

BAB II

KERANGKA TEORITIS

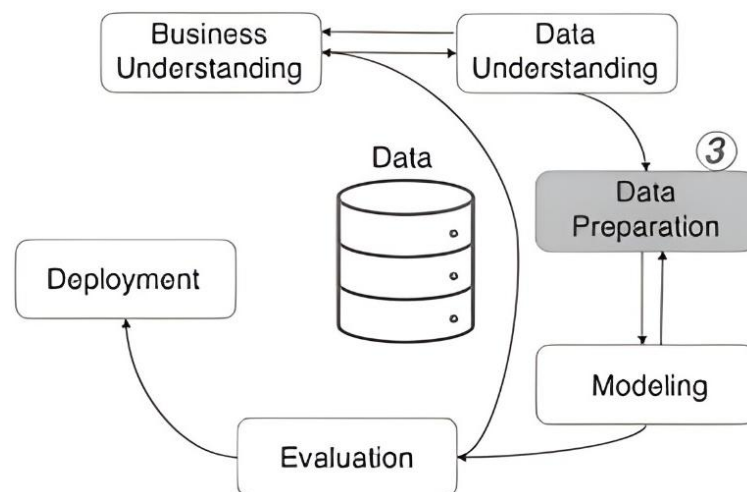
A. Landasan Teori

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

Data mining merupakan proses menggali pola atau pengetahuan dari kumpulan data berukuran besar untuk memperoleh informasi yang bermanfaat, menarik, bersifat non-trivial, implisit, serta sebelumnya belum diketahui namun berpotensi memiliki nilai guna (Urva et al., 2023, p. 6). Data yang dipandang sebagai objek informasi harus memiliki aturan atau minimal keterangan agar pengguna dapat menggunakan sesuai dengan peran objek data.

Salah satu standar proses penanganan data adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM) dengan mengimplementasikan enam tahapan pengelolaan data sebagai berikut (Urva et al., 2023, pp. 27–29):

- (1) *business understanding*: lingkup pemahaman kerangka bisnis mendasari dari analisis sebuah data;
- (2) *data understanding*: pengguna harus memahami substansi data sebagai bekal analisis. data yang memiliki relasi atribut dan nilai harus diterjemahkan agar antar entitas dapat memiliki benang merah;
- (3) *data preparation*: pengguna harus melakukan seleksi, pembersihan, konstruksi, integrasi, dan penyajian format data;
- (4) *modeling*: dilakukan *fitting* data atau pendekatan model dari data;
- (5) *evaluation*: pengujian ketepatan pemilihan model;
- (6) *deployment*: karakteristik yang dibangun harus dapat mengakomodir dari kebutuhan.



Gambar 2.1 Konsep CRISP-DM

Sumber: Fortino, 2023 (Urva et al., 2023, p. 28)

2. K-Nearest Neighbor (KNN)

Klasifikasi merupakan teknik penting dalam data mining yang berfungsi untuk mengelompokkan data ke dalam kelas atau kategori tertentu sesuai atribut yang dimilikinya. Tujuan utama teknik ini adalah membangun model yang mampu memprediksi label atau kategori dari data baru berdasarkan pola yang telah dipelajari dari data latih. Melalui analisis terhadap atribut yang memengaruhi proses pengelompokan, dapat diperoleh pemahaman mengenai faktor-faktor yang menentukan hasil klasifikasi (Setiawan et al., 2023, p. 6). Salah satu algoritma klasifikasi adalah *K-Nearest Neighbor* (KNN).

KNN adalah algoritma klasifikasi yang bekerja dengan memanfaatkan data berlabel sebagai dasar pembelajaran. Algoritma ini juga digunakan untuk menghitung probabilitas setiap kelas dari suatu data testing yang dipilih. Prosesnya dilakukan dengan mengukur jarak kesamaan antara data testing dengan data training. Setiap hasil perhitungan jarak kemudian dikaitkan dengan kelas asli dari data training yang terlibat dalam perhitungan tersebut. Selanjutnya ditentukan nilai K, yaitu parameter pada algoritma KNN yang digunakan dalam penelitian ini. Dari K data tersebut dihitung probabilitas kemunculan masing-masing kelas, dan kelas dengan probabilitas tertinggi dipilih sebagai prediksi untuk data testing (Wiliani & Fathurrahman, 2024, pp. 31–32).

Karena klasifikasi membutuhkan hasil akhir yang bersifat kategorikal, maka nilai probabilitas dari model dikonversi ke dalam bentuk biner (0 atau 1) menggunakan ambang batas (*threshold*), yang biasanya default-nya adalah 0,5. Jadi jika probabilitas > 0,5 maka dikategorikan sebagai 1. Dan jika probabilitas < 0,5 maka dikategorikan sebagai 0 (Az-zahrani et al., 2025). Sementara itu, untuk menghitung nilai jarak antara titik *data training* dan *data testing*, maka dapat digunakan rumus *Euclidean Distance* seperti berikut (Urva et al., 2023, pp. 20–21).

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \dots\dots\dots (2.1)$$

Keterangan:

d (x, y) = jarak

n = dimensi data

i = variabel data

x_i = sampel data

y_i = *data testing*

Berikut ini adalah langkah-langkah atau alur kerja KNN (Setiawan et al., 2023, p. 72):

- (1) menentukan parameter atau nilai k (jumlah ketetanggaan terdekat);
- (2) menghitung jarak objek baru terhadap *data training* yang diberikan atau jumlah k terdekat;
- (3) mengurutkan hasil pada langkah ke-2 secara *ascending*;
- (4) menetapkan kelas atau kategori ketetanggaan terdekat berdasarkan nilai k terdekat.

Contoh penerapan algoritma KNN (Setiawan et al., 2023, pp. 73–75) tentang klasifikasi kelas apakah A atau B (*dependent variables*), dengan objek *training* menggunakan 2 atribut yaitu usia dan pendapatan (*independent variables*) dimana ada 6 data yang sudah berlabel dan 1 data baru yang harus ditentukan kelasnya dari 2 kelas yaitu A dan B:

Usia	Pendapatan (Jutaan)	Kelas
38	4	A
34	5	A
34	6	B
37	7	A
35	6	B
37	5	A
33	8	?

langkah-langkah penyelesaiannya, yaitu:

- (1) menentukan nilai k, yaitu k=3;
- (2) menghitung jarak objek baru dengan *data training*, pada kasus ini menggunakan *euclidean*;

Usia	Pendapatan	Kelas	Jarak ke Data Baru (Usia = 33 dan Pendapatan = 8)
38	4	A	$= (38-33)^2 - (4-8)^2 = 41$
34	5	A	$= (34-33)^2 - (5-8)^2 = 10$
34	6	B	$= (34-33)^2 - (6-8)^2 = 5$
37	7	A	$= (37-33)^2 - (7-8)^2 = 17$
35	6	B	$= (35-33)^2 - (6-8)^2 = 8$
37	5	A	$= (37-33)^2 - (5-8)^2 = 7$

- (3) mengurutkan hasil pada langkah ke-2 secara *ascending*;

Usia	Pendapatan	Kelas	Jarak ke Data Baru (Usia = 33 dan Pendapatan = 8)	Perankingan
38	4	A	$= (38-33)^2 - (4-8)^2 = 41$	6
34	5	A	$= (34-33)^2 - (5-8)^2 = 10$	4
34	6	B	$= (34-33)^2 - (6-8)^2 = 5$	1
37	7	A	$= (37-33)^2 - (7-8)^2 = 17$	5
35	6	B	$= (35-33)^2 - (6-8)^2 = 8$	3
37	5	A	$= (37-33)^2 - (5-8)^2 = 7$	2

- (4) menetapkan kelas atau kategori ketetanggaan terdekat berdasarkan nilai K terdekat.

Usia	Pendapatan	Kelas	Jarak ke Data Baru (Usia = 33 dan Pendapatan = 8)	Perankingan	Klasifikasi
38	4	A	$= (38-33)^2 - (4-8)^2 = 41$	6	-
34	5	A	$= (34-33)^2 - (5-8)^2 = 10$	4	-
34	6	B	$= (34-33)^2 - (6-8)^2 = 5$	1	B
37	7	A	$= (37-33)^2 - (7-8)^2 = 17$	5	-
35	6	B	$= (35-33)^2 - (6-8)^2 = 8$	3	B
37	5	A	$= (37-33)^2 - (5-8)^2 = 7$	2	A

Sehingga diperoleh kelas B = 2 dan kelas A = 1, dengan *voting* maka diperoleh bahwa untuk data baru dengan usia 37 dan pendapatan 5 masuk kedalam kelas B.

3. Evaluasi Model Klasifikasi

Evaluasi model klasifikasi bertujuan untuk menunjukkan sejauh mana model mampu bekerja dengan baik dalam melakukan prediksi atau pengelompokan (klasifikasi) terhadap data yang tersedia (Fajri et al., 2024, p. 83). Salah satu metrik yang banyak digunakan untuk mengevaluasi model klasifikasi adalah *confusion matrix*.

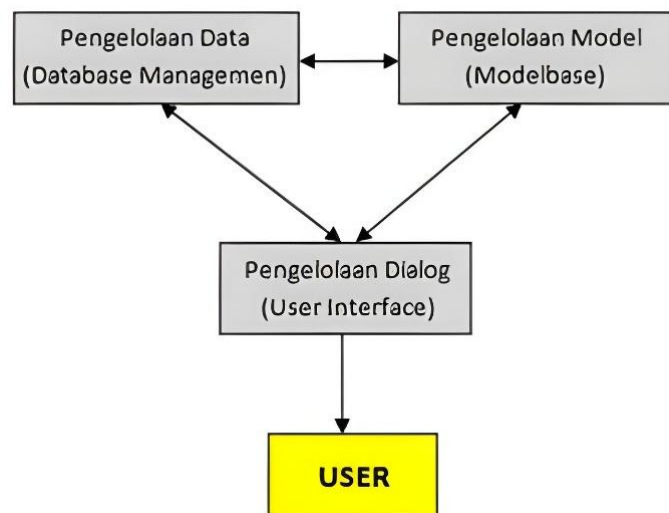
Confusion matrix merupakan metode evaluasi yang digunakan untuk menilai kinerja model klasifikasi dengan menampilkan jumlah prediksi yang benar maupun salah berdasarkan perbandingan dengan nilai aktual. Komponen dari *confusion matrix* meliputi (Fajri et al., 2024, p. 84):

- (a) *true positive* (TP): jumlah prediksi positif yang benar.
- (b) *true negative* (TN): jumlah prediksi negatif yang benar.
- (c) *false positive* (FP): jumlah prediksi positif yang salah.
- (d) *false negative* (FN): jumlah prediksi negatif yang salah.

Namun pada kasus data yang tidak seimbang, di mana jumlah sampel kelas tertentu jauh lebih banyak daripada kelas lainnya, teknik evaluasi model klasifikasi yang lebih tepat salah satunya adalah *balanced accuracy*. *Balanced accuracy* adalah rata-rata dari sensitivitas (*recall* untuk kelas positif) dan spesifisitas (*recall* untuk kelas negatif). Ini memberikan gambaran yang lebih akurat tentang kinerja model klasifikasi pada data yang tidak seimbang (Fajri et al., 2024, pp. 87–88).

4. Sistem Pendukung Keputusan (SPK)

Sistem Pendukung Keputusan (SPK) merupakan sebuah sistem interaktif yang dirancang untuk membantu proses pengambilan keputusan secara lebih tepat dengan memanfaatkan alternatif yang dihasilkan melalui perancangan model, pengolahan data, serta penyajian informasi (Aldo & Putra, 2020, p. 7). SPK terdiri atas tiga subsistem utama yang menentukan kemampuan teknisnya, yaitu subsistem pengelolaan data (*database*), subsistem pengelolaan model (*model base*), dan subsistem pengelolaan dialog (*user interface*) (Aldo & Putra, 2020, pp. 9–10).



Gambar 2.2 Komponen Sistem Pendukung Keputusan (SPK)

Sumber: (Aldo & Putra, 2020, p. 10)

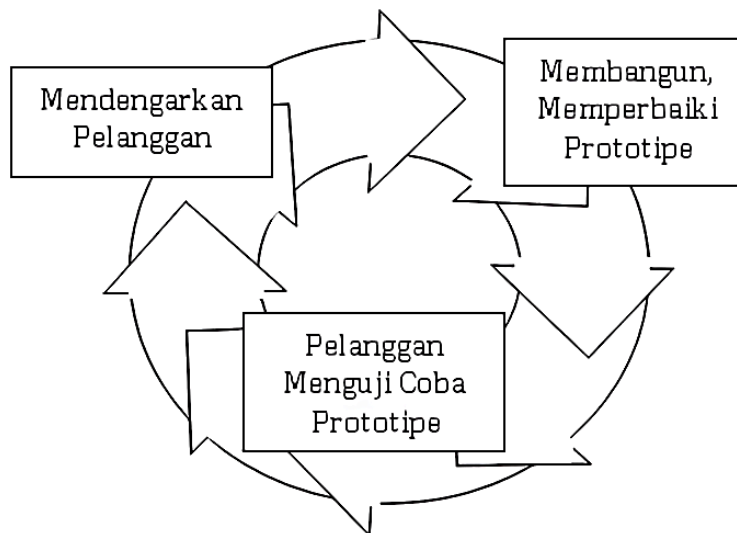
Dari hubungan tiga komponen sebagaimana yang ditunjukkan pada gambar 2.2, dapat dijelaskan sebagai berikut (Aldo & Putra, 2020, pp. 10–13):

- subsistem pengelolaan data (*database*), berfungsi sebagai menyediakan data-data yang diperlukan oleh *Database Management System* (DBMS).
- subsistem pengelolaan model (*model base*), memungkinkan pengambil keputusan menganalisis secara utuh dengan mengembangkan dan membandingkan alternatif solusi.
- subsistem pengelolaan dialog (*user interface*), komponen ini dibangun untuk memenuhi kebutuhan representasi dan mekanisme *control* selama proses analisis dalam SPK ditentukan dari kemampuan berinteraksi antara sistem yang terpasang dengan *user*.

5. *Prototype Software Development*

Prototype merupakan salah satu jenis model *Software Development Life Cycle* (SDLC) yang banyak digunakan. Model ini digunakan ketika pengguna

belum sepenuhnya mengetahui kebutuhan sistem secara detail sejak awal. Pada pendekatan ini, *prototype* dari produk akhir terlebih dahulu dibuat, kemudian diuji dan diperbaiki secara berulang berdasarkan masukan pengguna hingga diperoleh *prototype* akhir yang dapat diterima dan dijadikan acuan dalam pengembangan produk sesungguhnya (Hanifa et al., 2024, pp. 32–33).



Gambar 2.3 *Prototype Software Development*

Sumber: (Hanifa et al., 2024, p. 33)

6. **Business Process Modeling Notation (BPMN)**

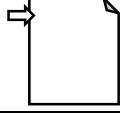
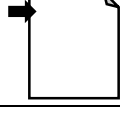
Business Process Modeling Notation (BPMN) merupakan standar pemodelan proses bisnis yang menyediakan simbol-simbol grafis untuk menggambarkan alur proses dalam bentuk diagram, sehingga dapat dipahami oleh seluruh pemangku kepentingan (*stakeholder*) dengan 5 kategori dasar notasi, yaitu (Wardani et al., 2021, p. 49-53):

flow objects, terdiri dari *event*, *activity*, dan *gateway*, sebagaimana ditunjukkan pada gambar 2.4.

Nama	Keterangan	Simbol
<i>Event</i>	Menunjukkan proses dimulai	
<i>Activity</i>	Menunjukkan kegiatan	
<i>Gateway</i>	Menunjukkan percabangan (pengambilan keputusan)	

Gambar 2.4 Notasi *Flow Objects* BPMN

data, terdiri dari *data object*, *data input* dan *data output*, sebagaimana ditunjukkan pada gambar 2.5.

Nama	Keterangan	Simbol
<i>Data Object</i>	Menunjukkan data yang dibutuhkan sebuah aktivitas	
<i>Data Input</i>	Menunjukkan data masukan yang dibutuhkan sebuah aktivitas	
<i>Data Output</i>	Menunjukkan data keluaran yang dihasilkan dari sebuah aktivitas	

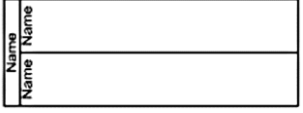
Gambar 2.5 Notasi Data BPMN

connecting objects, notasi yang dapat digunakan *flow objects* untuk terhubung satu sama lain berupa *sequence flow*, *message flow*, dan *association*, sebagaimana ditunjukkan pada gambar 2.6.

Nama	Keterangan	Simbol
<i>Sequence Flow</i>	Menunjukkan alur urutan aktivitas	
<i>Message Flow</i>	Menunjukkan alur pesan antar 2 partisipan atau antar pool	
<i>Association</i>	Menghubungkan artifak	

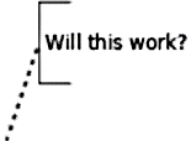
Gambar 2.6 Notasi *Connecting Objects* BPMN

swimlanes, terdiri dari *pool* untuk memisahkan organisasi yang berbeda dan *lane* untuk mengatur dan mengkategorikan kegiatan *swimlane* menurut fungsi atau peran, sebagaimana ditunjukkan pada gambar 2.7.

Nama	Keterangan	Simbol
<i>Pool</i>	Menggambarkan partisipan pada proses kolaborasi	
<i>Lane</i>	Mengelola dan mengkategorikan aktivitas	

Gambar 2.7 Notasi *Swimlanes* BPMN

artifacts, notasi untuk memberikan informasi yang lebih jelas kepada *stakeholder* agar lebih mudah dibaca dan dipahami, yang terdiri dari *group* dan *text annotation*, sebagaimana ditunjukkan pada gambar 2.8.

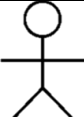
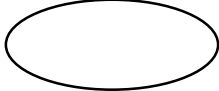
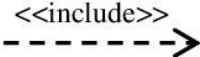
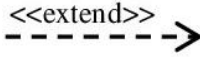
Nama	Keterangan	Simbol
<i>Group</i>	Memberi tanda untuk aktifitas yang bisa dikelompokkan	
<i>Text Annotation</i>	Menyediakan informasi tambahan berupa teks dari sebuah aktivitas	

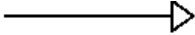
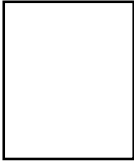
Gambar 2.8 Notasi *Artifacts* BPMN

7. Unified Modeling Language (UML)

Unified Modeling Language (UML) adalah sebuah bahasa pemodelan yang digunakan untuk menspesifikasikan, memvisualisasikan, membangun, serta mendokumentasikan berbagai aspek informasi yang dihasilkan dalam proses pengembangan perangkat lunak, baik untuk sistem perangkat lunak maupun sistem non-perangkat lunak lainnya (Habibi et al., 2020, p. 79). Ada beberapa diagram yang digunakan untuk proses tersebut, yaitu (Niqotaini et al., 2023, pp. 1–70):

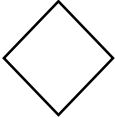
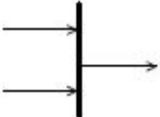
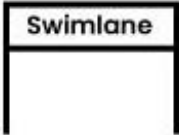
use case diagram, menggambarkan apa yang dilakukan pada sistem dan siapa yang dapat menerapkan fungsi dalam sistem tersebut (interaksi yang terjadi antara sistem dengan *user*), sebagaimana simbolnya ditunjukkan pada gambar 2.9.

Nama	Keterangan	Simbol
<i>Actor</i>	Orang, sistem atau proses ketika berinteraksi dengan <i>use case</i> atau sistem	
<i>Use case</i>	Fungsionalitas atau aksi yang ditampilkan sistem	
<i>Include</i>	Menghubungkan <i>use case</i> tambahan menuju <i>use case</i> yang ditambahkan. <i>Use case</i> yang ditambah akan membutuhkan <i>use case</i> ini agar dapat menjalankan fungsinya dalam <i>use case</i> tersebut. Arah panah <i>include</i> menuju ke <i>use case</i> tambahan	
<i>Extend</i>	Menghubungkan <i>use case</i> tambahan menuju <i>use case</i> yang ditambahkan. <i>Use case</i> yang ditambah dapat berdiri sendiri walaupun tanpa <i>use case</i> tambahan. Arah panah <i>extend</i> menuju ke <i>use case</i> tambahan	
<i>Association</i>	Menghubungkan antar <i>actor</i> dan <i>use case</i> yang saling berinteraksi	

Nama	Keterangan	Simbol
<i>Generalization</i>	Relasi umum dan tertentu antara dua <i>use case</i> yang salah satunya mempunyai fungsi global jika dibandingkan <i>use case</i> lainnya. Arah panah mengarah ke <i>use case</i> dimana dapat menjadi umum atau generalisasinya	
Sistem	Menspesifikasikan paket yang menampilkan sistem secara terbatas	

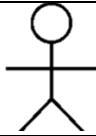
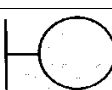
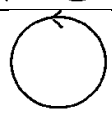
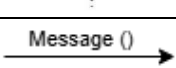
Gambar 2.9 Simbol *Use Case Diagram*

activity diagram, menggambarkan aliran kerja dengan menunjukkan urutan aktivitas seperti apa proses yang terjadi pada suatu sistem, sebagaimana simbolnya ditunjukkan pada gambar 2.10.

Nama	Keterangan	Simbol
<i>Initial State</i>	Status awal pada <i>activity diagram</i>	
<i>Activity</i>	Aktivitas atau kegiatan yang terjadi, diisi dengan kata kerja	
<i>Decision</i>	Proses memilih aktivitas lebih dari satu	
<i>Join</i>	Lebih dari satu aktivitas yang di gabungkan menjadi satu	
<i>Final State</i>	Status akhir yang menandakan bahwa urutan aktivitas telah selesai	
<i>Swimlane</i>	Pemisahan organisasi yang bertanggung jawab dalam aktivitas. Pengelompokkan aktivitas didasarkan oleh aktivitas <i>actor</i> dalam sebuah urutan yang sama	

Gambar 2.10 Simbol *Activity Diagram*

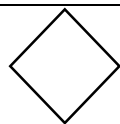
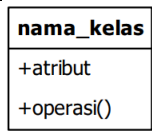
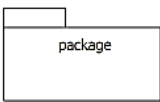
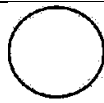
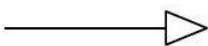
sequence diagram, menggambarkan logika sebuah fungsi atau operasi yang harus dijalankan pada sebuah sistem berdasarkan hubungan objek pada *use case*, sebagaimana simbolnya ditunjukkan pada gambar 2.11.

Nama	Keterangan	Simbol
<i>Actor</i>	Menggambarkan orang yang sedang berinteraksi dengan sistem	
<i>Entity Class</i>	Menggambarkan hubungan kegiatan yang akan dilakukan	
<i>Boundary Class</i>	Menggambarkan sebuah penggambaran dari form	
<i>Control Class</i>	Menggambarkan penghubung antara boundary dengan tabel	
<i>Focus Control & Life Line</i>	Menggambarkan tempat mulai dan berakhirnya sebuah message	
<i>Message</i>	Menggambarkan pengiriman pesan	

Gambar 2.11 Simbol *Sequence Diagram*

Sumber: (Habibi et al., 2020, p. 89)

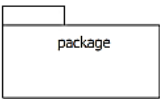
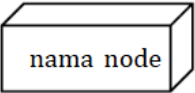
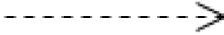
class diagram, menggambarkan kelas dan paket-paket yang berada dalam sebuah perangkat lunak atau sistem secara luas, sebagaimana simbolnya ditunjukkan pada gambar 2.12.

Nama	Keterangan	Simbol
<i>Association</i>	Relasi dari objek satu dengan lainnya	
<i>Public Association</i>	Asosiasi lebih dari dua objek dapat dihindari	
<i>Class</i>	Himpunan objek-objek yang berbagi atribut dan operasi yang sama	
<i>Package</i>	Bungkusan dari satu atau lebih kelas	
<i>Direct Association</i>	Relasi antar kelas dengan makna kelas yang satu digunakan oleh kelas yang lain	
<i>Interface</i>	Sama dengan konsep <i>interface</i> dalam pemrograman berorientasi objek	
<i>Generalization</i>	Relasi antar kelas dengan makna umum-khusus	

Nama	Keterangan	Simbol
<i>Dependency</i>	Relasi antar kelas dengan makna kebergantungan antar kelas	
<i>Aggregation</i>	Relasi antar kelas dengan makna semua bagian (<i>whole part</i>)	

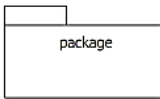
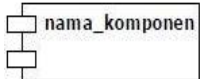
Gambar 2.12 Simbol *Class Diagram*

deployment diagram, menggambarkan hubungan antara perangkat lunak dan keras secara lengkap, sebagaimana simbolnya ditunjukkan pada gambar 2.13.

Nama	Keterangan	Simbol
<i>Package</i>	Alokasi <i>node</i>	
<i>Node</i>	Menunjukkan <i>hardware</i> tentang <i>software</i> yang tidak dibuat sendiri pada elemen yang memiliki fungsi terhadap perancangan, maka komponen memiliki kesamaan definisi dengan sebelumnya pada diagram komponen	
<i>Dependency</i>	Lambang hubungan antar elemen. Anak panah memberikan petunjuk tentang elemen yang dipakai	
<i>Link</i>	Hubungan antara <i>node</i>	

Gambar 2.13 Simbol *Deployment Diagram*

component diagram, menggambarkan komponen perangkat lunak dalam sistem dan keterkaitannya dalam sistem secara keseluruhan, sebagaimana simbolnya ditunjukkan pada gambar 2.14.

Nama	Keterangan	Simbol
<i>Package</i>	Letak komponen	
<i>Component System</i>	<i>Hardware</i> /objek pada sistem	
<i>Dependency</i>	Menggambarkan ketergantungan antar komponen. Arah panah berarti komponen yang dipakai	
<i>Interface</i>	Menggambarkan <i>interface</i> yang berguna agar tidak langsung menuju objek	
<i>Link</i>	Sebagai arah relasi antar komponen yang terhubung	

Gambar 2.14 Simbol *Component Diagram*

8. Pemrograman

Pemrograman adalah serangkaian perintah atau instruksi yang ditulis dalam bahasa yang dapat dipahami oleh komputer, seperti bahasa mesin atau biner, di mana komputer hanya mengenali angka 0 dan 1, kemudian diimplementasikan ke dalam perangkat komputer untuk melakukan tugas tertentu (Elfaladonna & Isa, 2024, p. 10). Melalui pemrograman, komputer dapat dipandu untuk menyelesaikan berbagai tugas atau proses secara cepat dan tepat sesuai perintah atau instruksi yang diberikan.

PHP (*Hypertext Preprocessor*) adalah bahasa pemrograman berbasis skrip yang berjalan di sisi server, di mana eksekusi kode dilakukan pada server dan hasilnya dikirim untuk ditampilkan melalui *browser*. PHP bisa diartikan sebagai bahasa pemrograman pelengkap yang digunakan untuk meningkatkan penggunaan tatanan HTML dan CSS untuk menghasilkan sebuah web yang dinamis dan interaktif (Fauzan & Nurhidayah, 2020, pp. 1–3).

Selain itu, ada juga bahasa pemrograman yang populer dan banyak digunakan, yaitu Python. Python merupakan bahasa pemrograman tingkat tinggi yang dapat mengeksekusi berbagai instruksi secara interpretatif dengan pendekatan berorientasi objek (OOP), serta memanfaatkan semantik dinamis untuk meningkatkan keterbacaan sintaks (Rahmadani et al., 2020, p. 10).

PHP digunakan dalam penelitian ini karena merupakan bahasa pemrograman *open source* yang cocok untuk pengembangan aplikasi berbasis web dan mudah diintegrasikan dengan *database* seperti MySQL untuk mengelola data pasien diabetes, khususnya diabetes melitus tipe 2 (DMT2). Sementara Python digunakan dalam penelitian ini karena mempermudah perhitungan dalam numerik dan memungkinkan implementasi algoritma KNN secara efisien.

9. Database

Dengan perkembangan teknologi, *database* telah berkembang dari sekadar penyimpanan data sederhana menjadi sistem yang kompleks untuk mendukung pengambilan keputusan, analisis data, dan otomatisasi proses bisnis. *Database* (basis data) didefinisikan sebagai kumpulan data yang terorganisir, disimpan, dan dikelola secara sistematis agar dapat diakses, diperbarui, dan dikelola dengan mudah (Usman et al., 2025, p. 1).

10. Diabetes Melitus

Penelitian ini dilakukan untuk mengidentifikasi penanganan pasien diabetes. Diabetes ada beberapa jenis, salah satunya adalah diabetes melitus. Diabetes melitus sendiri lebih dikenal dengan nama “penyakit gula” atau “kencing

manis”, merupakan kondisi dimana kadar gula darah lebih dari normal yang terjadi karena kelainan metabolisme sehingga terjadi peningkatan kadar gula darah (hiperglikemia) (Andriani, 2023, p. 13). Berikut ini adalah klasifikasi diabetes melitus menurut International Diabetes Federation & American Diabetes Association, yaitu (Andriani, 2023, pp. 13–15):

- (a) Diabetes melitus tipe 1, diakibatkan tidak adanya produksi insulin, biasanya diderita anak-anak atau remaja.
- (b) Diabetes melitus tipe 2, disebabkan ketidakcukupan atau ketidakefektifan kerja insulin (penderita dewasa).
- (c) Diabetes melitus kehamilan, diakibatkan gangguan hormonal, terjadi pada wanita hamil trimester ke-2 atau ke-3.
- (d) Diabetes melitus tipe lain, disebabkan karena pemakaian obat, bahan kimia, dan penyakit lain, atau bisa juga disebabkan faktor di luar pankreas.

B. Tinjauan Studi

Beberapa penelitian rujukan yang digunakan untuk penelitian dan pengembangan ini yaitu:

- (1) **Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma K-Nearest Neighbor (KNN)** (Oktaviana et al., 2024)

Penelitian ini memanfaatkan data dari Puskesmas Mlati II yang mencakup lima variabel, yaitu usia, tekanan darah, BMI, gula darah puasa, dan riwayat diabetes. *Dataset* diolah menggunakan teknik *MinMax Normalization* dan dibagi menjadi data training serta data testing dengan metode *Stratified 5-fold CV*. Dengan menerapkan algoritma KNN menggunakan parameter *Manhattan Distance* dan nilai $K=13$, penelitian ini berhasil mencapai akurasi 88%, presisi 83,54%, *recall* 87,5%, dan *F1-score* 85,36%.

- (2) **Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Penyakit Diabetes Melitus: Studi Kasus : Warga Desa Jatitengah** (Fasnuari et al., 2022)

Penelitian ini membahas penerapan algoritma KNN untuk mengklasifikasi penyakit DM. Penelitian ini menggunakan delapan variabel, seperti usia, riwayat keluarga, hasil cek gula darah, mudah haus, dan berat badan yang turun meskipun pola makan teratur. Dengan 81 data pelatihan dan 54 data pengujian, serta penggunaan *Euclidean Distance* dan $K=9$, penelitian ini mencapai akurasi sebesar 93%, presisi 100%, *recall* 60%, dan *F1-score* 75%.

(3) **Penerapan Metode *K-Nearest Neighbor* Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus** (Lestari et al., 2021)

Penelitian ini membahas penerapan algoritma KNN dalam membangun sistem pendukung keputusan untuk identifikasi dini penyakit DM. Penelitian ini menggunakan delapan variabel, seperti jumlah kehamilan, kadar glukosa, tekanan darah, indeks massa tubuh (IMT), dan riwayat keluarga. Dengan nilai $K=23$, penelitian ini mencapai tingkat akurasi sebesar 96%, menunjukkan efektivitas algoritma KNN dalam mendukung diagnosa dini.

(4) **Klasifikasi Potensi Penyakit Diabetes Melitus Tipe II Pada Pasien Menggunakan Algoritme KNN (*K-Nearest Neighbor*)** (Fiqri & Bhakti, 2024)

Penelitian ini menunjukkan bahwa algoritma KNN dapat digunakan untuk mendeteksi potensi DMT2 pada pasien. Dalam studi tersebut, algoritma KNN diuji dengan nilai $K=1$ dan $K=3$, menghasilkan akurasi tertinggi sebesar 85% pada $K=1$, yang menunjukkan bahwa parameter K memengaruhi efektivitas klasifikasi. Proses yang dilakukan meliputi pengolahan *dataset*, *preprocessing*, klasifikasi, evaluasi dengan *Confusion Matrix*, prediksi, dan penyimpanan model.

(5) **Penerapan Algoritma *K-Nearest Neighbor* Untuk Prediksi Akurasi Penyakit Diabetes** (Amin & Tampubolon, 2024)

Penelitian ini membahas pengembangan model prediksi diabetes dengan algoritma KNN berdasarkan dataset Pima Indians Diabetes. Penelitian ini menyoroti pentingnya *preprocessing* data, seperti penanganan nilai nol, imputasi data hilang, dan normalisasi fitur, yang berkontribusi pada peningkatan kinerja model. Dengan nilai K optimal sebesar 19, model tersebut mencapai akurasi 75,32% dalam memprediksi diabetes, dengan presisi 0,67 dan *recall* 0,57 untuk kasus positif.

(6) ***Supervised Learning* Metode *K-Nearest Neighbor* Untuk Prediksi Diabetes Pada Wanita** (Silalahi et al., 2023)

Penelitian ini mengimplementasikan algoritma KNN untuk klasifikasi DM pada wanita dengan dua kelas, yaitu DM Positif dan DM Negatif. Penelitian ini menggunakan parameter seperti kadar glukosa, tekanan darah, dan indeks massa tubuh. Hasil pengujian yang menunjukkan akurasi hingga 84% dengan peningkatan presisi dan *recall* pada beberapa skenario, memberikan bukti kuat akan efektivitas KNN dalam mendiagnosis penyakit DM.

(7) ***K-Nearest Neighbor Method for Early Detection of Diabetes Patients Based on Symptoms and Clinical Data*** (Rikatsih et al., 2024)

Penelitian ini menerapkan algoritma KNN untuk deteksi dini DM menggunakan data gejala dan klinis menunjukkan bahwa KNN efektif dalam mengklasifikasikan pasien menjadi dua kategori, yaitu diabetes dan non-diabetes. Hasil penelitian ini menunjukkan akurasi sebesar 98,27% dengan menggunakan teknik *K-Fold Cross-Validation*. Penelitian ini juga menekankan pentingnya pemilihan nilai K yang tepat untuk menghindari masalah *overfitting* atau *underfitting*, yang dapat mempengaruhi keakuratan model.

(8) ***Classification Of Diabetes Mellitus Disease At Rato Ebu Hospital-Indonesia Using The K-Nearest Neighbors Method Based On Missing Value*** (Putro et al., 2024)

Penelitian ini mengenai klasifikasi DM menggunakan KNN di Rumah Sakit Syarifah Ambami Rato Ebu Bangkalan menunjukkan bahwa algoritma KNN efektif untuk mendeteksi DM berdasarkan data pasien yang memiliki nilai hilang. Penelitian ini menggunakan teknik *data mining* yang meliputi pembersihan dan normalisasi data, serta pemilihan variabel yang signifikan untuk klasifikasi. Hasil eksperimen menunjukkan bahwa penggunaan rasio pembagian data 90:10 menghasilkan akurasi tertinggi sebesar 83%, tingkat kesalahan 16,67%, dan nilai AUC tertinggi 0,7407.

(9) ***Application Of Machine Learning K-Nearest Neighbour Algorithm To Predict Diabetes*** (Chandra & Nasien, 2023)

Penelitian ini menunjukkan bahwa KNN memiliki kinerja yang baik dalam memprediksi diabetes, dengan akurasi mencapai 93,58% dan tingkat kesalahan prediksi yang rendah yaitu 6,4% menggunakan data yang meliputi kadar glukosa, tekanan darah, usia, jenis kelamin, dan variabel lainnya.

(10) ***Implementation of K-Nearest Neighbour (KNN) Algorithm and Random Forest Algorithm in Identifying Diabetes*** (Diranisha et al., 2024)

Penelitian ini membandingkan dua algoritma, KNN dan *Random Forest*, dalam mengidentifikasi diabetes menggunakan data Pima Indian Diabetes. Hasil penelitian menunjukkan bahwa meskipun algoritma *Random Forest* memiliki akurasi yang sedikit lebih tinggi (77%) dibandingkan KNN (73%), KNN tetap menunjukkan kinerja yang baik dengan *precision* 87%, *recall* 73%, dan *F1-score* 79% pada nilai K = 7.

Penelitian rujukan yang digunakan dalam penelitian ini dirangkum kedalam sebuah tabel berikut yang berisi informasi mengenai peneliti, tahun, judul penelitian, jurnal sumber, dan kontribusi.

Tabel 2.1 Penelitian Rujukan

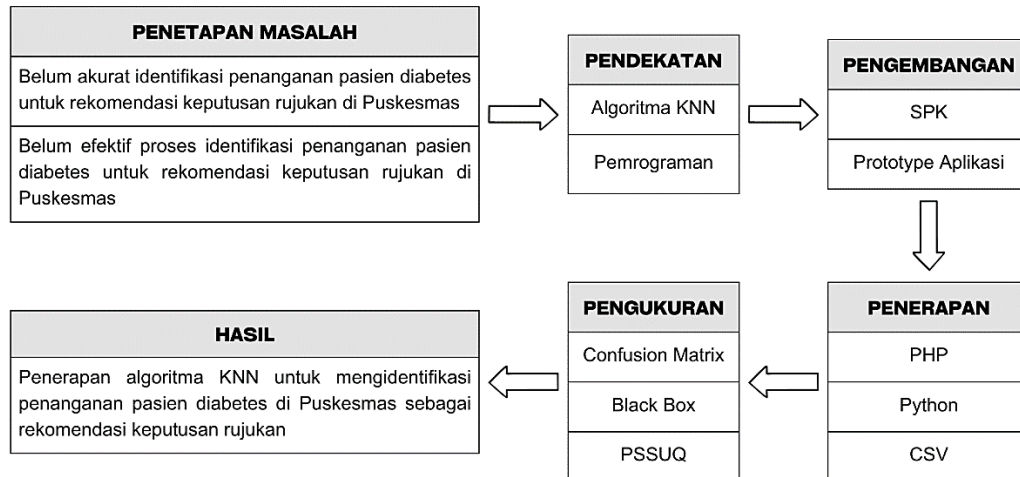
No	Peneliti /Tahun	Judul Penelitian	Jurnal Sumber	Kontribusi
1	Agfa Oktaviana, Dhina Puspasari Wijaya, Andri Pramuntadi, Dadang Heksaputra / 2024	Prediksi Penyakit Diabetes Melitus Tipe 2 Menggunakan Algoritma <i>K-Nearest Neighbor</i> (KNN)	MALCOM: <i>Indonesian Journal of Machine Learning and Computer</i> , Vol. 4 No. 3, Juli 2024, IRPI, DOI 10.57152 https://journal.irpi.or.id/index.php/malcom/article/view/1268/660	Kontribusi penelitian ini fokus pada identifikasi penanganan pasien diabetes untuk rekomendasi keputusan rujukan di Puskesmas menggunakan KNN dan implementasi berbasis web.
2	Happy Andrian Dwi Fasnuari, Haris Yuana, M. Taofik Chulkamdi / 2022	Penerapan Algoritma <i>K-Nearest Neighbor</i> Untuk Klasifikasi Penyakit Diabetes Melitus: Studi Kasus : Warga Desa Jatitengah	ANTIVIRUS: Jurnal Ilmiah Teknik Informatika, Vol. 16 No. 2, November 2022, UNISBA, DOI 10.35457 https://ejournal.unisbablitra.ac.id/index.php/antivirus/article/view/2445/1352	Kontribusi penelitian ini adalah memfokuskan pada identifikasi penanganan pasien diabetes yang sudah terdiagnosis, khususnya untuk rekomendasi keputusan rujukan di Puskesmas dengan pendekatan berbasis web.
3	Umikulsum Indah Lestari, Anis Yusrotun Nadhiroh, Cahyuni Nova / 2021	Penerapan Metode <i>K-Nearest Neighbor</i> Untuk Sistem Pendukung Keputusan Identifikasi Penyakit Diabetes Melitus	JATISI (Jurnal Teknik Informatika dan Sistem Informasi), Vol. 8 No. 4, Desember 2021, Universitas MDP, DOI 10.35957 https://jurnal.mdp.ac.id/index.php/jatisi/article/view/1235/509	Kontribusi penelitian ini adalah mengembangkan sistem yang membantu tenaga medis di Puskesmas dalam menentukan penanganan, khususnya rekomendasi keputusan rujukan pasien diabetes.
4	Mohammad Sholikhul Fiqri, Henny Dwi Bhakti / 2024	Klasifikasi Potensi Penyakit Diabetes Mellitus Tipe II Pada Pasien Menggunakan Algoritme KNN (<i>K-Nearest Neighbor</i>)	JATI (Jurnal Mahasiswa Teknik Informatika), Vol. 8 No. 4, Agustus 2024, Institut Teknologi Nasional Malang, DOI 10.36040 https://ejournal.itn.ac.id/index.php/jati/article/view/10133/5939	Kontribusi penelitian ini adalah menentukan penanganan pasien diabetes yang sudah terdiagnosis, khususnya rekomendasi keputusan rujukan di Puskesmas dengan sistem berbasis web.
5	Erika Tampubolon, Ruhul Amin / 2024	Penerapan Algoritma <i>K-Nearest Neighbor</i> Untuk Prediksi Akurasi Penyakit Diabetes	<i>Informatics and Digital Expert</i> (INDEX), Vol. 6 No. 1, Mei 2024, Universitas Perjuangan Tasikmalaya, DOI 10.36423 https://garuda.kemdikbud.go.id/documents/detail/4249215	Kontribusi penelitian ini adalah mengidentifikasi penanganan pasien diabetes di Puskesmas, khususnya dalam rekomendasi keputusan rujukan berdasarkan data pemeriksaan klinis.
6	Arina Prima Silalahi, Harlen Gilbert Simanullang, Marlyna Infryanty Hutapea / 2023	<i>Supervised Learning</i> Metode <i>K-Nearest Neighbor</i> Untuk Prediksi Diabetes Pada Wanita	METHOMIKA: Jurnal Manajemen Informatika & Komputerisasi Akuntansi, Vol. 7 No. 1, April 2023, Universitas Methodist Indonesia, DOI 10.46880 https://ejurnal.methodist.ac.id/index.php/methomika/article/view/1759/1307	Kontribusi penelitian ini adalah mengidentifikasi penanganan, khususnya untuk rekomendasi keputusan rujukan pasien di Puskesmas dengan memanfaatkan data klinis dan SPK medis.

No	Peneliti /Tahun	Judul Penelitian	Jurnal Sumber	Kontribusi
7	Nindynar Rikatsih, Mochammad Anshori, Risqy Siwi Pradini, Faurika / 2024	<i>K-Nearest Neighbor Method for Early Detection of Diabetes Patients Based on Symptoms and Clinical Data</i>	Inform: Jurnal Ilmiah Bidang Teknologi Informasi dan Komunikasi, Vol. 9 No. 2, Juli 2024, Universitas Dr. Soetomo, DOI 10.25139 https://ejournal.unitomo.ac.id/index.php/inform/article/view/8582/4155	Kontribusi penelitian ini adalah memanfaatkan data klinis pasien diabetes untuk menentukan penanganan, khususnya dalam rekomendasi keputusan rujukan di Puskesmas dengan pendekatan berbasis web.
8	Sigit Susanto Putro, Moh Abdan Syakura Putra, Doni Abdul Fatah, Yuli Panca Asmara, Hermawan Bin Fauzan, Eka Mala Sari Rochman, Aeri Rachmad / 2024	<i>Classification of diabetes mellitus disease at Rato Ebuh Hospital-Indonesia using the K-Nearest neighbors method based on missing value</i>	<i>BIO Web of Conferences</i> , Vol. 146, 2024, EDP Sciences, DOI 10.1051 https://www.bio-conferences.org/articles/bioconf/pdf/2024/65/bioconf_btmic2024_01081.pdf	Kontribusi penelitian ini adalah memberikan rekomendasi penanganan pasien diabetes, khususnya rekomendasi keputusan rujukan di Puskesmas dengan memanfaatkan data pemeriksaan klinis dan implementasi sistem berbasis web.
9	Jack Billie Chandra, Dewi Nasien / 2023	<i>Application Of Machine Learning K-Nearest Neighbour Algorithm To Predict Diabetes</i>	<i>International Journal of Electrical, Energy and Power System Engineering (IJEPPSE)</i> , Vol. 6 No. 2, Juni 2023, Universitas Riau, DOI 10.31258 https://ijeppse.id/journal/index.php/ijeppse/article/view/114/103	Kontribusi penelitian ini adalah mengidentifikasi penanganan pasien yang telah terdiagnosis diabetes, khususnya untuk rekomendasi keputusan rujukan di Puskesmas dengan pendekatan berbasis web.
10	Virly Diranisha, Agung Triayudi, Ratih Titi Komalasari / 2024	<i>Implementation of K-Nearest Neighbour (KNN) Algorithm and Random Forest Algorithm in Identifying Diabetes</i>	<i>SAGA: Journal of Technology and Information System</i> , Vol. 2 No. 2, Mei 2024, MDPI, DOI 10.58905 https://journal.mediadigitalpublikasi.com/index.php/saga/article/view/253/147	Kontribusi penelitian ini adalah memfokuskan penggunaan algoritma KNN untuk menentukan penanganan, terutama rekomendasi keputusan rujukan pasien diabetes di Puskesmas.

Berdasarkan penelitian rujukan pada tabel 2.1, dapat diketahui bahwa orisinalitas dan pembeda penelitian ini dengan penelitian sebelumnya terletak pada fokusnya terhadap pengambilan rekomendasi keputusan rujukan pasien diabetes di Puskesmas menggunakan KNN dengan probabilitas dan *threshold*, serta berkontribusi pada pengembangan *prototype* aplikasi berbasis web dengan data pasien diabetes, khususnya DMT2, menggunakan variabel yang lebih relevan berdasarkan indikasi rujukan pasien DMT2 dari PPK 1 ke PPK 2 menurut PNPK Tata Laksana DMT2 Dewasa, yaitu usia, tekanan darah sistol dan diastol, pemeriksaan gula, dan rujuk RS. Dengan demikian, penelitian ini memberikan pembaruan dari sisi penerapan, tujuan dan penggunaan KNN, serta bentuk implementasinya.

C. Kerangka Berfikir

Berdasarkan dukungan landasan teoritis yang diperoleh, permasalahan yang ada, dan penelitian-penelitian sebelumnya, berikut adalah kerangka yang diperoleh dalam eksplorasi teoritis yang menjadi referensi penelitian.



Gambar 2.15 Kerangka Berfikir

Kerangka berfikir pada gambar 2.15 dapat dijelaskan sebagai berikut:

- (1) Penetapan masalah penelitian ini yaitu belum akurat dan belum efektif dalam mengidentifikasi penanganan pasien diabetes untuk rekomendasi keputusan rujukan di Puskesmas;
- (2) Pendekatan penelitian ini menggunakan algoritma KNN dan pemrograman;
- (3) Pengembangan dilakukan dengan SPK dan *prototype* aplikasi;
- (4) Penerapan menggunakan bahasa pemrograman PHP dan Python serta file CSV;
- (5) Pengukuran pengujian dengan *Confusion Matrix*, *Black Box Testing*, dan PSSUQ;
- (6) Hasil yang ditargetkan adalah penerapan algoritma KNN untuk mengidentifikasi penanganan pasien diabetes di Puskesmas sebagai rekomendasi keputusan rujukan.

D. Hipotesis

K-Nearest Neighbor (KNN) adalah algoritma *machine learning* untuk klasifikasi menentukan label data baru berdasarkan mayoritas dari K tetangga terdekatnya. Beberapa penelitian dalam proses klasifikasi diperoleh rata-rata akurasi keberhasilannya yaitu 86,92%. Penelitian ini memilih motif pemecahan masalah yang relevan dengan sifat kerja KNN yaitu klasifikasi. Dengan pemahaman tersebut, maka dapat ditetapkan hipotesis dari penelitian dan pengembangan ini yaitu “penerapan KNN pada SPK diduga akurat dan efektif untuk mengidentifikasi penanganan pasien diabetes untuk rekomendasi keputusan rujukan di Puskesmas”.