BAB II KERANGKA TEORITIS

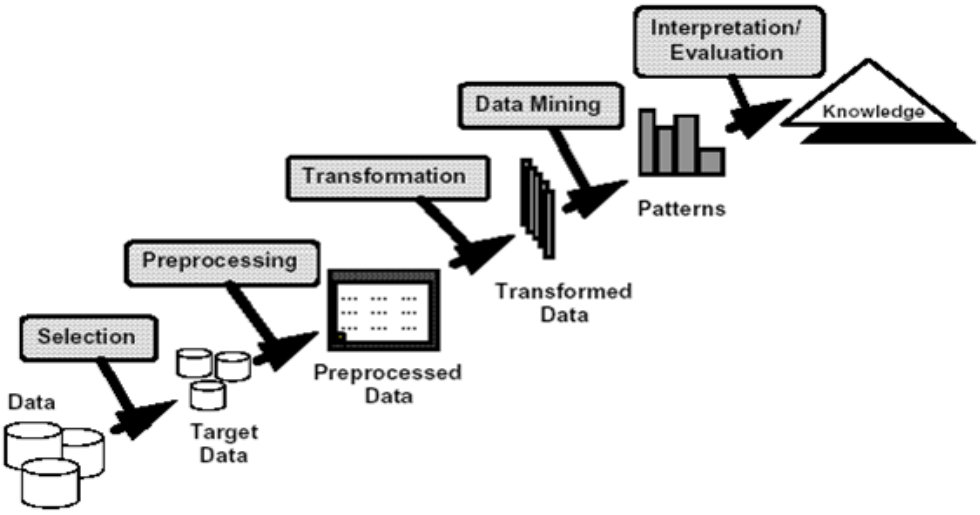
A. Landasan Teori

1. Data

(Andjarwati *et al*., 2021, p. 9) mengutarakan bahwa data adalah keterangan mengenai suatu hal dalam bentuk angka, kalimat, dan penjelasan yang dapat menunjukkan gambaran atau keadaan tentang sesuatu persoalan pada umumnya yang dikaitkan dengan waktu dan tempat. Menurut (Wijaya *et al*., 2024, p. 15) data adalah sekumpulan informasi atau fakta mentah dalam bentuk simbol, angka, kata-kata atau citra dan diperoleh melalui proses pengamatan atau dari sumber tertentu. Data adalah himpunan angka yang merupakan nilai dari unit sampel kita sebagai hasil mengamati dan mengukurnya (Harnani and Rasyid, 2015, p. 4).

2. Data Mining

Menurut (Rahayu *et al*., 2024, p. 2) data *mining* adalah proses memperoleh pengetahuan atau informasi berharga dari kumpulan data yang besar dan kompleks. (Jollyta *et al*., 2020, p. 119–120) mengutarakan bahwa data *mining* adalah proses mengidentifikasi pola, tren, dan informasi berharga dari data yang besar proses logis untuk menemukan informasi dan pola yang berguna dari kumpulan data yang sangat besar melalui pengumpulan, ekstraksi, analisis, dan penerapan statistik, yang juga dikenal sebagai *Knowledge Discovery, Knowledge Extraction, data/pattern analysis, and information harvesting.*



Gambar 2.1 Tahapan Proses KDD
Sumber : (Habibi and Suryansah, 2024, p. 7)

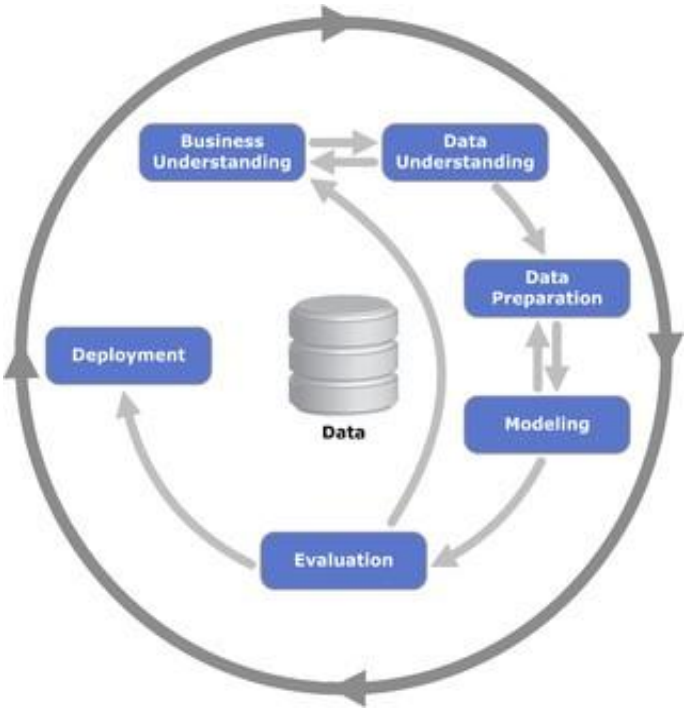
Knowledge Discovery in Database merupakan sebuah proses untuk mengungkap pengetahuan dalam basis data, yang melibatkan ekstraksi atau

identifikasi pola, pengetahuan, dan informasi dari kumpulan data besar (Sinurat, 2024, p. 4). Menurut (Habibi and Suryansah, 2020, p.7) KDD merupakan proses yang kompleks untuk mencari dan mengidentifikasi pola dalam data, di mana pola yang ditemukan harus valid, inovatif, bermanfaat, dan mudah dipahami. Gambar 2.1 menunjukkan pada proses KDD terdiri dari beberapa tahapan dalam (Habibi and Suryansah, 2020, p. 8–10), sebagai berikut:

(1) *data selection*, pada tahap ini, data operasional diseleksi dan disimpan dalam berkas terpisah untuk digunakan dalam proses data *mining*;
(2) *pre-processing/cleaning*, pada tahap ini, sebelum data *mining* perlu dilakukan data cleaning untuk menghapus duplikasi, memperbaiki inkonsistensi dan kesalahan, serta data enrichment dengan informasi relevan, termasuk data eksternal, untuk mendukung proses KDD;
(3) *transformation*, pada tahap ini, data yang dipilih ditransformasi agar sesuai untuk data *mining*, dengan proses kreatif yang bergantung pada jenis atau pola informasi yang dicari;
(4) *data mining*, pada tahap ini, pola atau informasi menarik ditemukan dalam data terpilih menggunakan teknik, metode, atau algoritma yang disesuaikan dengan tujuan dan proses KDD;
(5) *interpretation/evaluation*, pada tahap ini, pola informasi dari data *mining* disajikan dalam format yang mudah dipahami dan diperiksa apakah bertentangan dengan fakta atau hipotesis sebelumnya, dalam tahap interpretasi KDD.

3. Metode CRISP-DM

(Muttaqin *et al*., 2023, p. 18) mengutarakan bahwa *Cross-Industry Standard Process for Data Mining* atau CRISP-DM adalah salah satu model proses data *mining* (data *mining* framework) yang awalnya (1996) dibangun oleh 5 perusahaan yaitu Integral Solutions Ltd (ISL), Teradata, Daimler AG, NCR Corporation dan OHRA. Menurut (Muttaqin *et al*., 2023, p. 19) CRISP-DM merupakan model data *mining* model yang masih banyak diterapkan di kalangan industri secara luas, salah satunya karena kemampuannya dalam mengatasi berbagai masalah di proyek-proyek data *mining*.



Gambar 2.2 Tahapan Proses CRISP-DM

Sumber : (Sumadireja, 2020, p. 68)

Gambar 2.2 menunjukkan enam tahapan CRISP-DM dalam (Sumadireja *et al*., 2020, p. 68–70), sebagai berikut:

(1) *business understanding*, pada tahapan ini mencakup penentuan tujuan bisnis, evaluasi situasi saat ini, penetapan tujuan data *mining*, dan pengembangan rencana penelitian;

(2) *data understanding*, pada tahapan ini merupakan tahap pemahaman data yang diperoleh melalui observasi langsung dan analisis terhadap berkas data pelanggan pada objek penelitian;

(3) *data preparation,* pada tahapan ini merupakan tahap pengolahan data atau disebut juga tahapan persiapan data;

(4) *modelling,* pada tahapan ini dilakukan pengaplikasian teknik pemodelan, data yang dipilih pada fase persiapan digunakan sebagai parameter untuk melakukan peramalan dan optimasi;

(5) *evaluation*, pada tahapan ini diperlukan analisis yang menginterpretasikan hasil pengolahan data sebelumnya untuk menghasilkan peramalan dan optimasi nilai;

(6) *deployment*, pada tahapan ini mengatur dan mempresentasikan pengetahuan atau informasi yang diperoleh dalam format khusus agar dapat digunakan oleh pengguna.

4. Algoritma K-Means

Menurut (Arhami and Nasir, 2020, p.148) algoritma merupakan salah satu metode *clustering* yang sering digunakan untuk mengelompokkan data berdasarkan kesamaan karakteristik atau ciri-ciri yang serupa. Kelompok data yang terbentuk disebut *cluster*. Algoritma ini juga efisien dalam hal komputasi dan menghasilkan hasil yang baik jika *cluster* - *cluster* yang terbentuk padat, berbentuk hampir bulat, dan dapat memisahkan fitur-fitur ruang dengan baik. K-Means adalah algoritma yang sederhana, mudah dimengerti, dan mudah disesuaikan untuk mengolah data streaming. K-Means juga mudah diimplementasikan dan memiliki kompleksitas waktu serta ruang yang relatif rendah. Adapun langkah-langkah melakukan *clustering* dengan metode K-Means dalam (Jollyta *et al*., 2020, p. 119–120), sebagai berikut:

(1) pilih jumlah *cluster* K;

(2) inisialisasi K pusat *cluster* umumnya dilakukan secara acak dengan memberikan nilai awal berupa angka random;

(3) alokasikan setiap data ke *cluster* terdekat berdasarkan jarak ke pusat *cluster* , menggunakan rumus *Euclidean* untuk menentukan data masuk ke *cluster* mana, dengan rumus:

$$Di,j = \sqrt{((x_1 i - x_1 j)^2 + (x_2 i - x_2 j)^2 + \ldots + (x_k i - x_k j)^2)} \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots (2.1)$$

dimana:

D(*i,j*) = jarak data ke *i* ke pusat *cluster j;*

X(*k,i*) = data ke *i* pada atribut data ke *k;*

X(*k,j*) = titik pusat ke *j* pada atribut ke *k;*

(4) hitung ulang pusat *cluster* berdasarkan anggota saat ini, dengan menggunakan rata-rata atau median data dalam *cluster* tersebut;

(5) ulangi langkah ke 3 dan 4, apabila alokasi data *cluster* tidak berubah lagi maka proses *clustering* selesai.

Contoh penyelesaian perhitungan K-Means yang dapat dipaparkan dengan rumus *Euclidean Distance* pada sekelompok data nilai matakuliah mahasiswa (Jollyta *et al*., 2020, p. 123–129), sebagai berikut:

No	Nim	Nama	MPSI	IMK	POO II	Sister	Web & MM	APSI I	Instrumentasi	Dasar Akuntansi
1	004	Ando	75	71	71	81	72	74	72	76
2	016	Rudy	80	72	75	73	72	73	71	77

No	Nim	Nama	MPSI	IMK	POO II	Sister	Web & MM	APSI I	Instrumentasi	Dasar Akuntansi
3	017	Stef	73	80	74	81	88	85	73	78
4	026	Yuki	74	82	80	73	74	73	71	71
5	034	Dwi	78	78	77	87	86	78	75	87
6	035	Ana	71	71	66	78	71	71	75	71
7	036	Eric	73	71	71	85	76	76	71	71
8	045	Irana	73	72	66	73	73	76	71	76
9	049	Riza	75	72	71	80	73	77	72	73
10	051	N.Dwi	76	71	71	71	71	7	73	75
11	052	Tefa	72	67	71	71	73	77	75	80
12	055	Rosi	76	73	71	74	74	73	73	76

(1) sebelum diproses dengan algoritma K-Means, pastikan data nilai mahasiswa sudah lengkap sesuai dengan tahapan Knowledge Discovery in Database (KDD);

(2) selanjutnya, dilakukan Iterasi 1 pada data nilai mahasiswa dengan menentukan titik pusat *cluster* secara acak dari data, pada contoh ini, diambil 2 titik pusat *cluster* , yakni;
data nilai ketiga sebagai pusat *cluster* pertama = C1
C1 = 73 80 74 81 88 85 73 78
data nilai kelima sebagai pusat *cluster* kedua = C2
C2 = 78 78 77 87 86 78 75 87

(3) lakukan perhitungan kedekatan setiap data nilai mahasiswa dengan masing-masing pusat *cluster* menggunakan rumus *Euclidean Distance*, berdasarkan pusat *cluster* awal pada Iterasi 1;

C1	C2	Jarak Terpendek
10,4403	21,7256	10,44030651
0	24,5602	0
24,1454	14,5602	14,56021978
14,1774	26,0384	14,17744688
24,0208	0	0
15,6525	28,8097	15,65247584
16,4621	22,1359	16,46207763

C1	C2	Jarak Terpendek
11,8743	26,2298	11,87434209
11,1355	22,4722	11,13552873
6,85565	27,0924	6,8556546
12,2882	25,8457	12,28820573
6,55744	22,9783	6,557438524

C1 = (((nilai mhs-1 mk-1 dikurang nilai pusat *cluster* C1-1)^2) + ((nilai mhs-1 mk-1 dikurang nilai pusat *cluster* C1-2)^2) + ((nilai mhs-1 mk-1 dikurang nilai pusat *cluster* C1-3)^2) + ((nilai mhs-1 mk-1 dikurang nilai pusat *cluster* C1-8)^2))^0,5);

C1 = (((75-73)^2) + ((71-80)^2) + ((71-74)^2) + ((81-81)^2) + ((72-88)^2) + ((74-85)^2) + ((72-73)^2) + ((76-78)^2)^0,5

= 10,4403

Untuk menghitung jarak ke C2 caranya sama, tetapi menggunakan semua nilai dari pusat *cluster* C2 yang telah ditentukan sebelumnya;

C2 = (((75-78)^2) + ((71-78)^2) + ((71-77)^2) + ((81-87)^2) + ((72-86)^2) + ((74-78)^2) + ((72-75)^2) + ((76-87)^2)^0,5

= 21,7256

Kemudian cari jarak terpendek dari kedua *cluster* dengan mengambil nilai yang paling kecil dari setiap data (bisa dilakukan dengan rumus MIN);

(4) setelah nilai *cluster* diperoleh, langkah selanjutnya adalah mengelompokkan data untuk melihat jumlah data di tiap *cluster* (gunakan angka 1 atau 0 sebagai penanda, tergantung nilai terendah antara C1 dan C2, misalnya, jika nilai C1 lebih kecil (seperti 10,4403), maka beri tanda 1 pada C1), hasil tahap pengelompokan data pada Iterasi 1;

No	C1	C2
1	1	0
2	1	0
3	0	1
4	1	0
5	0	1
6	1	0
7	1	0
8	1	0
9	1	0
10	1	0

No	C1	C2
11	1	0
12	1	0

(5) langkah selanjutnya adalah menentukan nilai pusat *cluster* yang baru, yaitu dengan menjumlahkan semua data dalam tiap *cluster* (hanya data yang termasuk dalam *cluster* tersebut), lalu dibagi dengan jumlah datanya. Hasilnya adalah nilai rata-rata untuk setiap mata kuliah di masing-masing *cluster* ;

(6) setelah mendapatkan nilai pusat *cluster* yang baru, lakukan kembali perhitungan jarak seperti pada langkah 3 untuk Iterasi 2;

(7) lakukan kembali pengelompokan data *cluster* berdasarkan hasil perhitungan pada langkah 6, dengan hasil untuk Iterasi 2;

No	C1	C2
1	1	0
2	1	0
3	0	1
4	1	0
5	0	1
6	1	0
7	1	0
8	1	0
9	1	0
10	1	0
11	1	0
12	1	0

(8) selanjutnya, ditampilkan hasil akhir dari pengelompokan, untuk titik pusat *cluster* 1;

No	Nim	Nama	MPSI	IMK	POO II	Sister	Web & MM	APSI I	Instrumentasi	Dasar Akuntansi
1	004	Ando	75	71	71	81	72	74	72	76
2	016	Rudy	80	72	75	73	72	73	71	77
3	026	Yuki	74	82	80	73	74	73	71	71
4	035	Ana	71	71	66	78	71	71	75	71
5	036	Eric	73	71	71	85	76	76	71	71

No	Nim	Nama	MPSI	IMK	POO II	Sister	Web & MM	APSI I	Instrumentasi	Dasar Akuntansi
6	045	Irana	73	72	66	73	73	76	71	76
7	049	Riza	75	72	71	80	73	77	72	73
8	051	N.Dwi	76	71	71	71	71	7	73	75
9	052	Tefa	72	67	71	71	73	77	75	80
10	055	Rosi	76	73	71	74	74	73	73	76

hasil pengelompokan titik pusat *cluster* 2;

No	Nim	Nama	MPSI	IMK	POO II	Sister	Web & MM	APSI I	Instrumentasi	Dasar Akuntansi
3	017	Stef	73	80	74	81	88	85	73	78
5	034	Dwi	78	78	77	87	86	78	75	87

(9) berdasarkan hasil pengelompokan (*clustering*), diperoleh informasi bahwa sebagian besar mahasiswa (sekitar 83,3%) yang termasuk dalam *cluster* 1 memiliki rata-rata nilai keseluruhan di bawah 75, 16,6% atau 2 orang mahasiswa yang memiliki nilai di atas 75. Ini menunjukkan bahwa penguasaan mahasiswa terhadap 8 mata kuliah berada pada level menengah, dimana perlu dilakukan evaluasi terhadap metode pembelajaran yang digunakan pada mata kuliah tersebut.

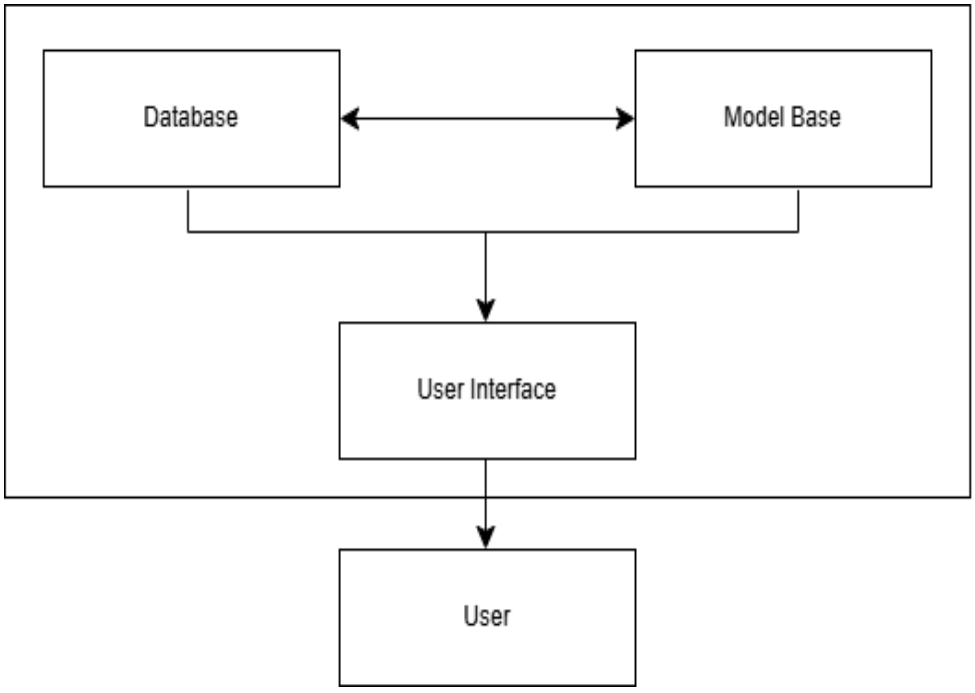
5. Clustering

(Setiawan et al., 2023, p. 7) mengutarakan bahwa c*lustering* merupakan metode dalam data *mining* yang digunakan untuk mengelompokkan data yang memiliki karakteristik sama ke dalam satu kelompok atau *cluster* berdasarkan kesamaan sifat atau ciri tertentu. Menurut (Mokoginta et al., 2024, p. 75) *clustering* mirip dengan klasifikasi, tetapi lebih fokus pada menemukan kesamaan antar objek, lalu mengelompokkannya berdasarkan perbedaan dengan objek lain. C*lustering* dilakukan dengan mengelompokkan data yang sejenis dari kumpulan data yang lebih besar. Teknik ini membantu menemukan kelompok-kelompok berdasarkan kesamaan. Dengan *clustering*, data minoritas yang tersebar bisa

dikumpulkan ke dalam satu kelompok besar yang memiliki kemiripan (Jollyta *et al*., 2020, p. 53).

6. Sistem Pendukung Keputusan (SPK)

Menurut (Mahendra *et al*., 2024) Sistem Pendukung Keputusan (SPK) atau Decision Support System (DSS) adalah sistem yang membantu pengambil keputusan dengan menyediakan solusi, rekomendasi, atau informasi untuk menangani masalah semi-terstruktur. (Ariantini *et al*., 2023, p. 1) mengutarakan bahwa Sistem Pendukung Keputusan (SPK) adalah suatu sistem yang dirancang untuk mendukung proses pengambilan keputusan dalam organisasi atau lingkungan tertentu. Sistem Pendukung Keputusan (SPK) adalah suatu sistem informasi berbasis komputer yang adaptif, interaktif, dan fleksibel, dirancang untuk membantu menyelesaikan masalah manajemen yang tidak terstruktur (Marie et al., 2023, p. 9).



Gambar 2.3 Komponen Utama SPK

Sumber : (Marie and Septiani, 2023, p. 10)

Gambar 2.3 menunjukkan tiga komponen utama Sistem Pendukung Keputusan dalam (Marie *et al*., 2023, p. 13–15), sebagai berikut:

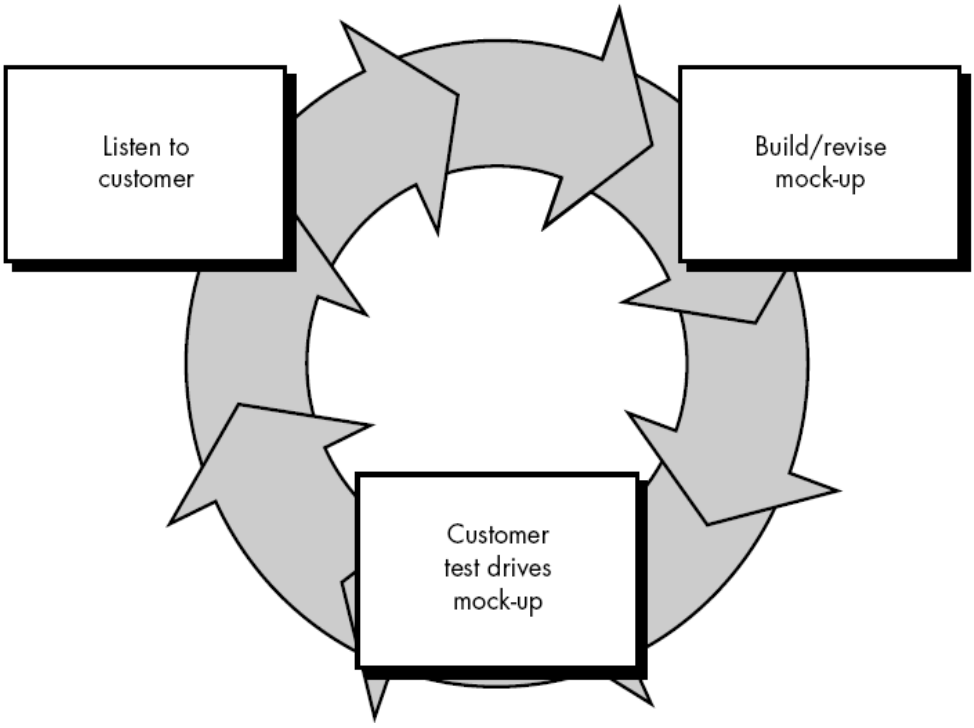
(1) sub sistem manajemen data, merupakan suatu sistem yang terdiri dari database yang berisi data sesuai kebutuhan dan dikelola menggunakan perangkat lunak yang disebut sistem manajemen database (DBMS);

(2) sub sistem manajemen model, merupakan suatu paket perangkat lunak yang memungkinkan integrasi berbagai jenis model, seperti model statistik, keuangan, atau model kuantitatif lainnya;

(3) sub sistem user interface, merupakan suatu perangkat lunak yang disebut sistem manajemen antarmuka pengguna (*User Interface Management System*/UIMS) memungkinkan pengguna berkomunikasi dan mengoperasikan SPK melalui subsite mini.

7. System Development Life Cycle (SDLC)

(Manuaba *et al*., 2023, p. 13) mengutarakan bahwa siklus pengembangan sistem (SDLC) adalah sebuah proses bagaimana sistem informasi mendukung kebutuhan bisnis dengan merancang, membangun, dan mengirimkan perangkat lunak kepada pengguna, di mana pemodelan dan pengembangan sebelumnya penting untuk menghasilkan perangkat lunak yang efektif. Menurut (Damayanthi *et al*., 2024, p. 16) Siklus Hidup Pengembangan Perangkat Lunak merupakan rangkaian langkah terstruktur yang diterapkan untuk merancang, mengembangkan, dan menguji perangkat lunak yang berkualitas tinggi**.** System Development Life Cycle (SDLC) adalah tahapan dan metodologi yang perlu dilakukan dalam mengembangkan perangkat lunak (Arifin et al., 2022, p. 35).



Gambar 2.4 Model Prototyping

Menurut (Rukmana *et al*., 2023, p. 143) model prototyping adalah model yang berfokus pada aktivitas pembuatan prototype sistem perangkat lunak. (De Fretes *et al*., 2024, p. 135) mengutarakan bahwa model pengembangan sistem kompleks di mana pengembang dan pengguna bersama-sama mengidentifikasi kebutuhan, dengan hanya mendefinisikan tujuan umum perangkat lunak tanpa merinci kebutuhan input, proses, dan output. Gambar 2.4 menunjukkan model prototyping dalam (Putra, 2021, p. 8), sebagai berikut:

(1) listen to customer, pada tahap ini mengumpulkan kebutuhan sistem yang akan dibangun dengan menganalisis sistem yang sedang berjalan untuk mengidentifikasi masalah yang dihadapi;
(2) build/revise mock-up, pada tahap ini melibatkan perancangan dan pembuatan prototipe sistem yang sesuai dengan kebutuhan yang telah ditetapkan;
(3) customer test drives mock-up, pada tahap ini, prototipe diuji oleh pengguna, dievaluasi untuk mengidentifikasi kekurangan, dan diperbaiki oleh pengembang hingga sesuai dengan kebutuhan pengguna sebelum melanjutkan ke tahap berikutnya.

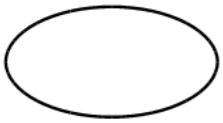
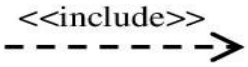
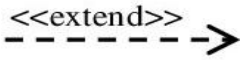
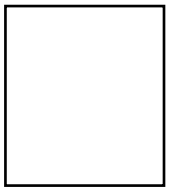
8. *Unified Modeling Language* (UML)

Menurut (Sari *et al*., 2024, p. 15) *Unified Modeling Language* (UML) adalah bahasa standar dalam industri yang digunakan untuk merancang, memvisualisasikan, dan mendokumentasikan sistem perangkat lunak. UML adalah bahasa berbasis grafis yang digunakan untuk memvisualisasikan, menentukan, membangun, dan mendokumentasikan sistem pengembangan perangkat lunak berbasis objek (Object-Oriented) (Sumirat *et al*., 2023, p. 73).

(Habibi *et al*., 2020, p. 79) mengutarakan bahwa UML adalah bahasa yang digunakan untuk menentukan secara rinci, menggambarkan, membangun, dan mencatat informasi yang dihasilkan atau diperlukan dalam proses pengembangan perangkat lunak, seperti sistem perangkat lunak, pemodelan bisnis, serta sistem non-perangkat lunak lainnya. Adapun diagram yang terdapat pada *Unified Modeling Language* (UML) dalam (Habibi *et al*., 2020, p. 80–93), sebagai berikut:

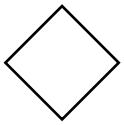
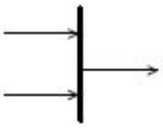
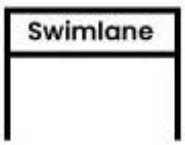
(a) *usecase* diagram, merupakan deskripsi yang menggambarkan fungsionalitas sistem serta interaksi antara pengguna dengan sistem sejak tahap awal pengembangan;

Nama	Simbol	Keterangan
Actor		menggambarkan orang, sistem atau proses ketika berinteraksi dengan *usecase* atau sistem;

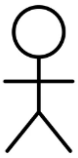
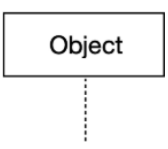
Nama	Simbol	Keterangan
Usecase		menggambarkan fungsionalitas atau aksi yang ditampilkan sistem;
Include		menggambarkan *usecase* tambahan diperlukan agar *usecase* utama dapat berfungsi, dengan arah panah include yang menunjuk ke *usecase* pendukung tersebut;
Extend		menggambarkan *usecase* tambahan tetap dapat berjalan secara mandiri meskipun tanpa *usecase* lain, dengan arah panah extend yang mengarah ke *usecase* tambahan tersebut;
Association		menggambarkan hubungan antar actor dan *usecase* yang saling berinteraksi;
Generalization		menggambarkan hubungan umum dan khusus antara dua *usecase*, di mana salah satunya memiliki fungsi lebih global dibandingkan yang lain, dengan arah panah menuju *usecase* yang berperan sebagai generalisasi;
Sistem		menggambarkan paket yang merepresentasikan sistem dalam lingkup yang terbatas;

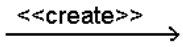
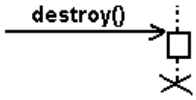
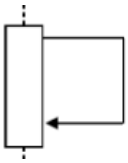
(b) *activity* diagram, merupakan alur aktivitas yang menggambarkan proses dalam sebuah sistem serta memberikan informasi menyeluruh, di mana *activity* diagram dapat disusun berdasarkan satu atau lebih usecase;

Nama	Simbol	Keterangan
Initial State		menggambarkan status awal pada diagram aktivitas biasanya ditunjukkan dengan satu titik awal, namun dalam beberapa kasus sebuah *activity* diagram dapat memiliki lebih dari satu status atau titik awal untuk menggambarkan

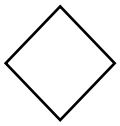
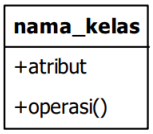
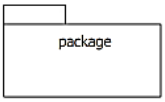
Nama	Simbol	Keterangan
		variasi proses yang dapat dimulai;
Activity		menggambarkan aktivitas atau kegiatan yang berlangsung, yang biasanya dituliskan menggunakan kata kerja;
Decision		menggambarkan proses dengan pilihan lebih dari satu aktivitas;
Join		menggambarkan beberapa aktivitas yang digabungkan menjadi satu alur;
Final State		menggambarkan status akhir yang ditunjukkan dengan simpul akhir, menandakan bahwa rangkaian aktivitas telah selesai atau berhenti;
Swimlane		menggambarkan pemisahan organisasi yang bertanggung jawab atas suatu aktivitas, di mana pengelompokan aktivitas dilakukan berdasarkan peran aktor dalam urutan yang sama;

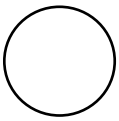
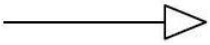
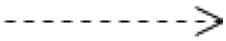
(c) *sequenc*e diagram, merupakan perilaku objek dalam usecase, termasuk siklus hidup objek serta pesan yang dikirim dan diterima oleh objek tersebut;

Nama	Simbol	Keterangan
Actor		menggambarkan orang, sistem, atau proses yang berinteraksi dengan usecase maupun sistem yang dikembangkan;
Lifeline		menggambarkan siklus hidup dari suatu objek;
Object		menggambarkan objek yang berinteraksi melalui pertukaran pesan;

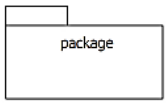
Nama	Simbol	Keterangan
Pesan tipe *create*		menggambarkan bahwa suatu objek mengakhiri keberadaan objek lain, dengan arah panah menuju objek yang diakhiri; umumnya jika terdapat *create* maka akan ada pula *destroy*;
Pesan tipe *destroy*		menggambarkan suatu objek yang setelah menjalankan operasi atau metode menghasilkan kembalian kepada objek lain, dengan arah panah menunjuk pada objek yang menerima hasil tersebut;
Pesan tipe *return*		menggambarkan suatu objek yang setelah menjalankan operasi atau metode memberikan hasil kembalian kepada objek tertentu, dengan arah panah menunjuk pada objek penerima kembalian;
Pesan tipe *self*		menggambarkan komponen berupa panah dengan garis putus-putus yang berfungsi untuk menunjukkan relasi antar objek itu sendiri;

(d) *class* diagram, merupakan diagram yang menampilkan kelas serta paket-paket yang ada dalam sebuah perangkat lunak;

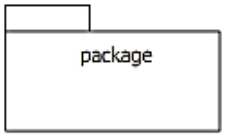
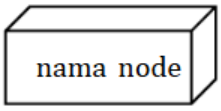
Nama	Simbol	Keterangan
Association		menggambarkan hubungan antara satu objek dengan objek lainnya;
Public Association		menggambarkan bahwa asosiasi yang melibatkan lebih dari dua objek sebaiknya dihindari;
Class		menggambarkan himpunan objek yang memiliki atribut dan operasi yang sama;
Package		menggambarkan package sebagai wadah yang berisi satu atau lebih kelas;
Direct Association		menggambarkan relasi antar kelas, di mana satu kelas digunakan oleh kelas

Nama	Simbol	Keterangan
		lain, dan asosiasi biasanya dilengkapi dengan multiplicity;
Interface		menggambarkan konsep yang serupa dengan interface dalam pemrograman berorientasi objek;
Generalization		menggambarkan relasi antar kelas dengan makna generalisasi–spesialisasi (umum–khusus);
Dependency		menggambarkan relasi antar kelas dengan makna adanya kebergantungan antara kelas satu dengan lainnya;
Aggregation		menggambarkan relasi antar kelas dengan makna tertentu sesuai hubungan yang terjadi;

(e) *component* diagram, merupakan visualisasi komponen perangkat lunak dalam sistem serta keterkaitan antar komponen dalam keseluruhan sistem;

Nama	Simbol	Keterangan
Package		menggambarkan posisi atau penempatan komponen dalam suatu sistem;
Component System		menggambarkan perangkat keras atau objek yang terdapat dalam sistem;
Dependency		menggambarkan hubungan ketergantungan antar komponen, di mana arah panah menunjukkan komponen yang digunakan;
Interface		menggambarkan antarmuka atau interface yang berfungsi sebagai perantara agar interaksi tidak langsung menuju objek;
Link		menggambarkan arah relasi antara komponen yang saling terhubung;

(f) *deployment* diagram, merupakan diagram yang merealisasikan alur proses dan hubungan antar elemen dalam suatu sistem, sehingga memudahkan pengguna dalam memahami sistem yang dibuat.

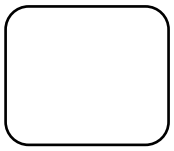
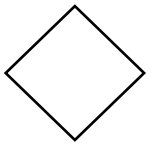
Nama	Simbol	Keterangan
Package		menggambarkan penempatan atau alokasi node dalam suatu sistem;
Node		menggambarkan perangkat keras yang menjalankan perangkat lunak yang tidak dikembangkan sendiri, di mana elemen tersebut memiliki peran dalam perancangan dan komponen tetap mempertahankan definisi yang sama seperti pada diagram komponen sebelumnya;
Dependency		menggambarkan hubungan antar elemen, di mana arah panah menunjukkan elemen yang digunakan;
Link		menggambarkan hubungan yang terjadi antara node;

9. Business Process Modeling Notation (BPMN)

(Nugroho, 2020, p. 26) mengutarakan bahwa *Business* P*rocess Modeling Notation* (BPMN) adalah Representasi grafis ini digunakan untuk menggambarkan proses bisnis dalam pemodelan proses bisnis dan menyediakan notasi standar yang mudah dipahami oleh semua pemangku kepentingan. Menurut (Nugroho, 2020, p. 26) BPMN juga sebagai alat untuk menjelaskan cara merancang proses bisnis dan mendeskripsikan secara teknis bagaimana proses bisnis dijalankan untuk keperluan otomatisasi. Adapun notasi yang terdapat pada *Business Process Modeling Notation* (BPMN) dalam (Wardani *et al*., 2021, p. 50), sebagai berikut:

(a) *flow objects,* merupakan elemen yang menggambarkan karakteristik dari suatu proses bisnis;

Nama	Simbol	Keterangan
Event		menunjukkan proses dimulai;

Nama	Simbol	Keterangan
Activity		menunjukkan kegiatan/aktivitas;
Gateway		menunjukkan percabangan atau pengambilan keputusan;

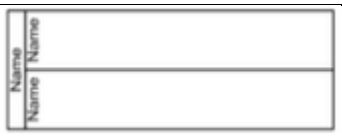
(b) *data type,* merupakan kebutuhan sistem informasi yang memiliki data input dan menghasilkan output;

Nama	Simbol	Keterangan
Data Object		menunjukkan data yang dibutuhkan untuk sebuah aktivitas;
Data Input		menunjukkan data masukkan yang dibutuhkan untuk sebuah aktivitas;
Data Output		menunjukkan data luaran yang dihasilkan dari sebuah aktivitas;

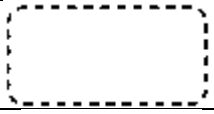
(c) connecting objects, merupakan elemen yang berfungsi untuk menghubungkan flow object;

Nama	Simbol	Keterangan
Sequence Flow		menunjukkan alur urutan aktivitas;
Message Flow		menunjukkan alur pesan antar 2 partisipan atau antar pool;
Association		menunjukkan alur yang menghubungkan artifak;

(d) swimlanes, merupakan pengelompokan dari berbagai elemen model yang digunakan untuk memisahkan dan mengorganisasi aktivitas berdasarkan peserta siapa yang bertanggung jawab atas setiap kegiatan;

Nama	Simbol	Keterangan
Pool		*pool* menunjukkan representasi partisipan pada proses kolaborasi;
Lane		*lane* menunjukkan mengelola dan mengkategorikan aktivitas;

(e) artifacts, merupakan elemen yang berfungsi untuk menyampaikan informasi tambahan tentang suatu proses. Bentuk dan penggunaannya bervariasi serta dapat disesuaikan dengan standar pemahaman BPMN yang diterapkan;

Nama	Simbol	Keterangan
Group		menunjukkan tanda untuk aktivitas yang bisa dikelompokkan;
Text Annotation		menunjukkan informasi tambahan berupa teks dari sebuah aktivitas;

10. Bahasa Pemrograman

Menurut (Hartatik *et al*., 2023, p.14–15) bahasa pemrograman adalah alat dengan sintaks berupa kumpulan perintah yang digunakan oleh programmer untuk memberikan instruksi kepada sistem, yang kemudian diterjemahkan menjadi logika yang dimengerti oleh sistem untuk menjalankan program. (Ariyadi et al., 2024, p. 1) mengutarakan bahwa bahasa pemrograman adalah sekumpulan aturan sistem dan sintaksis yang digunakan untuk mendefinisikan program sistem. Bahasa pemrograman adalah sebuah sarana yang digunakan untuk memungkinkan komunikasi antara manusia dan sistem (Ariyadi et al., 2024, p. 3).

(Maesaroh *et al*., 2024, p. 1) mengutarakan bahwa python adalah sistem pemrograman tingkat tinggi yang dirancang dengan filosofi mengutamakan keterbacaan kode dan sintaksis sederhana, sehingga menjadi pilihan favorit bagi pemula dan pengembang berpengalaman. Menurut (Surya *et al*., 2023, p. 1) python adalah sistem pemrograman yang serbaguna dan dapat digunakan untuk berbagai keperluan, seperti pengembangan website, analisis data, machine learning, data science, dan kecerdasan buatan (AI). Pemrograman phyton merupakan sebuah sistem pemrograman serbaguna yang mampu mengeksekusi berbagai instruksi secara langsung menggunakan metode pemrograman berorientasi objek serta sistem dinamis untuk meningkatkan keterbacaan sintaks (Setiawan and Vania, 2022, p. 4).

11. Database

(Hesananda, 2025, p. 8) mengutarakan bahwa database adalah kumpulan informasi terstruktur atau data yang terorganisir, biasanya disimpan dalam sistem komputer. Menurut (Putra *et al*., 2024, p. 89) database adalah kumpulan data yang terorganisir dalam format yang memudahkan pengambilan, penyimpanan, dan manipulasi informasi. Database adalah sekumpulan data yang tersusun secara terstruktur, diorganisir, dan disimpan dalam di satu tempat untuk memudahkan penyimpanan, pengelolaan, serta akses yang efisien (Munawar et al., 2023, p. 81).

12. Silhouette Coefficient

Menurut (Prihandoko *et al*., 2024, p. 31) *Silhouette Coefficient* adalah metrik yang digunakan untuk menilai seberapa efektif objek dikelompokkan dalam *cluster*. (Ma'ady *et al*., 2024, p. 152) mengutarakan bahwa *Silhouette Coefficient* adalah teknik yang digunakan untuk menentukan jumlah *cluster* optimal. Nilai *Silhouette Coefficient* tertinggi diperoleh dari penjumlahan nilai s yang telah dikonversi, dan jika jumlah tersebut lebih besar dibandingkan *cluster* lainnya, maka *cluster* tersebut dianggap terbaik karena memiliki lebih sedikit nilai s di bawah 0 (Mulaab, 2021, p. 87).

13. PSSUQ (Post-Study System Usability Questionnaire)

(Malik and Frimadani, 2023, p. 64) mengutarakan bahwa PSSUQ adalah kuesioner standar yang digunakan untuk mengukur seberapa mudah dan bermanfaat suatu antarmuka (UI/UX) bagi pengguna. Kuesioner ini membantu mengevaluasi pengalaman pengguna agar interaksi lebih optimal dan solusi yang diberikan lebih sesuai dengan kebutuhan mereka. Menurut (Wahyuningrum, 2023, p. 58) PSSUQ adalah alat ukur yang digunakan untuk menilai kepuasan dan kegunaan produk. Kuesioner ini membantu tim desain dan UX memahami apakah alur kerja sudah efektif dan memuaskan pengguna.

Post-Study System Usability Questionnaire (PSSUQ) merupakan instrumen penelitian yang dikembangkan oleh IBM untuk menilai usability suatu sistem (Pratasik, 2022, p. 7–8). Kuesioner ini berisi 16 pertanyaan dengan 7 pilihan jawaban, dan digunakan untuk mengevaluasi 4 aspek usability (Handayani *et al.*, 2022, p. 42–43). Aspek usability yang dinilai diantaranya adalah:

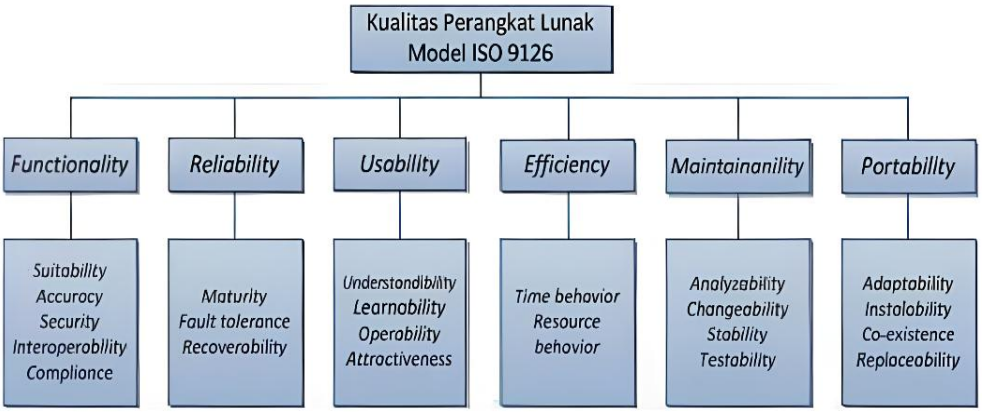
(1) *System Usefulness,* ukuran kepuasan pengguna terhadap kemampuan sistem dalam menjalankan fungsinya dengan baik;
(2) *Information Quality,* ukuran kepuasan pengguna terhadap kemampuan sistem dalam membantu menyelesaikan tugas melalui informasi dan navigasi;
(3) *Interface Quality*, ukuran kepuasan pengguna terhadap kenyamanan dan kemudahan menggunakan tampilan antarmuka sistem;

(4) *Overall Satisfaction,* ukuran kepuasan pengguna dilihat secara keseluruhan dari ketiga aspek di atas.

14. Pengujian ISO 9126

Menurut (Darujati and Azam, 2022, p. 122) ISO 9126 adalah standar internasional untuk menilai kualitas perangkat lunak. Standar ini menjelaskan model, karakteristik kualitas, dan metrik yang digunakan untuk mengevaluasi seberapa baik suatu produk software. (Oktarino *et al*., 2024, p. 276) mengutarakan bahwa ISO 9126 dipakai untuk mengatur kualitas dalam proyek pengembangan perangkat lunak. Jika sebuah proyek memenuhi standar ini, kualitasnya dianggap sangat baik, walaupun saat digunakan oleh pengguna masih bisa muncul beberapa kekurangan.

Kualitas perangkat lunak dapat diukur menggunakan ukuran dan metode tertentu, serta melalui pengujian perangkat lunak. Salah satu standar tolak ukur yang digunakan untuk menilai kualitas perangkat lunak adalah ISO 9126, yang dibuat oleh *International Organization for Standardization* (ISO) dan *International Electro Commission* (IEC) (Darujati and Azam, 2022, p. 86–87).



Gambar 2.5 Karakteristik Kualitas ISO 9126

Sumber : (Darujati and Azam, 2022, p. 86)

Gambar 2.5 menunjukkan 6 karakteristik faktor kualitas ISO 9126 dalam (Darujati and Azam, 2022, p. 86), sebagai berikut:

(1) *Functionality* (Fungsionalitas), yaitu kemampuan software untuk menyediakan fungsi sesuai kebutuhan dan memuaskan pengguna;

(2) *Reliability* (Kehandalan), yaitu kemampuan software untuk mempertahankan tingkat kinerja tertentu/performance dari software (ex: akurasi, konsistensi, kesederhanaan, toleransi kesalahan);

(3) *Usability* (Kebergunaan), yaitu kemampuan software untuk dipahami, dipelajari, digunakan, dan menarik bagi pengguna;
(4) *Efficiency* (Efisiensi), yaitu kemampuan software untuk memberikan kinerja yang sesuai dan relatif terhadap jumlah sumber daya yang digunakan pada saat keadaan tersebut (ex: efisiensi penyimpanan);
(5) *Maintainability* (Pemeliharaan), yaitu kemampuan software untuk dimodifikasi. Modifikasi meliputi koreksi, perbaikan atau adaptasi terhadap perubahan lingkungan, persyaratan, dan spesifikasi fungsional (ex: konsistensi);
(6) *Portability* (Portabilitas), yaitu kemampuan software untuk ditransfer dari satu lingkungan ke lingkungan lain atau kemampuan software beradaptasi saat digunakan di area tertentu (ex:self-documentation,teratur).

15. Tentang Status Gizi Balita

(Yunawati *et al*., 2023, p. 20) mengutarakan bahwa status gizi merupakan faktor penting dalam mencapai derajat kesehatan yang optimal dan faktor penting untuk kesehatan optimal, mendukung pertumbuhan anak, daya tahan tubuh, kecerdasan, dan produktivitas, sedangkan gizi buruk, akibat ketidakseimbangan antara asupan dan kebutuhan nutrisi serta rendahnya kualitas makanan masih menjadi permasalahan di berbagai daerah di Indonesia. Menurut (Yunawati *et al*., 2023, p. 44) masalah status gizi pada balita dapat dianalisis melalui pengukuran antropometri seperti usia, berat badan (BB), dan tinggi badan (TB), dengan menggunakan timbangan digital untuk balita yang dapat berdiri, timbangan dacin untuk yang berusia di bawah 2 tahun, length-board untuk mengukur panjang badan, dan microtoise untuk tinggi badan, yang hasilnya disajikan dalam indikator antropometri seperti BB/U, TB/U, dan BB/TB.

16. Tentang Pemberian Penyuluhan Kesehatan

Penelitian ini bertujuan untuk memberikan penentuan terkait dalam pemberian penyuluhan. Menurut (Ropii *et al*., 2024, p. 50) Penyuluhan kesehatan merupakan proses memberikan informasi dan edukasi untuk meningkatkan pemahaman, sikap, dan perilaku masyarakat tentang kesehatan, dengan teknik yang disesuaikan untuk mendorong perubahan perilaku dan meningkatkan kesehatan secara keseluruhan. (Ropii *et al*., 2024, p. 50) mengutarakan bahwa efektivitas penyuluhan dapat ditingkatkan dengan melibatkan komunitas, memilih metode yang tepat, dan memanfaatkan teknologi terkini, seperti pendidikan berbasis komunitas yang dilakukan oleh lembaga kesehatan atau organisasi non-pemerintah untuk meningkatkan pengetahuan, sikap, dan perilaku kesehatan masyarakat.

B. Tinjauan Studi

Penelitian ini didasarkan dari penelitian sebelumnya dan penting untuk mengkaji informasi dari penelitian sebelumnya yang relevan. Penelitian sebelumnya sudah banyak digunakan dengan kasus yang sama dan metode yang sama, yaitu algoritma K-Means. Kajian tersebut dapat menjadi bahan pertimbangan dalam mengidentifikasi perbedaan antara penelitian terdahulu dan penelitian yang akan dilakukan. Penelitian rujukan ini diperlukan sebagai landasan dalam melaksanakan penelitian, khususnya mengenai persebaran status gizi balita untuk penentuan pemberian penyuluhan. Penelitian sebelumnya berfungsi sebagai sumber pengetahuan yang berkontribusi pada pengembangan penelitian ini. Berikut adalah penelitian-penelitian yang menjadi acuan dalam penulisan ini:

(1) ***Clustering*** **Data Antropometri Balita Untuk Menentukan Status Gizi Balita Di Kelurahan Jumput Rejo Sukodono Sidoarjo** (Ali, 2020)

Penelitian ini menggunakan teknik data *mining* dengan algoritma K-Means untuk mengelompokkan data. Data menggunakan 3 parameter parameter di antaranya ukuran antropometri yaitu, berat badan menurut umur (BB/U), tinggi badan menurut umur (TB/U). Hasil pengelompokan menunjukkan terdapat 5 *cluster* , yaitu: *cluster* 1 dengan 37 balita, *cluster* 2 sebanyak 30 balita, *cluster* 3 berisi 28 balita, *cluster* 4 dengan 33 balita, dan *cluster* 5 sebanyak 22 balita. Hasil analisis data antropometri balita di Desa Jumput Rejo Sukodono menunjukkan bahwa sebanyak 37 balita memiliki status gizi buruk, 30 balita mengalami gizi kurang, 28 balita memiliki gizi baik, 33 balita memiliki gizi lebih, dan 22 balita mengalami obesitas dari total 150 data antropometri balita.

(2) Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih) (Ni Komang Sri Julyantari *et al*., 2021)

Penelitian ini menggunakan metode K-Means agar dapat memberikan informasi kelompok *cluster* status gizi balita pada Banjar Titih. Data yang digunakan data sampel berasal dari 3 posyandu yaitu, Titih Kelod, Titih Tengah, dan Titih Kaler dengan jumlah data 46 balita. Hasil dari penelitian ini nilai gizi balita Banjar Titih dapat diklasterisasikan menggunakan metode K-Means dengan 3 buah parameter yaitu berat badan, jenis kelamin dan umur. Berdasarkan hasil perhitungan pengelompokan data status gizi yang dibagi menjadi 3 kelompok antara lain gizi normal, gizi buruk dan gizi lebih pada Banjar Titih didapatkan 3 buah *cluster* yaitu dengan presentase gizi normal 47.83%, gizi buruk 30.43%, dan gizi lebih 21.74%.

(3) Penentuan Kelompok Status Gizi Balita dengan Menggunakan Metode K-Means (Purwaningrum *et al*., 2021)

Penelitian ini menggunakan algoritma K-Means untuk klasterisasi status gizi balita di Posyandu Tanjung 1 RW 06, Desa Pepelegi, Sidoarjo, dengan tujuan agar penanganan gizi pada balita dapat lebih tepat sasaran. Pengelompokan status gizi dilakukan berdasarkan data berat badan, tinggi badan, dan umur balita yang tercatat pada bulan Februari 2018. Total sampel yang digunakan adalah 82 data yang berasal dari 82 balita yang berbeda, dan menghasilkan 3 *cluster* berdasarkan perhitungan menggunakan metode *Elbow* dan *Silhouette Coefficient*. Hasil klasterisasi ini dapat menjadi acuan bagi kader posyandu dalam memberikan pelayanan kepada balita, sehingga pelayanan yang diberikan dapat lebih tepat sasaran dan maksimal.

(4) Analisa Perbandingan Algoritma *Clustering* untuk Pemetaan Status Gizi Balita di Puskesmas Pasir Jaya (Fatonah dan Pancarani, 2022)

Penelitian ini menggunakan algoritma K-Means untuk mengimplementasikan algoritma klasterisasi, yaitu K-Means dan Fuzzy C-Means, dengan tujuan untuk mengevaluasi status gizi balita secara umum. Hasil evaluasi ini diharapkan dapat menjadi dasar bagi petugas kesehatan di puskesmas untuk melakukan pencegahan dini terhadap masalah gizi buruk. Data yang digunakan dalam penelitian ini merupakan data kuantitatif sekunder, yaitu data status gizi anak yang diambil dari Puskesmas Pasir Jaya, Cikupa pada tahun 2019, yang terdiri dari 1562 sampel. Hasil penelitian ini didasarkan pada beberapa parameter, antara lain berat badan, tinggi badan, data posyandu, dan umur. Berdasarkan hasil perhitungan, algoritma K-Means menghasilkan validasi terbaik sebesar 0,79, sedangkan algoritma Fuzzy C-Means menghasilkan validasi sebesar 0,78. Dari kedua algoritma tersebut, K-Means lebih efektif dalam menggambarkan status gizi balita di Puskesmas Pasir Jaya, karena memperoleh validasi tertinggi yaitu 0,79 dan menghasilkan 4 *cluster* yaitu gizi kurang, gizi obesitas, gizi buruk, dan gizi baik.

(5) Application of *Clustering*-Based Data *Mining* for the Assessment of Nutritional Status in Toddlers at Community Health Centers (Fianty *et al*., 2023)

Penelitian ini menggunakan algoritma *clustering* K-Means untuk mengelompokkan data balita berdasarkan status gizinya, dengan mempertimbangkan empat indeks antropometri (Berat/Berat Usia, Tinggi/Berat Usia, Berat/Tinggi, dan Indeks Massa Tubuh/Berat Usia) sebagai atribut. Data yang digunakan diperoleh dari Puskesmas Kelapa Dua Tangerang. Hasil penelitian menunjukkan terbentuknya tiga c*luster* dengan karakteristik yang berbeda, yang dianalisis lebih mendalam menggunakan teknik data *mining*

dengan algoritma *clustering* K-Means untuk memperoleh pemahaman lebih mengenai status gizi balita.

(6) Pengelompokan Status Gizi pada Anak Usia 3-5 Tahun Menggunakan Metode K-Means *Clustering* (Dewi *et al*., 2024)

Penelitian ini menggunakan metode K-Means *Clustering* untuk mendukung perhitungan dari setiap kriteria dan atribut yang telah ditentukan. Parameter yang digunakan yaitu, umur, tinggi badan dan berat badan. Berdasarkan hasil klasterisasi, terdapat tiga kategori penilaian, yaitu: Gizi Baik, Gizi Kurang, dan Gizi Buruk. Dari penelitian ini, dapat disimpulkan bahwa penerapan data *mining* dalam pengelompokan status gizi pada anak usia 3-5 tahun di Puskesmas Medan Amplas menggunakan metode K-Means *Clustering* telah berhasil dilakukan. Hal ini ditunjukkan dengan semakin mudahnya pengguna sistem dalam mengelompokkan status gizi anak, serta meningkatnya efisiensi dan akurasi dari hasil yang diperoleh.

(7) Implementasi Algoritma K-Means untuk Prediksi Status Gizi Balita pada Tiga Puskesmas di Kecamatan Simanindo (Napitu *et al*., 2024)

Penelitian ini menggunakan algoritma K-Means untuk memprediksi status gizi balita. Data yang digunakan diperoleh dari tiga Puskesmas di Kecamatan Simanindo, yaitu Puskesmas Ambarita, Puskesmas Tuktuk, dan Puskesmas Tomok. Data yang digunakan sebanyak 955 mencakup umur, tinggi badan, dan berat badan. Data dikelompokkan menjadi 4 *Cluster*, yaitu sangat kurang, kurang, normal, dan risiko lebih. Hasil penelitian menunjukkan bahwa algoritma K-Means efektif dalam memprediksi status gizi pada balita.

(8) Implementasi Metode K-Means *Clustering* untuk Menentukan Kondisi Gizi Balita (Studi Kasus : Puskesmas Mamsena) (Eko *et al*., 2024)

Penelitian ini menggunakan metode K-Means *Clustering* untuk mengelompokkan informasi yang sama menjadi sekumpulan informasi ke dalam kelompok-kelompok tertentu. Data ini menggunakan pengukuran antropometri yaitu jenis kelamin, usia, berat badan, tinggi badan, dan lingkar lengan. Dari 348 data balita yang digunakan terdapat 181 balita laki-laki dan 167 balita kelamin perempuan. Hasil penelitian terdapat 5 kelompok yaitu, kurang sehat, gizi kurang, gizi cukup, gizi baik dan obesitas dan teknik K-Means *Clustering* terbukti metode yang efektif dan efisien untuk menganalisis data.

(9) Data *Mining* and Visualization for Toddler Nutrition Monitoring in Community Health Centers (Fianty *et al*., 2024)

Penelitian ini menggunakan metode K-Means *Clustering* untuk mengelompokkan data balita berdasarkan status gizi yang dianalisis melalui 4 indeks antropometri,

yaitu Berat Badan/Umur, Tinggi Badan/Umur, Berat Badan/Tinggi Badan, dan Indeks Massa Tubuh/Umur. Data balita yang digunakan berasal dari Puskesmas dengan periode pencatatan dari Februari 2021 hingga Februari 2023, mencakup rata-rata 2.911 entri per tahun yang disimpan dalam format Microsoft Excel. Analisis ini menghasilkan tiga kategori kelompok, yaitu kelompok dengan sistem gizi yang baik, kelompok dengan rasio berat badan/tinggi badan yang rendah, dan kelompok dengan sistem tinggi badan/umur yang tidak memadai. Dataset yang digunakan mencakup atribut seperti jenis kelamin, tanggal lahir, desa/kelurahan, berat badan, dan tinggi badan. Hasil penelitian menunjukkan bahwa data balita dapat dikelompokkan menjadi 2 *cluster* : *cluster* 1 dan *cluster* 2. Melalui K-Means *Clustering*, penelitian berhasil mengkategorikan data balita menjadi 2 *cluster* sehingga membuka peluang untuk memahami karakteristik unik dari setiap *cluster* .

(10) Analysis of Malnutrition Status in Toddlers Using the K-Means Algorithm Case Study in DKI Jakarta Province (Sintawati dan Mariskhana, 2024)

Penelitian ini bertujuan untuk mengklasifikasikan status gizi buruk pada masyarakat anak-anak menggunakan algoritma K-Means, dengan studi kasus di DKI Jakarta. Bagian ini menganalisis hasil K-Means *Clustering* mengenai status gizi anak di DKI Jakarta dan membandingkannya dengan temuan penelitian terbaru. Penelitian ini menggunakan algoritma K-Means untuk mengklasifikasikan status gizi balita di DKI Jakarta, membagi 6 menjadi 3 *cluster* berdasarkan prevalensi status gizi kurang, normal dan stunting. Penerapan K-Means dalam penelitian ini memberikan pendekatan yang efisien untuk mengidentifikasi kelompok daerah yang memerlukan perhatian lebih dalam memerangi gizi buruk pada anak.

Pada Tabel 2.1 dirangkum penelitian rujukan yang digunakan dalam studi ini ke dalam sebuah tabel, mencakup informasi tentang peneliti, tahun, judul penelitian, sumber jurnal, serta kontribusi masing-masing penelitian, sebagai berikut:

Tabel 2.1 Tinjauan Studi

No	Peneliti/Tahun	Judul Penelitian	Sumber Jurnal	Kontribusi
1.	Amir Ali (2020)	*Clustering* Data Antropometri Balita untuk Menentukan Status Gizi Balita di Kelurahan Jumput Rejo Sukodono Sidoarjo	Jurnal Teknik Informatika dan Sistem Informasi, Vol. 7, No. 3, Desember 2020, DOI:10.35957/jatisi.v7i3.530, E-ISSN:2503-2933, P-ISSN:2407-4322,	Berkontribusi pada penerapan data *mining* dengan menggunakan Algoritma K-Means *Clustering*

No	Peneliti/Tahun	Judul Penelitian	Sumber Jurnal	Kontribusi
			https://jurnal.mdp.ac.id/index.php/jatisi/article/view/530/181	
2.	Ni Komang Sri Julyantari, I Komang Budiarta, Ni Made Dewi Kansa Putri (2021)	Implementasi K-Means Untuk Pengelompokan Status Gizi Balita (Studi Kasus Banjar Titih)	Jurnal Janitra Informatika dan Sistem Informasi, Vol. 1, No. 2 - Oktober 2021, DOI: 10.25008/janitra.v1i2.134, E-ISSN:2775-9490, https://janitra.org/index.php/home/article/view/134/14	Berkontribusi pada hasil *cluster* yaitu gizi kurang, gizi normal dan gizi lebih dengan menggunakan Algoritma K-Means
3.	Oktania Purwaningrum, Yudha Yunanto Putra, Amalia Anjani Arifiyanti (2021)	Penentuan Kelompok Status Gizi Balita dengan Menggunakan Metode K-Means	Jurnal Ilmiah Teknologi Informasi Asia, Vol.15, No.2 - Oktober 2021, E-ISSN:2580-8397; P-ISSN:0852-730X, https://jurnal.stmikasia.ac.id/index.php/jitika/article/view/594	Berkontribusi pada hitungan *Silhouette Coefficient* dengan menggunakan algoritma K-Means
4.	Nenden Siti Fatonah, Trisnansita Kintan Pancarani (2022)	Analisa Perbandingan Algoritma *Clustering* untuk Pemetaan Status Gizi Balita di Puskesmas Pasir Jaya	Teknologi Informasi dan Komunikasi, Vol. 18, No. 1, Januari 2022, E-ISSN:2721-0936, https://jurnal.untag-sby.ac.id/index.php/KONVERGENSI/article/view/5497	Berkontribusi pada hasil *cluster* yaitu gizi kurang, gizi obesitas, gizi buruk dan gizi baik dengan menggunakan Algoritma K-Means
5.	Melissa Indah Fianty, Monika Evelin Johan, Azka Aulia, Mella Margareta Veronica (2023)	Application of *Clustering*-Based Data *Mining* for the Assessment of Nutritional Status in Toddlers at Community Health Centers	Journal Information System and Informatics, Vol. 5, No. 4, December 2023, DOI:10.51519/journalisi.v5i4.586, E-ISSN:2656-4882, P-ISSN:2656-5935, https://journal-isi.org/index.php/i	Berkontribusi pada tahapan proses KDD yaitu Data Transformation dengan menggunakan atribut seperti TB/U, BB/U, dan TB/BB

No	Peneliti/Tahun	Judul Penelitian	Sumber Jurnal	Kontribusi
			si/article/view/586/280	
6.	Siska Dewi, Yohanni Syahra, Faisal Taufik (2024)	Pengelompokan Status Gizi pada Anak Usia 3-5 Tahun Menggunakan Metode K-Means *Clustering*	Jurnal Sistem Informasi Triguna Dharma (JURSI TGD), Vol. 3, Vol. 2, Maret 2024, DOI: 10.53513/jursi.v3i2.5939, E-ISSN:2503-2933, P-ISSN : 2828-1004, https://ojs.trigunadharma.ac.id/index.php/jsi/article/view/5939	Berkontribusi pada penerapan data *mining* dengan menggunakan Algoritma K-Means *Clustering*
7.	Stifani Napitu, Hanna Dewi M. Hutabarat (2024)	Implementasi Algoritma K-Means untuk Prediksi Status Gizi Balita pada Tiga Puskesmas di Kecamatan Simanindo	Jurnal Ilmu Komputer dan Teknologi, Vol. 5, No. 3, Oktober 2024, DOI:10.35960/ikomti.v5i3.1648, ISSN: 2746-4237 https://ejournal.uhb.ac.id/index.php/IKOMTI/article/view/1648	Berkontribusi pada penerapan data *mining* dengan menggunakan Algoritma K-Means *Clustering*
8.	Yulita Ejo, Yasinta Oktaviana Legu Rema, Hevi Herlina Ullu, Budiman Bas o (2024)	Implementasi Metode K-Means *Clustering* untuk Menentukan Kondisi Gizi Balita (Studi Kasus : Puskesmas Mamsena)	Jurnal Tekno Kompak, Vol. 19, No. 1, Februari 2025, DOI:10.33365/jtk.v19i1.4717, E-ISSN:2656-3525, P-ISSN: 1412-9663 https://ejurnal.teknokrat.ac.id/index.php/teknokompak/article/view/4717	Berkontribusi dalam bahasa pemrograman *Python* dengan menggunakan algoritma K-Means
9.	Melissa Indah Fianty, Monika Evelin Johan, Fahmy Rinanda Saputri, Fidelia Rahmah (2024)	Data Mining and Visualization for Toddler Nutrition Monitoring in Community Health Centers	Journal of System and Management Sciences, Vol.14, No. 7, Oktober 2024, DOI:10.33168/JS	Berkontribusi pada tahapan proses KDD yaitu Data Transformation Dengan menggunakan

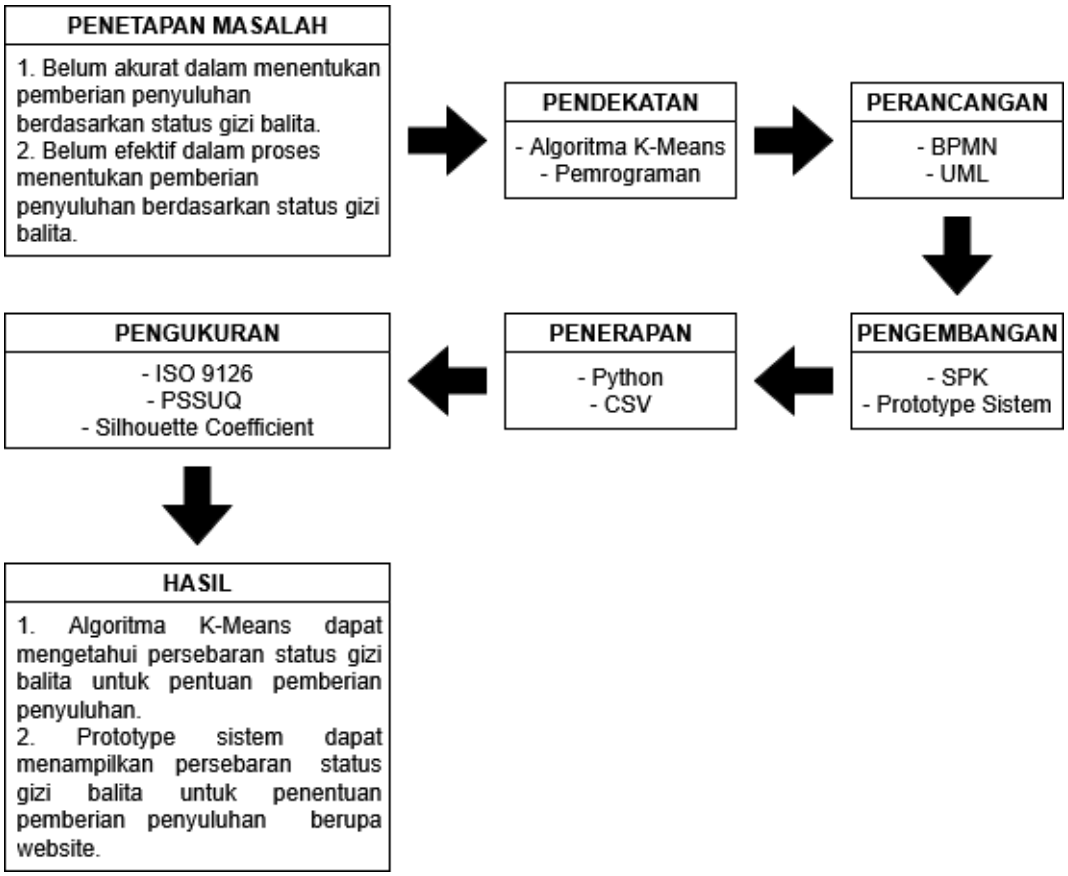
No	Peneliti/Tahun	Judul Penelitian	Sumber Jurnal	Kontribusi
			MS.2024.0703, E-ISSN: 1818-0523, P-ISSN: 1816-6075 https://www.aasmr.org/jsms/Vol14/No.7/Vol.14.No.7.03.pdf	atribut seperti TB/U, BB/U, dan TB/BB
10.	Ita Dewi Sintawati, Widiarina, Kartika Mariskhana (2024)	Analysis of Malnutrition Status in Toddlers Using the K-Means Algorithm Case Study in DKI Jakarta Province	Jurnal dan Penelitian Teknik Informatika, Vol. 8, No. 4, Oktober 2024,DOI:10.33395/sinkron.v8i4.14087, E-ISSN:2541-2019, P-ISSN: 2541-044X https://jurnal.polgan.ac.id/index.php/sinkron/article/view/14087	Berkontribusi pada penerapan data *mining* dengan menggunakan Algoritma K-Means *Clustering*

Berdasarkan tinjauan studi yang sudah dipaparkan pada Tabel 2.1 didapatkan pengetahuan yang dijadikan rujukan dalam pelaksanaan penelitian ini. Penelitian yang akan dilakukan berfokus pada penggunaan metode K-Means untuk memetakan persebaran status gizi balita untuk penentuan pemberian penyuluhan agar lebih tepat sasaran. Orisinalitas penelitian terdapat pada penggunaan variabel usia, berat badan, tinggi badan, z score bb/u, z score tb/u dan z score bb/tb. Kontribusi baru dalam penelitian ini yaitu sistem informasi dengan metode K-Means untuk memetakan persebaran status gizi balita untuk penentuan pemberian penyuluhan agar lebih tepat sasaran di masing-masing posyandu. Kontribusi dalam rujukan penelitian ini memberikan pengetahuan ilmu mengenai permasalahan pada persebaran status gizi balita untuk penentuan pemberian penyuluhan dengan proses perhitungan algoritma K-Means. Penelitian dilakukan karena ada pengembangan dari penelitian sebelumnya dengan menggunakan algoritma K-Means yang akan diuji dengan *Silhouette Coefficient* untuk mengevaluasi kualitas hasil *clustering.* Salah satu jurnal yang menjadi rujukan paling penting sekaligus dasar pemikiran dalam penelitian ini adalah jurnal berjudul **"*Clustering* Data Antropometri Balita untuk Menentukan Status Gizi Balita di Kelurahan Jumput**

Rejo Sukodono Sidoarjo". Jurnal tersebut sangat relevan karena membahas pendekatan K-Means dalam konteks status gizi balita, khususnya dalam menentukan status gizi balita. Relevansi tersebut memberikan landasan teoritis sekaligus menjadi alasan utama peneliti untuk mengembangkan pendekatan serupa, namun dengan fokus berbeda yaitu pada pemetaan persebaran status gizi balita untuk penentuan pemberian penyuluhan.

C. Kerangka Berpikir

Kerangka pemikiran adalah rancangan yang menjadi dasar pemikiran dalam sebuah penelitian. Kerangka ini dirancang untuk mewakili konsep pemecahan masalah yang mencakup objek penelitian dan metode yang digunakan. Gambar 2. 6 merupakan kerangka pemikiran yang digunakan untuk memecahkan masalah dalam penelitian ini, sebagai berikut:



Gambar 2.6 Kerangka Pemikiran

Penjelasan kerangka pemikiran pada Gambar 2.6 yaitu, sebagai berikut:

(1) permasalahan yang terjadi pada penelitian ini yaitu, belum akurat dalam penentuan pemberian penyuluhan berdasarkan status gizi dan belum efektif dalam proses menentukan pemberian penyuluhan berdasarkan status gizi balita;
(2) pendekatan penelitian dilakukan untuk menemukan solusi dari permasalahan yang ada, sehingga dilakukan dengan menggunakan algoritma K-Means dan pemrograman;
(3) tahap pengembangan dilakukan dengan membangun sistem pendukung keputusan dan membuat prototype sistem;
(4) penerapan penelitian dilaksanakan melalui penggunaan sistem pemrograman *Python* dan file CSV sebagai sumber data yang digunakan;
(5) penerapan penelitian juga dilakukan dengan menguji menggunakan *ISO 9126*, PSSUQ, dan *Silhouette Coefficient*;
(6) hasil dari penelitian dan pengembangan ini menunjukkan bahwa algoritma K-Means dapat mengetahui persebaran gizi kurang dan gizi lebih berdasarkan status gizi balita untuk penentuan pemberian penyuluhan;
(7) hasil dari penelitian dan pengembangan ini prototype sistem berupa website yang dapat menampilkan persebaran status gizi.

D. Hipotesis

Algoritma K-Means merupakan algoritma dalam data *mining* yang digunakan untuk mengelompokkan data ke dalam beberapa kelompok (*cluster*) secara sederhana dan cepat berdasarkan kesamaan karakteristik dan data yang mirip akan dikelompokkan ke dalam *cluster* yang sama. Merujuk pada penelitian sebelumnya yang juga menggunakan algoritma K-Means menunjukkan tingkat keberhasilan sebesar 79%**,** dengan pengujian yang digunakan *Silhouette Coefficient*. Algoritma K-Means bekerja dengan cara mengelompokkan data, maka pada penelitian ini juga berkaitan dengan pengelompokan data tentang pemetaan persebaran status gizi balita untuk penentuan pemberian penyuluhan. Karena pemantauan status gizi balita sangat penting untuk mendeteksi masalah gizi sejak dini dan memberikan intervensi yang tepat. Berdasarkan hal tersebut, hipotesis penelitian ini menyatakan bahwa penerapan algoritma K-Means dalam sistem pendukung keputusan (SPK) diduga akurat dan efektif untuk memetakan persebaran status gizi balita, sehingga penentuan pemberian penyuluhan menjadi lebih tepat sasaran.