

## **BAB II KERANGKA TEORITIS**

### **A. Landasan Teori**

#### **1. Data Mining**

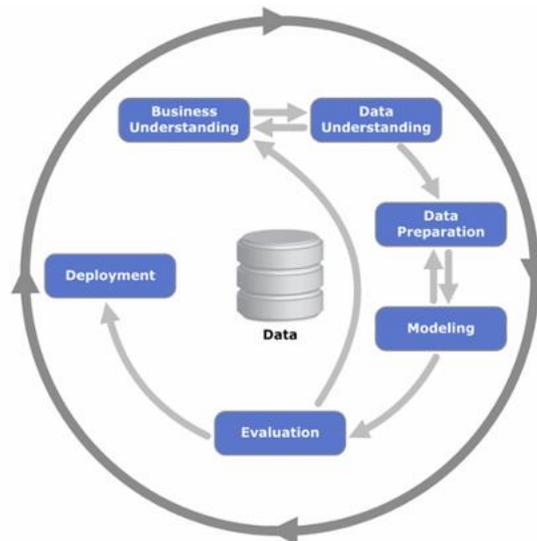
Menurut (Wahyudi & rekan, 2020, p. 2) Data mining mewarisi banyak aspek dan teknik dari bidang-bidang ilmu yang sudah mapan terlebih dahulu. (Much Aziz Muslim, 2019, p. 2) mengutarakan bahwa Data mining, secara sederhana merupakan suatu langkah ekstraksi untuk mendapatkan informasi penting yang sifatnya implisit dan belum diketahui. Data mining, yang sering disebut Knowledge Discovery in Database (KDD), adalah proses mengumpulkan, menggunakan, dan menganalisis data historis untuk menemukan hubungan, pola, atau keteraturan dalam kumpulan data yang sangat besar (Wahyudi & rekan, 2020, p. 1).

Tahapan data mining menurut (Wahyudi & rekan, 2020, p. 3)

- (1) Pembersihan data (data cleaning) Pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan;
- (2) Integrasi data (data integration) Integrasi data merupakan penggabungan data dari berbagai database ke dalam satu database baru;
- (3) Seleksi data (data selection) Data yang ada pada database sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari database;
- (4) Transformasi data (data transformation) Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam data mining;
- (5) Proses mining Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data;
- (6) Evaluasi pola (pattern evaluation) Untuk mengidentifikasi pola-pola menarik ke dalam knowledge base yang ditemukan;
- (7) Presentasi pengetahuan (knowledge presentation) Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan yang diperoleh pengguna.

## 2. Crisp DM

(Cross-Industry Standard Process for Data Mining) CRISP-DM menurut (Sulianta, 2024, p. 35) adalah sebuah kerangka kerja atau metodologi yang digunakan dalam data mining dan analisis data. CRISP-DM dapat disesuaikan dengan berbagai proyek data mining dan merupakan pendekatan yang fleksibel (Sulianta, 2024, p. 37). Menurut (Osei-Bryson & Barclay, 2015, p. 19) Model CRISP-DM dapat menggambarkan berbagai putaran umpan balik agar menekankan bagaimana fase-fase tertentu harus ditinjau kembali untuk memanfaatkan informasi atau pengetahuan baru yang diperoleh pada fase berikutnya.



Gambar 2. 1 Siklus Hidup Data Mining

Sumber: (Sulianta, 2024)

Menurut (Sulianta, 2024, p. 35) CRISP-DM terdiri dari enam tahap utama yang dijelaskan sebagai berikut :

- (1) Pemahaman Bisnis (Business Understanding) Tahap ini melibatkan pemahaman mendalam tentang masalah yang ingin dipecahkan dengan data mining, Ini melibatkan interaksi dengan pemangku kepentingan untuk mendefinisikan tujuan bisnis, target, dan kendala yang relevan;
- (2) Pemahaman Data (Data Understanding) Pada tahap ini, harus mengumpulkan dan memahami data yang relevan untuk proyek data mining, didalamnya

mencakup eksploitasi sumber data, eksplorasi data awal, dan pemahaman karakteristik data;

- (3) Persiapan data (Data Preparation) Tahap ini melibatkan pembersihan, transformasi, dan pemodelan data yang akan digunakan dalam proses data mining, proses ini termasuk penghapusan data yang tidak relevan, penggabungan sumber data, dan pelabelan data;
- (4) Modeling di tahap ini, teknik data mining yang sesuai dan mengembangkan model analisis data, ini dapat melibatkan teknik seperti regresi, klasifikasi, klasterisasi, atau asosiasi, tergantung pada tujuan proyek;
- (5) Evaluasi (Evaluation) Setelah mengembangkan model, harus menguji dan mengevaluasinya untuk memastikan bahwa hal tersebut memenuhi tujuan bisnis dan memiliki kinerja yang baik, evaluasi melibatkan penggunaan teknik statistika dan teknik validasi untuk mengukur kualitas model;
- (6) Implementasi (Deployment) Tahap akhir melibatkan mengimplementasikan model ke dalam lingkungan yang sebenarnya sebagai bagian dari aktivitas organisasi atau perusahaan, ini dapat mencakup pengkodean model, pengintegrasian ke dalam sistem, dan pelatihan personel yang terlibat.

### **3. Algoritma K-Means**

Algoritma K-Means merupakan metode non-hierarki yang pada awalnya mengambil sebagian banyaknya komponen populasi untuk dijadikan pusat kluster awal, pada tahap ini pusat kluster dipilih secara acak dari sekumpulan populasi data, berikutnya, K-Means menguji setiap komponen dalam populasi data dan menandainya ke salah satu pusat kluster yang telah ditetapkan berdasarkan jarak minimal antar komponen, posisi pusat kluster akan dihitung kembali sampai semua komponen data digolongkan ke dalam tiap pusat kluster, dan kemudian posisi pusat kluster baru akan dibuat (Wahyudi & rekan, 2020, p. 6). Pengelompokan K-means merupakan metode populer untuk analisis data jika tujuannya adalah untuk mengidentifikasi struktur yang mendasari dalam kumpulan data (Permana & rekan, 2023, p. 62). (Permana & rekan, 2023) mengutarakan bahwa Algoritma k-means dinilai cukup efisien, dengan ketentuan k adalah jumlah cluster yang terbentuk, n adalah jumlah objek data, dan t adalah jumlah iterasi.

Berikut ini adalah contoh kasus teknik klastering K-Means, (Vulandari, 2017, pp. 54-62). Untuk menghitung nilai BMI (Body Mass Index) atau ukuran tubuh dan ukuran kerangka dengan menggunakan persamaan (1) dan (2). Hasil dari perhitungan dapat dilihat pada tabel dibawah ini:

Mhs	BMI	Ukuran Kerangka
1	22,21	11,64
2	43,25	8,95
3	19,71	10,93
4	21,05	10,38
5	17,93	12,85
6	17,72	12
7	18,71	11,53
8	25,86	9,33
9	19,15	11,8
10	18,42	11,2
11	22,94	10,6
12	26,89	10,44
13	24,91	10,63
14	22,99	11,47
15	26,81	9,17
16	19,14	12,11
17	21,09	10,67
18	18,71	12,36
19	20,58	10,8
20	27,66	9,94

Digunakan algoritma klasifikasi K-Means untuk mengelompokan data yang ada. Data yang ada akan dikelompokan menjadi 3 kelompok. Adapun langkah dari pengelompokan data adalah sebagai berikut :

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

(1) Tentukan pusat kluster secara acak, misalkan kita tentukan c1 = (20,9); c2 = (23,10); dan c3 = (27,11)

(2) Hitung jarak setiap data yang ada terhadap setiap pusat cluster. Misalkan untuk menghitung jarak data mahasiswa pertama dengan pusat cluster pertama adalah :

$$d_{11} = \sqrt{22,21 - 20)^2 + (11,64 - 9)^2} = 4,97$$

Jarak data mahasiswa pertama dengan pusat cluster kedua:

$$d_{12} = \sqrt{22,21 - 23)^2 + (11,64 - 10)^2} = 2,04$$

Jarak data mahasiswa pertama dengan pusat cluster ketiga:

$$d_{13} = \sqrt{22,21 - 27)^2 + (11,64 - 11)^2} = 1,99$$

Mhs	BMI	UK	C1	C2	C3
1	22,21	11,64	4,97	2,04	1,91
2	43,25	8,95	25,25	22,28	19,36
3	19,71	10,93	2,58	1,6	4,29
4	21,05	10,38	3,34	0,38	3,02
5	17,93	12,85	3,85	4,19	6,35
6	17,72	12	3,01	3,85	6,36
7	18,71	11,53	2,63	2,76	5,32
8	25,86	9,33	7,87	4,91	2,5
9	19,15	11,8	3,03	2,58	4,91
10	18,42	11,2	2,24	2,84	5,58
11	22,94	10,6	5,19	2,03	1,13
12	26,89	10,44	9,01	5,91	2,95
13	24,91	10,63	7,1	3,96	0,99
14	22,99	11,47	5,56	2,47	1,12
15	26,81	9,17	8,82	5,87	3,36
16	19,14	12,11	3,31	2,81	4,98
17	21,09	10,67	3,51	0,67	2,93
18	18,71	12,36	3,43	3,29	5,46
19	20,58	10,8	3,14	0,91	3,43
20	27,66	9,94	9,71	6,66	3,81

(3) Suatu data akan menjadi anggota dari suatu cluster yang memiliki jarak terkecil dari pusat clusternya. Data pertama, jarak terkecil diperoleh pada cluster ketiga, sehingga data pertama akan menjadi anggota dari cluster ketiga. Posisi cluster selengkapnya dapat dilihat pada tabel sebagai berikut:

Mhs	BMI	UK	C1	C2	C3
1	22,21	11,64			*
2	43,25	8,95			*
3	19,71	10,93		*	
4	21,05	10,38		*	
5	17,93	12,85	*		
6	17,72	12	*		
7	18,71	11,53	*		
8	25,86	9,33			*
9	19,15	11,8		*	
10	18,42	11,2	*		
11	22,94	10,6			*
12	26,89	10,44			*
13	24,91	10,63			*
14	22,99	11,47			*
15	26,81	9,17			*
16	19,14	12,11		*	
17	21,09	10,67		*	
18	18,71	12,36		*	
19	20,58	10,8		*	
20	27,66	9,94			*

(4) Hitung pusat cluster baru. Untuk cluster pertama, ada 4 data yaitu data ke-5, 6, 7 dan 10, sehingga:

$$C_{11} = \frac{17,93 + 17,72 + 18,71 + 18,42}{4} = 18,19$$

$$C_{12} = \frac{12,85 + 12,00 + 11,53 + 11,20}{4} = 11,89$$

Untuk cluster kedua, ada 7 data yaitu data ke-3, 4, 9, 16, 17, 18 dan 19:

$$C_{21} = \frac{(19,71 + 21,05 + 19,15 + 19,14 + 21,09 + 18,71 + 20,58)}{7} = 19,92$$

$$C_{22} = \frac{(10,93 + 10,38 + 11,80 + 12,11 + 10,67 + 12,36 + 10,8)}{7} = 11,29$$

Untuk cluster ketiga, ada 9 data yaitu data ke-1, 2, 8, 11, 12, 13, 14, 15, 20:

$$C_{31} = \frac{(22,21 + 43,25 + 25,86 + 22,94 + 26,89 + 24,91 + 22,99 + 26,81 + 27,66)}{9} = 27,06$$

$$C_{32} = \frac{(11,64 + 8,95 + 9,33 + 10,6 + 10,44 + 10,63 + 11,47 + 9,17 + 9,94)}{9} = 10,24$$

- (5) Ulangi langkah 2 hingga posisi data sudah tidak mengalami perubahan  
Berikut adalah tabel posisi cluster pada iterasi ke-2

Mhs	BMI	UK	C1	C2	C3
1	22,21	11,64		*	
2	43,25	8,95			*
3	19,71	10,93		*	
4	21,05	10,38		*	
5	17,93	12,85	*		
6	17,72	12	*		
7	18,71	11,53	*		
8	25,86	9,33			*
9	19,15	11,8		*	
10	18,42	11,2	*		
11	22,94	10,6		*	
12	26,89	10,44			*
13	24,91	10,63			*
14	22,99	11,47		*	
15	26,81	9,17			*
16	19,14	12,11	*		
17	21,09	10,67		*	
18	18,71	12,36	*		
19	20,58	10,8		*	

Mhs	BMI	UK	C1	C2	C3
20	27,66	9,94			*

Berikut adalah tabel posisi cluster pada iterasi ke-3

Mhs	BMI	UK	C1	C2	C3
1	22,21	11,64		*	
2	43,25	8,95			*
3	19,71	10,93		*	
4	21,05	10,38		*	
5	17,93	12,85	*		
6	17,72	12	*		
7	18,71	11,53		*	
8	25,86	9,33			*
9	19,15	11,8	*		
10	18,42	11,2	*		
11	22,94	10,6		*	
12	26,89	10,44			*
13	24,91	10,63		*	
14	22,99	11,47		*	
15	26,81	9,17			*
16	19,14	12,11	*		
17	21,09	10,67		*	
18	18,71	12,36	*		
19	20,58	10,8		*	
20	27,66	9,94			*

Berikut adalah tabel posisi cluster pada iterasi ke-8

Mhs	BMI	UK	C1	C2	C3
1	22,21	11,64	*		
2	43,25	8,95			*
3	19,71	10,93	*		
4	21,05	10,38	*		

Mhs	BMI	UK	C1	C2	C3
5	17,93	12,85	*		
6	17,72	12	*		
7	18,71	11,53	*		
8	25,86	9,33		*	
9	19,15	11,8	*		
10	18,42	11,2	*		
11	22,94	10,6		*	
12	26,89	10,44		*	
13	24,91	10,63		*	
14	22,99	11,47		*	
15	26,81	9,17		*	
16	19,14	12,11	*		
17	21,09	10,67	*		
18	18,71	12,36	*		
19	20,58	10,8	*		
20	27,66	9,94		*	

Berikut adalah tabel posisi cluster pada iterasi ke-9

Mhs	BMI	UK	C1	C2	C3
1	22,21	11,64	*		
2	43,25	8,95			*
3	19,71	10,93	*		
4	21,05	10,38	*		
5	17,93	12,85	*		
6	17,72	12	*		
7	18,71	11,53	*		
8	25,86	9,33		*	
9	19,15	11,8	*		
10	18,42	11,2	*		
11	22,94	10,6		*	
12	26,89	10,44		*	

Mhs	BMI	UK	C1	C2	C3
13	24,91	10,63		*	
14	22,99	11,47		*	
15	26,81	9,17		*	
16	19,14	12,11	*		
17	21,09	10,67	*		
18	18,71	12,36	*		
19	20,58	10,8	*		
20	27,66	9,94		*	

Karena pada iterasi ke-8 dan ke-9 posisi cluster tidak berubah, maka iterasi dihentikan dan hasil akhir yang diperoleh adalah 3 cluster:

- (a) Cluster pertama memiliki pusat (19,53; 11,52) yang dapat diartikan sebagai kelompok mahasiswa dengan BMI normal dan kerangka besar.
- (b) Cluster kedua memiliki pusat (25,44; 10,22) yang dapat diartikan sebagai kelompok mahasiswa dengan BMI obesitas sedang dan kerangka sedang.
- (c) Cluster ketiga memiliki pusat (43,25; 8,95) yang dapat diartikan sebagai kelompok mahasiswa dengan BMI obesitas berat dan kerangka kecil.

#### 4. Clustering

Clustering atau klasterisasi adalah metode pengelompokan data Menurut (Tan, 2006, p. 125) clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa cluster atau kelompok sehingga data dalam satu cluster memiliki tingkat kemiripan yang maksimum dan data antar cluster memiliki kemiripan yang minimum. Clustering atau pengelompokan merupakan suatu teknik yang dapat dipakai untuk mengidentifikasi kelompok yang dihasilkan dari pengelompokan item yang lebih kecil berdasarkan kesamaannya (Permana & rekan, 2023, p. 27). Clustering merupakan teknik dengan cara mengelompokkan data secara otomatis tanpa diberitahukan label kelasnya (Much Aziz Muslim, 2019, p. 45).

(Permana & rekan, 2023, p. 61) mengutarakan bahwa Clustering merupakan salah satu teknik dasar untuk mendapatkan wawasan tentang sifat dasar dan struktur data dengan tujuan mengorganisasikan sekumpulan data ke dalam cluster-cluster yang elemen-elemennya mirip satu sama lain dan berbeda dengan yang ada di cluster lain. Menurut (LAROSE, 2005, p. 16) Clustering berbeda dengan klasifikasi karena tidak

ada variabel target untuk clustering. Clustering sering kali dilakukan sebagai langkah awal dalam proses penambangan data, dan cluster yang dihasilkan digunakan sebagai masukan lebih lanjut ke dalam teknik hilir yang berbeda (LAROSE, 2005, p. 17).

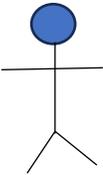
## 5. Unified Modelling Language (UML)

UML (Unified Modeling Language) merupakan alat bantu yang sudah menjadi standar dalam dunia pengembangan sistem perangkat lunak berorientasi objek (Anardani, 2019, p. 13). UML dapat digunakan sebagai pemodelan untuk semua jenis aplikasi perangkat lunak serta dalam bahasa pemrograman apapun (Anardani, 2019, p. 14). (Hasanah & Untari, 2020, p. 64) mengutarakan bahwa UML muncul karena adanya kebutuhan pemodelan visual untuk menspesifikasikan, menggambarkan, membangun, dan mendokumentasikan sistem perangkat lunak.

UML memiliki diagram-diagram dalam pembuatan aplikasi berorientasi objek, yaitu (Rosa & M. Shalahuddin, 2016):

### (a) Use Case Diagram

Digunakan untuk mengetahui fungsi apa saja yang ada di dalam sebuah sistem informasi dan siapa saja yang berhak menggunakan fungsi-fungsi tersebut;

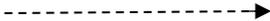
Notasi	Keterangan	Simbol
Actor	Pengguna sistem, misalnya manusia, aplikasi atau objek lain	

Notasi	Keterangan	Simbol
Use Case	Dinyatakan dengan menggunakan kata kerja di tengah lingkaran elips	
Asosiasi / Association	Sebuah garis yang berfungsi menghubungkan actor dengan use case	

(b) Class Diagram

Diagram kelas atau class diagram menggambarkan struktur sistem dari segi pendefinisian kelas-kelas yang akan dibuat untuk membangun sistem;

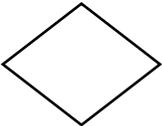
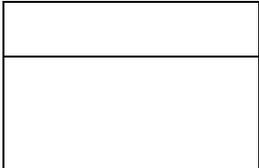
Notasi	Keterangan	Simbol			
Class	Menggambarkan class pada struktur system	<table border="1" data-bbox="1112 1207 1364 1344"> <tr> <td>Nama_kelas</td> </tr> <tr> <td>+atribut</td> </tr> <tr> <td>+operasi()</td> </tr> </table>	Nama_kelas	+atribut	+operasi()
Nama_kelas					
+atribut					
+operasi()					
Association	Relasi antar kelas dengan makna umum				

Notasi	Keterangan	Simbol
Dependency	Ketergantungan antar satu class dengan class yang lain	
Generalization	Relasi antar class ke beberapa class	

(c) Activity Diagram

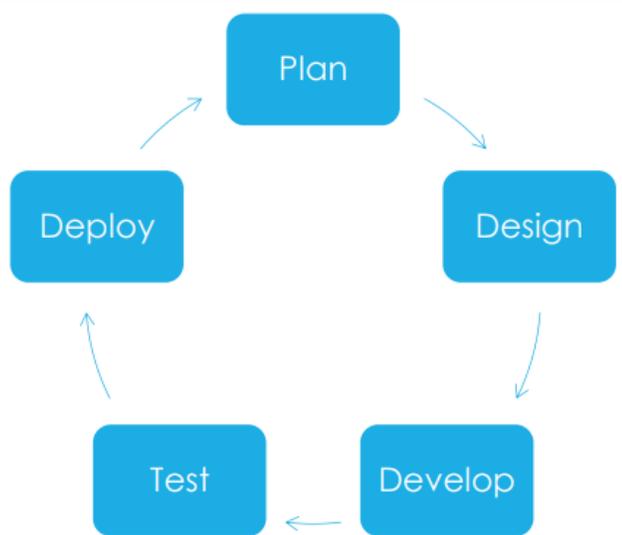
Diagram aktivitas atau activity diagram menggambarkan workflow (aliran kerja) atau aktivitas dari sebuah sistem atau proses bisnis yang ada pada perangkat lunak;

Notasi	Keterangan	Simbol
Initial state	Titik awal untuk suatu diagram aktivitas	
Final state	Titik akhir yang dilakukan oleh sistem	
Activity	Menggambarkan sebuah aktivitas	

Notasi	Keterangan	Simbol
Decision	Pilihan untuk pengambilan keputusan	
Swimlane	Memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi	

## 6. System Development Life Cycle (SDLC)

(Hartono, 2021, p. 62) Mengutarakan bahwa Model SDLC atau Sekuensial Linier ini mengusulkan sebuah pendekatan perkembangan perangkat lunak yang sistematis dan sekuensial yang diawali pada tingkat dan kemajuan sistem pada seluruh analisis, desain, kode, pengujian, dan pemeliharaan. Menurut (Tiky, 2016, p. 7) Siklus Hidup Pengembangan Perangkat Lunak terdiri dari langkah-langkah dan aktivitas terperinci yang menjelaskan bagaimana merancang, mengembangkan, memelihara, mengganti, mengubah, menyempurnakan, menguji, atau bahkan meluncurkan perangkat lunak sehingga Kegiatan-kegiatan tersebut dapat diuraikan menjadi tingkat yang sangat rinci tetapi pada saat yang sama dapat dikelompokkan ke dalam lima (5) kategori inti: Merencanakan, Merancang, Mengembangkan, Menguji dan Menerapkan.



*Gambar 2. 2 SDLC Konsep*

*Sumber: (Hasanah & Untari, 2020)*

Dalam rekayasa perangkat lunak, konsep SDLC mendasari berbagai jenis metodologi pengembangan perangkat lunak dan Metodologi-metodologi ini membentuk suatu kerangka kerja untuk perencanaan dan pengendalian pembuatan sistem informasi, yaitu proses pengembangan perangkat lunak, terdapat beberapa model SDLC yang dapat digunakan, semuanya mempunyai kelebihan dan kekurangan masing-masing pada tiap tahapannya, hal yang paling penting yaitu mengenali type pelanggan/ customer dan memilih menggunakan model SDLC yang sesuai dengan karakter pelanggan dan sesuai dengan karakter pengembang (Hasanah & Untari, 2020, p. 20).

## **7. Python**

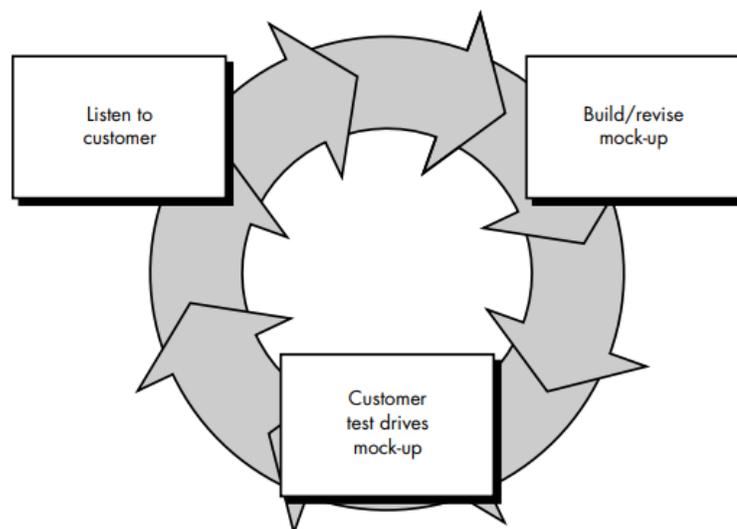
Python merupakan bahasa pemrograman dinamis, tingkat tinggi, dimana merupakan bahasa pemrograman interpreter yaitu bahasa yang mengubah source code menjadi machine code secara langsung ketika program dijalankan (Suharto, 2023, p. 1). Menurut (Ma'arif, 2020, p. 1) Python dapat digunakan untuk Pengembangan web (sisi server), pengembangan perangkat lunak atau membuat aplikasi (software), menyelesaikan persamaan Matematika, pembuatan skrip sistem dan pemrograman mikrokontroler (MicroPython). Python mendukung pemrograman

berorientasi objek, pendekatan pemrograman prosedural dan menyediakan alokasi memori dinamis (Suharto, 2023, p. 10).

Menurut (Suharto, 2023, p. 20) Bahasa pemrograman python merupakan bahasa dengan pertumbuhan tercepat dan dapat mengembangkan aplikasi apa pun. Variabel Python dikenal sebagai pengidentifikasi dan digunakan untuk menyimpan nilai (Suharto, 2023, p. 34). (Suharto, 2023, p. 43) mengutarakan bahwa Python merupakan bahasa yang diketik secara dinamis; sehingga tidak perlu mendefinisikan tipe variabel saat mendeklarasikannya.

## 8. Model Prototype

Pembuatan prototype dimulai dengan mengumpulkan persyaratan. (Pressman, 2002, p. 30)Pengembang dan klien bertemu untuk menetapkan tujuan umum, mengidentifikasi gambaran kebutuhan dari segi input dan format output serta gambaran interface, kemudian dilakukan perancangan cepat. Hasil dari perancangan cepat tersebut nantinya akan diuji dan dievaluasi. Penjelasan lengkap pada metode prototype dapat digambarkan sebagai berikut:



*Gambar 2. 3 Model Prototype*

Sumber: (Pressman, 2002)

Tahapan – tahapan model prototype (Pressman, 2002, p. 31);

- (1) Listen to customer; Dalam tahapan ini dilakukan pengumpulan kebutuhan sistem yang akan dibuat dengan cara mendengar keluhan dari pelanggan dan memastikan bahwa sistem yang dibuat memenuhi kebutuhan, maka perlu mengetahui terlebih dahulu bagaimana sistem yang sedang berjalan untuk kemudian dapat mengidentifikasi masalah yang terjadi;
- (2) Build/revise mock-up; Pada tahapan ini, prototype sistem dibuat dan dirancang dengan menyesuaikan kebutuhan sistem yang telah ditentukan pada tahapan sebelumnya;
- (3) Pada tahap ini, prototype yang dibuat, diuji coba oleh pengguna, prototype tersebut kemudian dievaluasi dan dilakukan perbaikan jika masih ada kekurangan atau belum memenuhi kebutuhan pengguna dan mendengarkan keluhan dari pelanggan untuk memperbaiki prototype yang ada.

## **9. Pengertian Data**

Menurut (Priadana & Sunarsi, 2021, p. 20) Data sebagai informasi faktual (misalnya pengukuran atau statistik) yang dipergunakan sebagai dasar untuk penalaran, diskusi, atau perhitungan, misalnya dalam penelitian ilmiah dan data sebagai kenyataan-kenyataan murni yang belum diberi penafsiran apapun, belum diubah, atau belum dimanipulasi, namun telah tersusun dalam sistematika statistika tertentu, sistematika tersebut bisa mengikuti dasar kronologis (waktu), spasial (tempat), peristiwa, pokok soal, atau dasar lainnya yang sesuai dengan tujuan penelitian. Menurut (Pressman, 2002, p. 2) Data yang diperoleh melalui penelitian itu adalah data empiris (teramati) yang mempunyai kriteria tertentu yaitu valid. Data yang telah diperoleh dari penelitian dapat digunakan untuk memahami, memecahkan dan mengantisipasi masalah (Sugiyono, 2013, p. 3).

## **10. Sistem Pendukung Keputusan (SPK)**

Sistem Pendukung Keputusan (SPK), Menurut (Mashuri & Mujianto, 2021, p. 5) dapat diartikan atau didefinisikan sebagai sebuah sistem yang dapat dan mampu memberikan solusi atau kemampuan baik kemampuan pemberian solusi atau pemecahan masalah maupun kemampuan mengkomunikasikan terhadap masalah masalah semi-terstruktur, dengan kata lain secara khusus, SPK dideskripsikan atau

dijelaskan sebagai sebuah sistem yang dapat mensupport kerja seorang pengambil keputusan dalam memecahkan / memberikan solusi terhadap masalah yang bersifat semiterstruktur melalui cara memberikan informasi ataupun saran menuju pada keputusan tertentu. Sistem Pendukung Keputusan merupakan sistem yang membantu pengambilan keputusan dengan melengkapi informasi dari data yang telah diolah dengan benar dan dibutuhkan untuk membuat keputusan tentang suatu masalah dengan lebih cepat dan akurat (Yuswardi & dkk, 2022, p. 3). Simanullang & Simorangkir pada (Yuswardi & dkk, 2022, p. 8) menjelaskan bahwa keputusan dirancang secara spesifik guna mendukung seseorang yang harus mengambil keputusan-keputusan tertentu.

Menurut (Yuswardi & dkk, 2022, p. 16) dalam melakukan pendukung keputusan ada beberapa tahapan yang harus dilalui, antara lain adalah:

- (1) Tahap Pemahaman (Intelligence Phase) Pada tahap ini melakukan proses penelusuran untuk memetakan tingkat masalah (problem), serta mampu mengenali permasalahan yang terjadi. Input data yang diperoleh nantinya diproses dan diuji cobakan dalam rangka mendukung proses indentifikasi masalah;
- (2) Tahap Perencanaan (Design Phase) Pada tahap ini proses pengembangan pencarian solusi alternatif yang sangat mungkin untuk diambil. Diperlukan proses verifikasi dan validasi untuk dapat mengetahui tingkat keakuratan pada model yang diteliti;
- (3) Tahap Pemilihan (Choice Phase) Merupakan pemilihan fungsi untuk memilih berbagai solusi alternatif yang dapat dipilih, serta dimunculkan pada fase perencanaan dengan memperhatikan kriteria berdasarkan tujuan utama;
- (4) Tahap Implementasi (Implementasi Phase) Merupakan penerapan dengan melakukan penyesuaian rancangan sistem yang telah dibuat pada beberapa fase sebelumnya.

## **B. Tinjauan Pustaka**

Penelitian rujukan yang digunakan pada peneliti ini adalah:

- (1) Penelitian yang dilakukan oleh (Febrianto & rekan, 2021) dengan judul, "**PENERAPAN METODE K-MEANS UNTUK CLUSTERING PENGUNJUNG PERPUSTAKAAN ITN MALANG**"; perpustakaan Insitut Teknologi Nasional Malang merupakan salah satu fasilitas sarana dan prasarana yang menyediakan berbagai layanan penyediaan

informasi akademik, adapun terdapat koleksi seperti hasil-hasil penelitian, makalah, artikel, skripsi, dan tesis lulusan pascasarjana, terdapat data transaksi peminjaman selama ini tidak dimanfaatkan lebih lanjut untuk digali juga pola informasi yang mungkin berharga untuk digunakan dalam mengevaluasi koleksi buku maupun dokumentasi lain; peneliti menggagas menghasilkan aplikasi yang dapat mengelompokkan pengunjung dan peminjam menggunakan metode K-Means, membantu petugas perpustakaan untuk mengevaluasi informasi kunjungan dan peminjaman buku di perpustakaan ITN Malang; penelitian ini merupakan penelitian Research and Development yang berbasis aplikasi berbasis website, metode yang digunakan adalah metode K-Means, kebutuhan pada produk skripsi yang dikembangkan mencakup sistem hanya dapat digunakan diperangkat komputer, sistem hanya dapat diakses oleh petugas perpustakaan ITN Malang; hasil penelitian ini adalah produk berupa aplikasi website, produk mempunyai fitur yaitu sistem pada website dapat memberikan analisis informasi pengunjung dan peminjam, berdasarkan pengujian fungsional sistem seluruhnya berhasil dan berjalan dengan baik, berdasarkan pengujian pengguna diketahui hasil presentase responden 50% Sangat Baik, 48% Baik dan 2% Kurang Baik, berdasarkan penelitian disimpulkan bahwa semua fitur dapat berjalan dengan baik pada browser Mozilla Firefox 83.0 dan Google Chrome 87.0.4280.88, berdasarkan pengujian pengguna mayoritas menilai sangat baik terhadap penggunaan aplikasi keuntungan yang didapatkan jika menggunakan aplikasi ini yaitu dapat memudahkan dalam memberikan seran untuk pengadaan buku secara komputerisasi dan kerugian jika tidak menggunakan aplikasi ini yaitu dalam penentuan untuk pengadaan buku masih secara manual;

- (2) Penelitian yang dilakukan oleh (Kaisi. & rekan, 2021) dengan judul, **“PENGELOMPOKKAN JUMLAH PENGUNJUNG PERPUSTAKAAN DI DKI JAKARTA MENGGUNAKAN METODE K-MEANS”**; perpustakaan adalah suatu ruangan pengoleksian buku yang berfungsi sebagai wahana pendidikan, penelitian, informasi dan rekreasi untuk meningkatkan kecerdasan dan keberdayaan bangsa; salah satu jenis perpustakaan adalah perpustakaan daerah yang berkedudukan di tiap provinsi di Indonesia misalnya di DKI Jakarta; banyaknya pengunjung perpustakaan yang tercatat pada daftar buku pengunjung perpustakaan dapat diolah menjadi sebuah data yang menghasilkan informasi profil pengunjung perpustakaan dan alamat pengunjung perpustakaan; apabila dilakukan pengolahan data pada sumber data

tersebut maka dapat diketahui berbagai informasi yang bermanfaat dalam membantu pengelompokan jumlah pengunjung perpustakaan; oleh sebab itu, diperlukan pengelompokan jumlah pengunjung perpustakaan dengan menggunakan teknik data mining clustering dan metode K-Means; KMeans merupakan salah satu metode Clustering nonhierarki yang mempartisi data ke dalam bentuk satu atau lebih cluster sehingga data dengan karakteristik yang sama dikelompokkan dalam satu cluster yang sama pula; dari 276 data dibentuk 2 cluster yang telah dievaluasi menggunakan Davies Bouldin Index dengan nilai 0.363 yang mana cluster 0 menjadi cluster rendah dengan jumlah anggota sebanyak 74 data sedangkan cluster 1 menjadi cluster tinggi dengan jumlah anggota 202 data;

- (3) Penelitian yang dilakukan oleh (Fakhri & rekan, 2021) dengan judul, "**OPTIMALISASI PELAYANAN PERPUSTAKAAN TERHADAP MINAT BACA MENGGUNAKAN METODE K-MEANS CLUSTERING**"; Knowledge Discovery in Database (KDD) merupakan proses analisa yang terstruktur bertujuan mendapatkan informasi yang baru dan benar, menemukan pola dari data yang kompleks, dan bermanfaat; data mining merupakan inti dari proses KDD; clustering merupakan metode data mining yang cocok untuk pengoptimalisasikan pelayanan perpustakaan dikarenakan dapat mengklusterisasikan buku dengan dengan efektif dan efisien, dengan algoritma K-Means data dapat di clustering dan informasi dari setiap nilai centroid dari setiap cluster; pelayanan perpustakaan dapat mengoptimalisasikan penempatan buku sehingga santri bisa dengan cepat mencari buku sesuai dengan minat bacanya dengan lebih efektif dan bisa tertarik dengan buku yang lain karena berada dalam satu pengelompokan; sedangkan untuk pihak perpustakaan bisa memprioritaskan untuk pengadaan buku selanjutnya; optimalisasi pelayanan perpustakaan di cluster dengan menggunakan metode K-Means; clustering minat baca memiliki kriteria jumlah kesediaan buku, buku yang di pinjam, dan lama buku di pinjam; data buku di clustering menjadi 3 yaitu sangat diminati, diminati, dan kurang diminati; setelah melakukan proses perhitungan dari 40 sampel jenis buku maka menghasilkan 6 kali iterasi, dan di dapatkan hasil akhir 3 clustering yaitu cluster 1 sebanyak 4 buku yang sangat diminati, cluster 2 sebanyak 20 buku yang diminati, dan cluster 3 sebanyak 16 buku yang kurang diminati; penelitian ini dapat dijadikan sebagai acuan rekomendasi untuk pengoptimalisasikan pelayanan perpustakaan baik untuk tata letak maupun pengadaan buku dengan memprioritaskan jenis buku yang sangat diminati;

- (4) Penelitian yang dilakukan oleh (Zaliman & rekan, 2020) dengan judul, **“SISTEM PENENTUAN LOKASI PUSAT LAYANAN TERPADU BAGI PENDERITA PENYAKIT DEMAM BERDARAH DENGAN MENGGUNAKAN K-MEANS CLUSTERING“**; puskesmas merupakan kesatuan organisasi fungsional yang menyelenggarakan upaya kesehatan yang bersifat menyeluruh, terpadu, merata dapat diterima dan terjangkau oleh masyarakat; fungsi puskesmas adalah memberi pelayanan kesehatan kepada masyarakat melalui program Upaya Kesehatan Masyarakat (UKM) dan Upaya Kesehatan Perseorangan (UKP) yang merupakan fasilitas di barisan terdepan dalam memberikan layanan kesehatan pada masyarakat khususnya pencegahan dan pengobatan penyakit; penyakit terbagi dari 3 macam yaitu penyakit menular atau penyakit yang disebabkan oleh kuman yang menyerang tubuh manusia; pada penelitian ini akan berupaya melakukan penanganan terhadap penyakit menular yaitu tepatnya demam berdarah dengue (DBD); demam berdarah atau demam dengue (disingkat DBD) adalah infeksi yang disebabkan oleh virus dengue; nyamuk atau/ beberapa jenis nyamuk menularkan (atau menyebarkan) virus dengue; maka analisis yang terkomputerisasi menggunakan software data mining yang menunjang arus data dan informasi sesuai dengan kebutuhan penanganan penyakit demam berdarah dari proses-proses tersebut dan diperlukan pemilihan metode yang lebih cocok digunakan yaitu dengan menggunakan K-Means clustering untuk mendapatkan geolocation pasien DBD;
- (5) Penelitian yang dilakukan oleh (Anwar & rekan, 2020) dengan judul, **“PENENTUAN DAERAH PRIORITAS PELAYANAN AKTA KELAHIRAN DENGAN METODE K-NN DAN K-MEANS“**; data Kependudukan adalah data perseorangan atau data agregat yang terstruktur sebagai hasil dari kegiatan Pendaftaran Penduduk dan Pencatatan Sipil; Akta Kelahiran adalah Akta Catatan Sipil hasil pencatatan peristiwa kelahiran seseorang bayi yang dilaporkan kelahirannya akan terdaftar dalam Kartu Keluarga dan diberi Nomor Induk Kependudukan (NIK) sebagai dasar untuk memperoleh pelayanan masyarakat lainnya; dari jumlah pelaporan akta kelahiran yang terintegrasi Sistem Informasi Administrasi Kependudukan (SIAK) tahun 2018 yang berjumlah 570.637 terdapat 503.946 yang dilaporkan terlambat dan hanya 66.691 yang dilaporkan secara umum; clustering merupakan metode yang digunakan untuk mengelompokan suatu data yang mirip dengan lainnya dalam satu kelompok ataupun ketidak miripan data

terhadap kelompok lainnya; K-Nearest Neighbor adalah sebuah metode untuk melakukan klasifikasi terhadap objek berdasarkan data pembelajaran yang jaraknya paling dekat dengan data uji k-means merupakan metode yang digunakan untuk membagi sejumlah objek ke dalam kelompok berdasarkan kategori-kategori yang ada dengan melihat titik tengah; dalam praproses data mining, data dibersihkan dengan mengisi data yang kosong dengan data yang paling mendominasi, dan menyeleksi atribut dengan menggunakan metode information gain; berdasarkan metode k-nearest neighbor untuk memprediksi keterlambatan pelaporan dan metode k-means untuk mengelompokkan daerah yang menjadi prioritas pelayanan dengan 10.000 data kependudukan akta kelahiran tahun 2019 memiliki performa yang cukup baik menghasilkan prediksi dengan akurasi 74,00 % dan dengan K=2 pada k-means menghasilkan index davies bouldin sebesar 1,179;

- (6) Penelitian yang dilakukan oleh (Kurniawan & rekan, 2023) dengan judul, **“KLASTERISASI WILAYAH PRIORITAS VAKSIN MENGGUNAKAN ALGORITMA K-MEANS CLUSTERING”**; algoritma K-Means Clustering adalah metode klastering yang dilakukan secara partisi (partitional clustering); penggunaan klaster bermaksud untuk mempartisi sejumlah objek ke dalam kelompoknya dimana setiap objek ke dalam kelompok yang terdekat sehingga akan menghasilkan kelompok dengan perbedaan yang signifikan; selain itu upaya yang telah dilakukan oleh pemerintah untuk mencegah transmisi penyakit yang lebih luas antara lain dengan menerapkan pembatasan sosial berskala besar dan memonitor wilayah-wilayah yang banyak terjadi migrasi penduduk lokal maupun asing serta melakukan vaksinasi; melihat data banyaknya orang yang tertular, meninggal dunia akibat virus Covid-19 sampai sembuh yang terjadi di berbagai penjuru daerah di Indonesia; untuk itu perlu dibuat klasterisasi daerah dari zona merah, kuning dan hijau yang berarti untuk zona merah sendiri berarti daerah itu adalah daerah bahaya dan daerah dengan jumlah terjangkit virus dengan jumlah cukup banyak. Penelitian ini bertujuan untuk memecahkan permasalahan yaitu menghasilkan aplikasi yang dapat memberikan informasi tentang wilayah yang di prioritaskan vaksin di Sumatera Utara, melakukan klasterisasi daerah Sumatera Utara, mengetahui wilayah prioritas vaksin dengan menggunakan algoritma K-Means; untuk mengetahui hasil dari penggunaan aplikasi algoritma K-Means yaitu aplikasi dapat mengelompokkan kasus covid 19 setiap Kabupaten/Kota yang terdapat pada Provinsi Sumatera Utara ke dalam klasterisasi C1 (Tinggi), C2 (Sedang) dan C3 (Rendah);

berdasarkan hasil klasterisasi, Medan berada pada klasterisasi C1 dengan nilai jarak 0.00 sehingga dapat diprioritaskan dalam kegiatan vaksinasi covid 19; pada klasterisasi C2 (Sedang) terdapat satu Kabupaten, yaitu Deli Serdang dengan nilai iterasi ke-4 dengan jarak 0.00; pada klasterisasi C3 (Rendah) terdapat 31 Kabupaten/Kota Data kasus covid 19 dari Maret - November 2021 di Provinsi Sumatera Utara dilakukan perhitungan sebanyak 4 iterasi hingga tidak ada lagi perubahan data dalam proses klasterisasi;

- (7) Penelitian yang dilakukan oleh (Syalwa & rekan, 2023) dengan judul, **“PENERAPAN ALGORITMA K-MEANS UNTUK PENENTUAN WILAYAH PENJUALAN POTENSIAL PADA PERUSAHAAN JASA CLEANING SERVICE”**; promosi merupakan hal penting bagi perusahaan dalam mendorong penjualan serta membantu konsumen mengenal dan mengingat merek perusahaan; apabila tanpa promosi yang tepat, suatu bisnis tidak mungkin dikenal konsumen sehingga dapat merugikan pertumbuhan dan penjualan merek; hal tersebut sama dengan kondisi terjadi di perusahaan cleaning service yang berlokasi di Jawa Timur dimana perusahaan memiliki permasalahan pada saat menerima pelanggan di beberapa tempat yang berbeda di Jawa Timur perusahaan yakni tidak mempunyai informasi tentang area potensial untuk melakukan promosi penawaran jasanya sehingga hal tersebut berdampak bagi perusahaan yaitu tidak dapat merancang promosi yang tepat sasaran dan perusahaan bisa melewatkan peluang bisnis yang berharga; tujuan dari penelitian ini adalah untuk melakukan pengolahan data penjualan perusahaan dari tahun 2021 sampai dengan tahun 2023 dengan data mining menggunakan pendekatan clustering k-means mengelompokkan 3 cluster yakni Cluster 0 (C0) wilayah kurang berpotensi, Cluster 1 (C1) cukup berpotensi, Cluster 2 (C2) sangat berpotensi; berdasarkan hasil dari pengolahan data didapatkan Cluster 0 berjumlah 4 wilayah kurang berpotensi yakni wilayah Mojokerto, Jombang, Pasuruan, Malang; Cluster 1 berjumlah 3 wilayah cukup berpotensi yakni wilayah Madura, Gresik, Lamongan; pada cluster 2 berjumlah 2 wilayah sangat berpotensi, yakni wilayah Sidoarjo, dan Surabaya;
- (8) Penelitian yang dilakukan oleh (Indiani & rekan, 2022) dengan judul, **“PENENTUAN LOKASI ATM BANK SYARIAH INDONESIA DI WILAYAH JEMBER KOTA MENGGUNAKAN K-MEANS CLUSTERING”**; semua transaksi tersebut tidak terlepas dari peran ATM sebagai pendukung peningkatan keuangan badan usaha; oleh karena

itu, keberadaan ATM saat ini telah menjadi kebutuhan umum masyarakat modern; saat ini, jumlah penduduk Kabupaten Jember sebanyak 2.568,88 jiwa, dan telah berdiri 10 ATM, yang berarti rasio jumlah penduduk dan jumlah ATM tidak seimbang, sehingga perlu dilakukan pengelompokan data agar diperoleh lokasi pendirian ATM yang dapat menjangkau desa-desa di sekitarnya; penelitian ini mengkaji metode K-Means Clustering dalam kasus penentuan lokasi ATM berdasarkan jarak dari pusat keramaian; perhitungan jarak menggunakan Google Maps K-Means Clustering digunakan untuk mengelompokkan 40 titik data pusat keramaian tiap desa menjadi 6 cluster; penelitian ini mengasumsikan bahwa centroid adalah lokasi pendirian ATM dan anggota cluster adalah desa yang dapat dijangkau oleh centroid. Hasil pengelompokan C1-C6 adalah Pasar Tanjung, Pasar Pakusari, MTSN 2 Jember, KUA Arjasa, SMPN 3 Arjasa, dan SMAN 4 Jember secara berurutan;

(9) Penelitian yang dilakukan oleh (Vlachou, Georgaki, Pagkratis, & Poulos, 2021) dengan judul, **“EVALUASI PENGELOMPOKAN TAK TERAWASI PADA LAYANAN PERPUSTAKAAN UMUM”**; tugas ini meneliti kasus evaluasi layanan perpustakaan umum; untuk tujuan ini digunakan indeks standar ISO, yang kami sesuaikan dengan indeks MOPAB (Total Quality Unit of Academic Perpustakaan); secara keseluruhan, kami menggunakan tujuh indeks; tujuan dari tugas ini adalah agar indeks-indeks ini dapat dinilai secara global dengan menggunakan pendapat para ahli; perumusan di atas dibuat dengan klasifikasi pengelompokan tanpa pengawasan, yang diimplementasikan oleh algoritma K-Means; hasil implementasi ini menunjukkan bahwa tingkat keberhasilannya melebihi 99%; hasil ini menunjukkan bahwa metode ini valid dan dapat digunakan untuk penilaian; hal di atas tujuan di atas pasti dapat dicapai juga;

(10) Penelitian yang dilakukan oleh (Malicay, Talirongan, Talirongan, & Saren, 2021) dengan judul, **“PENGELOMPOKAN KESADARAN LAYANAN E-GOVERNMENT PADA UNIT PEMERINTAH DAERAH MELALUI ALGORITMA K-MEANS”**; departemen Dalam Negeri dan Pemerintah Daerah (DILG) Republik Filipina menggunakan perangkat kinerja untuk mengukur unit-unit pemerintah daerah (LGU) kepada masyarakatnya; upaya ini merupakan inisiatif DILG untuk memastikan bahwa masyarakat merasakan layanan dari pemerintah daerah dan untuk memastikan tingkat kesadaran mereka; studi ini menggunakan kumpulan data gender dan status

pekerjaan dari responden yang menggunakan alat kinerja dan menggunakan pengelompokan K-Means untuk memberikan klaster yang sama dari kesadaran masyarakat berdasarkan jenis kelamin dan status pekerjaan; hasilnya, laki-laki lebih sadar akan layanan seperti tata kelola dan tanggap darurat, pekerjaan umum dan infrastruktur, pengelolaan lingkungan, serta promosi ekonomi dan investasi promosi; di sisi lain, perempuan lebih sadar akan layanan seperti kesehatan, dukungan untuk pendidikan, dan kesejahteraan social; dalam hal status pekerjaan, mereka yang bekerja mengetahui layanan seperti tata kelola dan tanggap darurat, pekerjaan umum dan infrastruktur, dan promosi ekonomi dan investasi; sebaliknya, para penganggur mengetahui layanan seperti kesehatan, pendidikan, kesejahteraan sosial, dan pengelolaan lingkungan.

*Tabel 2. 1 Tinjauan Studi*

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
1	Andy Febrianto (2021)	Penerapan Metode K-Means Untuk Clustering Pengunjung Perpustakaan ITN Malang	Pengelompokan data Pengunjung Perpustakaan	JATI (Jurnal Mahasiswa Teknik Informatika) Volume 5 No. 1 Maret 2021, <a href="https://ejournal.itn.ac.id/index.php/jati">https://ejournal.itn.ac.id/index.php/jati</a>	Menghasilkan informasi berupa pengelompokan ke dalam 3 cluster
2	Esi Faisan Kaisi Rahmat (2021)	Pengelompokan Jumlah Pengunjung Perpustakaan Di DKI Jakarta Menggunakan Metode K-Means	Pengelompokan jumlah Pengunjung Perpustakaan	eProsiding Sistem Informasi (POTENSI) Volume 2 No. 1 Juni 2021, <a href="https://eprosiding.ars.ac.id/index.php/psi/article/view/378">https://eprosiding.ars.ac.id/index.php/psi/article/view/378</a>	Sebagai acuan Dalam Penggunaan Metode Data Mining CRISP-DM
3	Dwiki Aulia Fakhri (2021)	Optimalisasi Pelayanan Perpustakaan terhadap Minat Baca Menggunakan Metode K-Means Clustering	Kurang Optimalnya Pelayanan Perpustakaan	Jurnal Informasi dan Teknologi Volume 3 No. 3 2021, <a href="https://jidt.org/jidt/article/view/137">https://jidt.org/jidt/article/view/137</a>	Menghasilkan 3 cluster dengan menggunakan metode K-Means untuk optimalisasi pelayanan perpustakaan

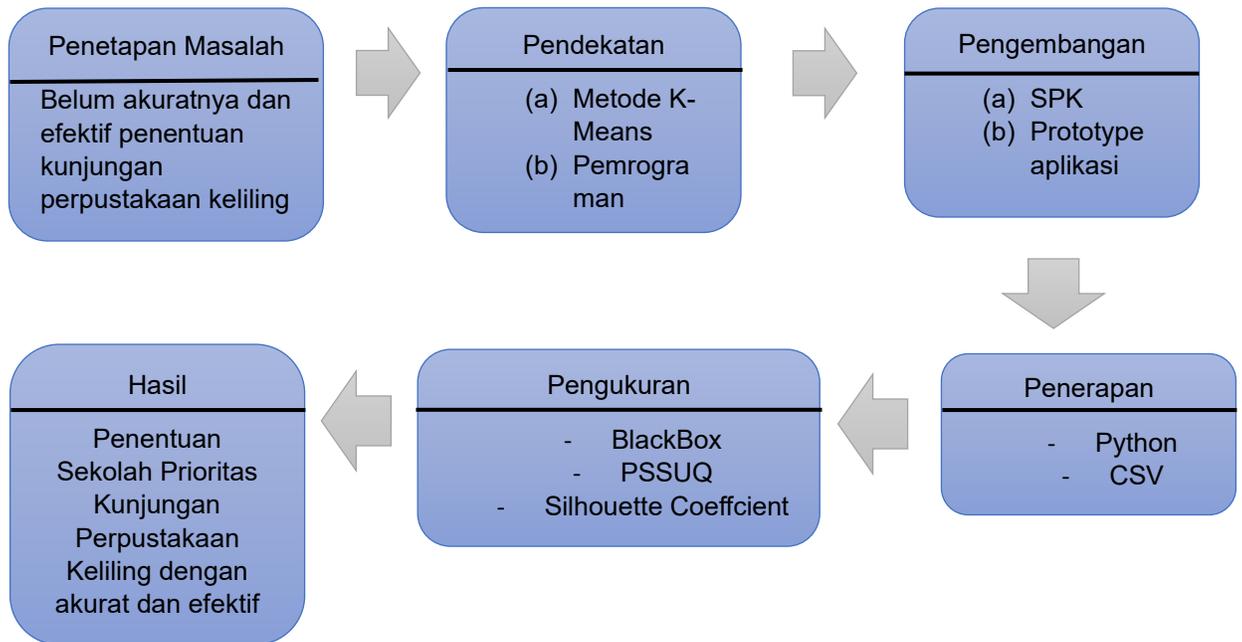
No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
4	Iski Zaliman (2020)	Sistem Penentuan Lokasi Pusat Layanan Terpadu Bagi Penderita Penyakit Demam Berdarah Dengan Menggunakan K-Means Clustering	Kurang terpadu dan menyeluruh pelayanan puskesmas untuk masyarakat	Jurnal Ilmiah Informatika Global Volume 11 No. 2 Desember 2020, <a href="https://ejournal.uigm.ac.id/index.php/IG/article/view/1225">https://ejournal.uigm.ac.id/index.php/IG/article/view/1225</a>	Sebagai acuan bahwa algoritma K-Means dapat digunakan untuk sistem penentuan lokasi
5	Ade Muchlis Maulana Anwar (2020)	Penentuan Daerah Prioritas Pelayanan Akta Kelahiran Dengan Metode K-NN Dan K-Means	Keterlambatan pelaporan akta kelahiran	KOMPUTASI (Jurnal Ilmiah Ilmu Komputer dan Matematika) Volume 17 No. 1 Januari 2020, <a href="https://journal.unpak.ac.id/index.php/komputasi/article/view/1884">https://journal.unpak.ac.id/index.php/komputasi/article/view/1884</a>	Algoritma K-Means menghasilkan 3 cluster dan tingkat akurasi sebesar 74,00%
6	Rahmat Kurniawan (2023)	Klasterisasi Wilayah Prioritas Vaksin Menggunakan Algoritma K-Means Clustering	Banyaknya kasus orang tertular, meninggal dunia akibat virus Covid-19	KLIK: Kajian Ilmiah Informatika dan Komputer Volume 4 No. 3 Desember 2023, <a href="https://djournals.com/klik/article/view/1334">https://djournals.com/klik/article/view/1334</a>	Melakukan klasterisasi wilayah menggunakan algoritma K-Means
7	Reynalda Vonna Syalwa (2023)	Penerapan Algoritma K-Means Untuk Penentuan Wilayah Penjualan Potensial Pada Perusahaan Jasa Cleaning Service	Kurangnya informasi tentang area potensial untuk melakukan promosi	TEMATIK: Jurnal Teknologi Informasi Komunikasi (e-journal) Volume 10 No. 2 (2022), <a href="https://jurnal.plb.ac.id/index.php/tematik/article/view/1518/753">https://jurnal.plb.ac.id/index.php/tematik/article/view/1518/753</a>	Metode K-Means dapat digunakan sebagai metode untuk pengelompokan wilayah
8	Nila Al Indiani (2022)	Penentuan Lokasi ATM Bank Syariah Indonesia Di Wilayah Jember Kota Menggunakan K-Means Clustering	Kurang terjangkau nya mesin ATM Bank Syariah Indonesia di Kabupaten Jember	Seminar Nasional Matematika, Geometri, Statistika, dan Komputasi SeNa-MaGeStiK 2022, <a href="https://jurnal.unej.ac.id/index.php/prosiding/article/view/33496">https://jurnal.unej.ac.id/index.php/prosiding/article/view/33496</a>	Sebagai acuan bahwa algoritma K-Means dapat digunakan untuk sistem penentuan lokasi
9	Sofia Vlachou	Evaluasi Pengelompokan Tak Terawasi pada Layanan Perpustakaan Umum	Mengelompokan layanan yang tak terawasi pada perpustakaan umum	WSEAS TRANSACTIONS on ADVANCES in ENGINEERING EDUCATION, <a href="https://wseas.com/journals/articles.php?id=3680">https://wseas.com/journals/articles.php?id=3680</a>	Melakukan pengelompokan terhadap layanan perpustakaan umum dan menghasilkan tingkat keberhasilan 99%

No	Nama Peneliti	Judul Penelitian	Permasalahan	Jurnal	Kontribusi
10	Charies Malicay (2021)	Pengelompokan Kesadaran Layanan E-Government pada Unit Pemerintah Daerah melalui Algoritma K-Means	Mengelompokan kesadaran layanan E-Government pada pemerintah daerah	International Journal of Research Publications (IJRP.ORG), <a href="https://ijrp.org/paper-detail/1764">https://ijrp.org/paper-detail/1764</a>	Menggunakan algoritma K-Means pada pengelompokan kesadaran layanan E-Government

Berdasarkan tinjauan pustaka yang sudah dipaparkan pada tabel 2.1 didapatkan pengetahuan yang dijadikan rujukan dalam pelaksanaan penelitian ini. Kontribusi dalam rujukan penelitian ini memberikan pengetahuan ilmu mengenai permasalahan pada penentuan sekolah prioritas kunjungan perpustakaan keliling dengan proses perhitungan K-Means. Dalam penelitian ini lebih menekankan terhadap variabel yang digunakan yaitu, nama satuan pendidikan, bentuk pendidikan, status sekolah, kecamatan, jumlah judul buku, jumlah eksemplar, jumlah anggota, jumlah siswa, jumlah pustakawan, jumlah tenaga teknis, keterangan prioritas. Dalam referensi yang di dapat yang berkaitan dengan metode serta permasalahan penelitian & pengembangan saat ini, tidak ditemukan kesamaan secara spesifik. Dengan demikian penelitian ini dapat memberikan kontribusi baru terkait dengan penerapan metode k-means untuk penentuan sekolah prioritas kunjungan perpustakaan keliling.

### C. Kerangka Pemikiran

Metode K-Means bekerja mempartisi data ke dalam bentuk satu atau lebih cluster yang dapat dijadikan metode pemecahan masalah layanan perpustakaan keliling, kerangka pemikiran dibawah ini dibuat berdasakan masalah-masalah yang terjadi;



Gambar 2.4 Kerangka Pemikiran

Berdasarkan gambar 2.4 tahap dalam kerangka pemikiran adalah;

- (1) Penetapan Masalah, yaitu belum akurat dan efektif dalam penentuan kunjungan perpustakaan keliling.
- (2) Pendekatan, penelitian dengan metode *K-Means* terhadap data yang telah dikumpulkan.
- (3) Pengembangan, melakukan perancangan use case dan prototype aplikasi.
- (4) Penerapan, menggunakan Bahasa pemrograman python dan CSV sebagai data yang digunakan.
- (5) Pengukuran, menggunakan blackbox dan PSSUQ.
- (6) Hasil, prototype yang menampilkan penentuan sekolah prioritas kunjungan perpustakaan keliling secara tepat dan efektif.

#### **D. Hipotesis**

Metode K-Means merupakan salah satu metode data clustering non hirarki yang berusaha mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik cluster yang sama ke dalam satu cluster dan data yang mempunyai karakteristik cluster yang berbeda ke dalam cluster yang lain. Telah dilakukan penelitian sebelumnya oleh (Anwar & rekan, 2020) dengan judul Penentuan Daerah Prioritas Pelayanan Akta Kelahiran Dengan Metode K-NN Dan K-MEANS pada penelitian tersebut didapatkan hasil dari 10.000 data kependudukan akta kelahiran tahun 2019 memiliki performa yang cukup baik menghasilkan prediksi dengan akurasi 74,00 % dan dengan K=2 pada k-means menghasilkan index davies bouldin sebesar 1,179. Berdasarkan penelitian sebelumnya maka hipotesis dalam penelitian ini adalah metode K-Means diduga dapat menentukan sekolah prioritas kunjungan perpustakaan keliling.