaran penyakit pada pasien BPJS di RSUP Haji Adam Malik Medan. Hasilnya mengidentifikasi tiga kluster dominan: Kluster 1 (24 pasien) dengan penyakit gigi, Kluster 2 (28 pasien) dengan penyakit saluran pernapasan atas, dan Kluster 3 (23 pasien) dengan diare, menunjukkan keberhasilan algoritma dalam mengidentifikasi kecenderungan penyakit signifikan.
- (j) **Segmentasi Pelanggan Penjualan Online Menggunakan Metode K-means Clustering** (oleh : Candra Hafidz Ardana dkk, 2024). Penelitian ini menggunakan K-Means Clustering untuk segmentasi pelanggan penjualan online, dengan tujuan memahami preferensi dan perilaku pelanggan demi strategi pemasaran yang efektif. Meskipun nilai Silhouette Score tidak disebutkan, hasil evaluasi menunjukkan K-Means efektif mengelompokkan data pelanggan, esensial untuk pengambilan keputusan bisnis yang lebih baik di industri ini.

Tabel 2 . 14 Tinjauan Pustaka

No	Peneliti	Judul	Jurnal Sumber	Kontribusi
1.	Reno Supardi, Indra Kanedi (2020)	Implementasi Metode Algoritma K-Means Clustering Pada Toko Eidelweis	(Jurnal Teknologi Informasi) Vol.4, No.2, Desember 2020 Link : https://www.neliti.com/publications/495510/implementasi-metode-algoritma-k-means-clustering-pada-toko-eidelweis	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada Teknik perhitungan manual menggunakan K-Means Clustering

No	Peneliti	Judul	Jurnal Sumber	Kontribusi
2.	Novia Nur Dzuhriati, Lugi Prayoto, Bagas Tanjung Mardiyanto, Heliyanti Susana (2021)	Analisis Pemetaan Wilayah Menggunakan Metode <i>Algoritma K-Means</i>	Jurnal Ilmiah Betrik, Volume.12, No.03, Desember 2021 Link : https://media.neliti.com/media/publications/460086-none-05336f9b.pdf	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada penyediaan kerangka kerja metodologis dan validasi implementasi <i>Algoritma K-Means</i> untuk pemetaan wilayah berbasis data
3.	Catrine Wahana ¹ , Desiyanna Lasut ² (2024)	Penerapan Metode K-Means Clustering Untuk Manajemen Stok Gudang	Akselerator : Jurnal Sains Terapan dan Teknologi, Vol. 5 No. 2 (2024): Engineering for Data Mining Link : https://jurnal.ubd.ac.id/index.php/aksel/article/view/2385	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada Tampilan Data Clustering di Perancangan Aplikasinya
4.	Mohammad Ferdiansyah, Umi Chotijah (2024)	Implementasi Algoritme <i>K-Means++</i> Untuk <i>Clustering</i> Penjualan Bahan Bangunan	Jurnal Ilmiah Teknik Informatika dan Komunikasi Vol. 4 No. 1 Maret 2024 E-ISSN: 2827-7945 / P-ISSN: 2827-8127, Hal 181-193 Link : http://eprints.umg.ac.id/9617/	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada Pemilihan centroid/pusat awal Cluster

No	Peneliti	Judul	Jurnal Sumber	Kontribusi
5.	Choirun Nisa, Wiyli Yustanti (2021)	Studi Perbandingan Algoritma Klustering Dalam Pengelompokan Persediaan Produk (Studi Kasus : Subdirektorat Perencanaan Sarana Prasarana Dan Logistik PTN X)	<i>JEISBI: Volume 02 Number 03, 2021 ISSN 2774-3993 (Journal of Emerging Information Systems and Business Intelligence)</i> Link : https://ejournal.unesa.ac.id/index.php/JEISBI/article/view/41103	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada teknik penulisan Hasil dan Pembahasan
6.	Very Kurnia Bakti, Jatmiko Indriyatno (2017)	Klasterisasi Dokumen Tugas Akhir Menggunakan <i>K-Means Clustering</i> sebagai Analisa Penerapan Sistem Temu Kembali	KOPERTIP: Jurnal Ilmiah Manajemen Informatika dan Komputer ISSN: 2549-211X Vol. 01, No. 01, Februari 2017, pp. 31-34 Link : https://www.neliti.com/publications/227019/klasterisasi-dokumen-tugas-akhir-menggunakan-k-means-clustering-sebagai-analisa	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada teknik Pengumpulan Data Penelitian
7.	Dwiki Krisnanda Wardy, I Ketut Gede Darma Putra, Ni Kadek Dwi Rusjyanthi (2022)	Clustering Artikel pada Portal Berita Online Menggunakan Metode K- Means	JITTER- Jurnal Ilmiah Teknologi dan Komputer Vol. 3, No. 1 April 2022 Link : https://www.neliti.com/publications/432889/clustering-artikel-pada-portal-berita-online-menggunakan-metode-k-means	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada Proses Penilaian nilai Cluster menggunakan visualisasi berupa <i>bar plot</i>

No	Peneliti	Judul	Jurnal Sumber	Kontribusi
8.	Yessica Putri Santoso, Marlina, dan Halim Agung (2018)	Implementasi Metode <i>K-Means Clustering</i> pada Sistem Rekomendasi Dosen Tetap Berdasarkan Penilaian Dosen	Jurnal Informatika Universitas Pamulang ISSN: 2541-1004 Vol. 3, No. 4, Desember 2018 e-ISSN: 2622-4615 Link : https://openjournal.unpam.ac.id/index.php/informatika/article/view/2133	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada teknik Analisis dan Pembahasan
9.	Parasian D P Silitonga, S.Kom., M.Cs, Romanus Damanik S.Kom., M.Kom (2016)	Implementasi Algoritma K-Means Clustering Pada Analisis Penyebaran Penyakit Pasien Pengguna Badan Penyelenggara Jaminan Sosial Kesehatan (Bpjs) (Studi Kasus : Rumah Sakit Umum Pusat Haji Adam Malik Medan)	<i>Jurnal Teknik Informatika Unika St. Thomas (JTIUST), Volume 01 Nomor 02, Desember 2016, 2548-1916</i> Link : https://ejournal.ust.ac.id/index.php/JTIUST/article/view/34	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada teknik Penulisan Analisis Masalah

No	Peneliti	Judul	Jurnal Sumber	Kontribusi
10.	Candra Hafidz Ardana, Adlian Aldita Alif Aisyah Ainur Khoyum, Muhammad Faisal (2024)	Segmentasi Pelanggan Penjualan <i>Online</i> Menggunakan Metode <i>K means Clustering</i>	JISKA (Jurnal Informatika Sunan Kalijaga) Vol. 9, No. 1, JANUARI, 2024, Pp. 1 – 9 ISSN: 2527 – 5836 (print) 2528 – 0074 (online) Link : https://ejournal.uin-suka.ac.id/saintek/JISKA/article/view/4039	Kontribusi jurnal tersebut dalam Penelitian ini adalah pada teknik Penentuan Jumlah Cluster

Berdasarkan tinjauan pustaka dari 10 (sepuluh) jurnal yang sudah dipaparkan pada tabel 2.14, maka diambil 1 (satu) jurnal yang menjadi rujukan penelitian sebelumnya dengan judul “ANALISIS PEMETAAN WILAYAH MENGGUNAKAN METODE ALGORITMA *K-MEANS*”. Didapatkan kesamaan penerapan metode *K-means*, sementara sisi Originalitas dari penelitian yang akan dilakukan terletak pada objek *clustering* yang berbeda, dimana pada penelitian sebelumnya berfokus pada pengelompokkan desa dan kelurahan berdasarkan potensi wisata. Sementara dari sisi kontribusi baru penelitian ini adalah terletak pada penggunaan metode *K-Means* untuk memetakan wilayah berdasarkan data akta kelahiran sebagai bentuk analisis dan peningkatan pelayanan publik di bidang kependudukan. Data Akta Kelahiran yang digunakan bersifat administratif dan demografis, yaitu jumlah pencatatan kelahiran pada masing-masing wilayah. Dengan demikian, data akta kelahiran berorientasi pada pemenuhan hak sipil dan pelayanan dasar masyarakat sehingga masyarakat semua memiliki akta kelahiran, sekaligus terciptanya pelayanan yang lebih merata dan tepat sasaran.

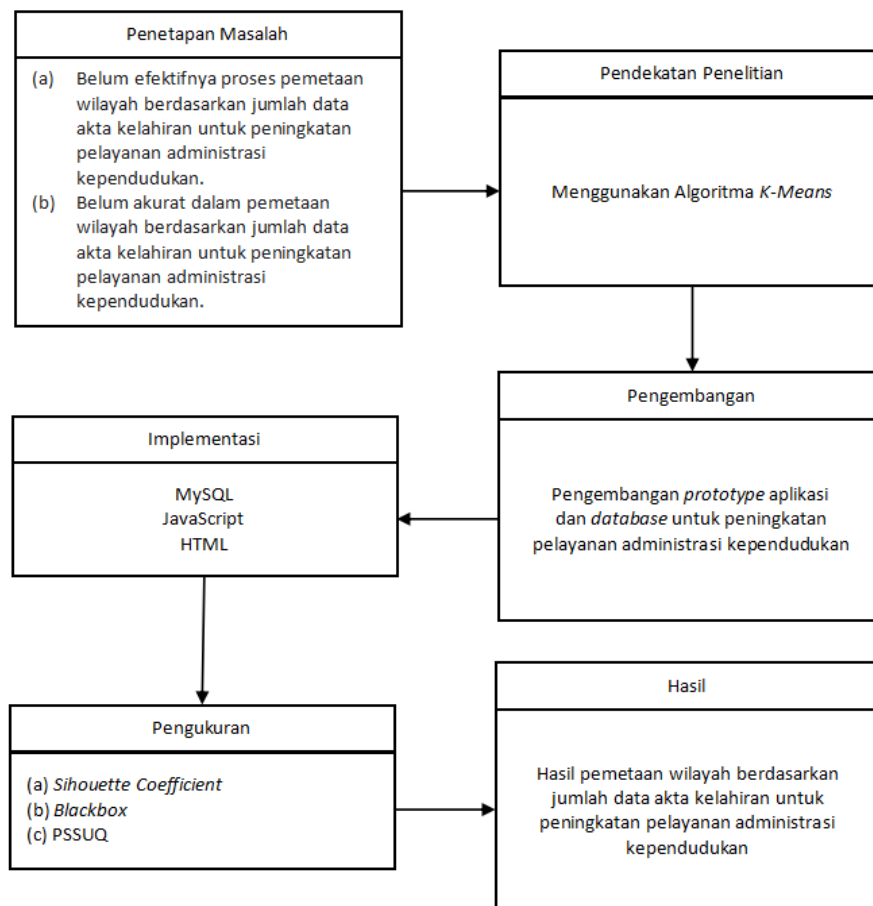
C. Kerangka Berfikir

Berikut Gambar 2.5 berisi tentang 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.

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 efektifnya proses pemetaan wilayah berdasarkan jumlah data akta kelahiran untuk peningkatan pelayanan administrasi kependudukan.

- (b) Belum akurat dalam pemetaan wilayah berdasarkan jumlah data akta kelahiran untuk peningkatan pelayanan administrasi kependudukan.
- (2) Pada penelitian ini pendekatan yang dilakukan yaitu menggunakan pendekatan Algoritma *K-Means*.
- (3) Penelitian ini menggunakan data berupa data excel.
- (4) Pengembangan yang dilakukan yaitu menggunakan MySQL, JavaScript dan HTML.
- (5) Setelah itu dilakukan pengukuran dengan menggunakan algoritma *Silhouette Coefficient* sebagai uji keakuratan hasil penerapan Algoritma *K-Means*, pengujian *Blackbox* sebagai uji sistem dan PSSUQ untuk uji kebergunaan.
- (6) Hasil dari penelitian ini adalah pemetaan wilayah berdasarkan jumlah data akta kelahiran untuk peningkatan pelayanan administrasi kependudukan dan *prototype* aplikasi penerapan Algoritma *K-Means*.



Gambar 2 . 5 Kerangka Berfikir

D. Hipotesis Penelitian

Telah dilakukan penelitian oleh Novia Nur Dzuhriati dkk., (2021), yang berjudul “ANALISIS PEMETAAN WILAYAH MENGGUNAKAN METODE ALGORITMA *K-MEANS*” yang dipublikasikan dalam Jurnal Ilmiah Betrik, Volume. 12, No. 03, pada Desember 2021. Penelitian tersebut menunjukkan bahwa metode *K-Means Clustering* mampu menghasilkan pengelompokan wilayah berdasarkan tingkat potensi wisata yang memiliki karakteristik berbeda-beda dan metode *K-Means Clustering* mampu memberikan hasil pemetaan yang optimal dalam mengidentifikasi wilayah yang memiliki potensi wisata untuk dikembangkan.

Berdasarkan penelitian tersebut, maka hipotesis dari penelitian ini adalah metode *K-Means Clustering* diduga mampu menghasilkan cluster wilayah yang optimal berdasarkan kedekatan nilai jumlah data akta kelahiran dengan pusat *cluster (centroid)* sehingga hasil pengelompokan dapat digunakan untuk mendukung pengambilan keputusan dalam peningkatan pelayanan administrasi kependudukan.