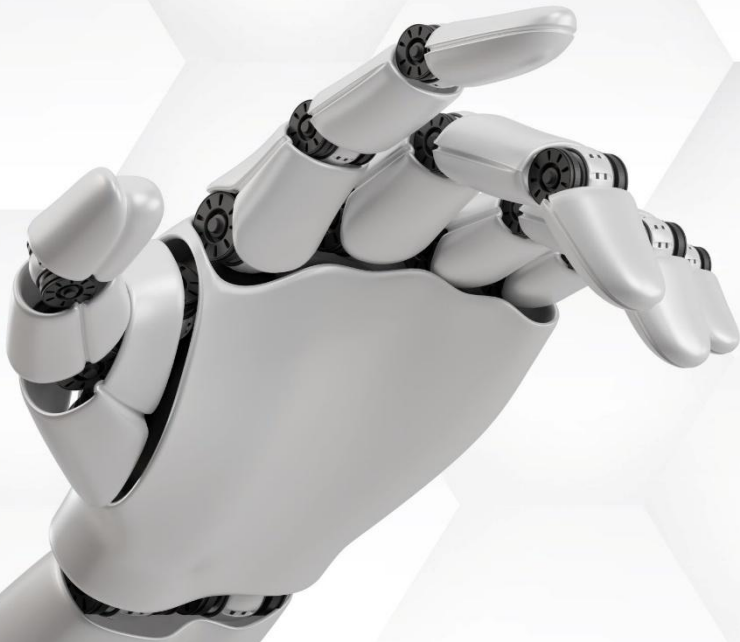


# **PREDIKSI RISIKO DIABETES DINI BERBASIS K-NEAREST NEIGHBORS (K-NN): STUDI IMPLEMENTASI PADA PLATFORM WEB**



**PENULIS :**

**FAHMI RUZIQ, S.T., M.KOM.  
M. RHIFKY WAYAHDI, S.KOM., M.KOM.**

# **PREDIKSI RISIKO DIABETES DINI BERBASIS K-NEAREST NEIGHBORS (K-NN): STUDI IMPLEMENTASI PADA PLATFORM WEB**

**Penulis**

**Fahmi Ruziq, S.T., M.Kom.**

**M. Rhifky Wayahdi, S.Kom., M.Kom.**

**PENERBIT:**



**HADLA**  
MEDIA INFORMASI

**PREDIKSI RISIKO DIABETES DINI BERBASIS K-NEAREST NEIGHBORS (K-NN):  
STUDI IMPLEMENTASI PADA PLATFORM WEB**

Tim Penulis:

**Fahmi Ruziq, S.T., M.Kom.  
M. Rhifky Wayahdi, S.Kom., M.Kom.**

Desain Cover:

**Sulaiman**

Tata Letak:

**Sulaiman**

ISBN:

-

Cetakan Pertama:

**Januari, 2026**

Hak Cipta 2026, Pada Penulis

---

Hak Cipta Dilindungi Oleh Undang-Undang

---

**Copyright © 2026**

**by HADLA Media Informasi**

All Right Reserved

**Dilarang keras menerjemahkan, memfotokopi, atau memperbanyak sebagian atau seluruh isi buku ini tanpa izin tertulis dari Penerbit.**

**PENERBIT:**



Website: [www.media.hadlacorp.com](http://www.media.hadlacorp.com)

# PRAKATA

Bismillahirrahmanirrahim

Alhamdulillah, dengan izin Allah SWT penulisan buku ini akhirnya dapat diselesaikan. Semoga kehadirannya membawa manfaat, baik bagi kalangan akademisi maupun masyarakat luas yang peduli pada kesehatan.

Sering buang air kecil di malam hari, rasa haus yang tak kunjung hilang, berat badan turun tanpa sebab, hingga luka kecil yang tak kunjung sembuh—semua itu bukan sekadar keluhan sepele. Ia adalah bahasa tubuh yang berusaha memberi tanda. Dalam penelitian ini, gejala seperti polyuria, polydipsia, polyphagia, visual blurring, obesitas, hingga iritabilitas dikaji lebih dalam dan dianalisis menggunakan pendekatan machine learning. Dari sana, kita dapat memahami bahwa tubuh senantiasa memberikan sinyal, hanya saja sering kali kita terlambat mendengarkannya.

Namun, bagaimana kita dapat mengenali risiko lebih dini, sebelum semuanya terlambat? Bagaimana gejala yang berserakan bisa dirangkai menjadi sebuah gambaran yang jelas? Di sinilah sains data mengambil peran. Melalui buku ini, saya mencoba menghadirkan sebuah pendekatan baru: memanfaatkan algoritma machine learning—khususnya metode K-Nearest Neighbors (KNN)—untuk membaca pola gejala dan memprediksi risiko diabetes dengan lebih objektif.

KNN, secara sederhana, bekerja dengan mencari kedekatan antar individu berdasarkan gejala yang dialami. Sama halnya seperti seorang dokter yang membandingkan kondisi pasien dengan kasus serupa sebelumnya, algoritma ini "bertanya" pada tetangga terdekatnya: apakah pola yang muncul lebih mirip penderita berisiko tinggi atau individu sehat? Dengan cara inilah, data yang semula abstrak dapat berubah menjadi wawasan yang bermakna.

Buku ini tidak hanya berhenti pada analisis data, penelitian ini diwujudkan dalam bentuk sebuah sistem berbasis website. Melalui sistem ini, siapa pun dapat memasukkan gejala yang dirasakan—seperti polyuria, polydipsia, polyphagia, obesitas, hingga delayed healing—dan mendapatkan prediksi risiko diabetes secara cepat. Dengan cara ini, teknologi machine learning tidak lagi hanya menjadi konsep di atas kertas, tetapi hadir nyata dalam kehidupan sehari-hari, memberi manfaat langsung bagi masyarakat.

Diabetes sendiri merupakan salah satu masalah kesehatan global yang jumlah penderitanya terus meningkat dari tahun ke tahun. Organisasi kesehatan dunia (WHO) mencatat bahwa penyakit ini menjadi salah satu penyebab utama kematian dini, menurunkan kualitas hidup, serta menambah beban biaya kesehatan baik bagi keluarga maupun negara. Fakta ini menunjukkan bahwa upaya deteksi dini menjadi sangat penting, agar risiko dapat dikenali lebih awal sebelum berkembang menjadi kondisi kronis yang sulit ditangani.

Oleh karena itu, buku ini diharapkan dapat memberikan kontribusi pada dua sisi: di ranah akademik sebagai referensi ilmiah tentang penerapan machine learning dalam bidang kesehatan, serta di ranah praktis sebagai gambaran bagaimana teknologi sederhana berbasis website dapat membantu masyarakat melakukan pemeriksaan mandiri. Saya berharap buku ini tidak hanya dibaca oleh para peneliti atau mahasiswa, tetapi juga oleh praktisi kesehatan, pengambil kebijakan, maupun masyarakat umum yang ingin memahami cara tubuh mereka berbicara melalui gejala yang dialami.

Penyusunan buku ini tentu tidak terlepas dari dukungan berbagai pihak. Untuk itu, saya ingin menyampaikan terima kasih yang tulus kepada keluarga, sahabat, rekan sejawat, serta para pembimbing yang senantiasa memberi masukan dan semangat. Tidak lupa, apresiasi juga saya sampaikan kepada setiap individu yang terlibat secara langsung maupun tidak langsung dalam penelitian ini, karena tanpa dukungan mereka buku ini tidak akan pernah terwujud.

Akhirnya, semoga buku ini tidak hanya menjadi bacaan, tetapi juga pengingat bahwa tubuh kita selalu berbicara—dan dengan bantuan sains, kita bisa belajar mendengarkannya lebih baik. Semoga buku ini dapat menjadi inspirasi untuk melanjutkan riset, memperluas pemahaman, serta mendorong kesadaran masyarakat dalam menjaga kesehatan. Selamat membaca, semoga bermanfaat.

Salam  
Penulis

PRAKATA.....	iii
DAFTAR ISI.....	v

BAB I PENDAHULUAN.....	1
1.1. Latar Belakang dan Urgensi Deteksi Dini Diabetes Melitus (DM).....	2
1.1.1. Skala Krisis Global dan Beban Ekonomi.....	3
1.1.2. Tantangan Deteksi dan Gejala Asimtomatik.....	3
1.1.3. Urgensi Deteksi Dini sebagai Intervensi Kunci.....	3
1.2. Proyeksi Peningkatan Prevalensi DM dan Beban Kesehatan Global.....	4
1.2.1. Tren Proyeksi Prevalensi yang Mengkhawatirkan.....	5
1.2.2. Beban Ekonomi dan Biaya Langsung ( <i>Direct Cost</i> ).....	5
1.2.3. Konsekuensi Sosial dan Dampak pada Kualitas Hidup.....	6
1.3. Komplikasi Serius DM: Kardiovaskular, Nefropati, dan Neuropati.....	6
1.3.1. Penyakit Kardiovaskular (Komplikasi Makrovaskular).....	6
1.3.2. Nefropati Diabetik dan Gagal Ginjal.....	7
1.3.3. Neuropati Diabetik dan Kerusakan Saraf.....	8
1.4. Peran Teknologi Informasi dan <i>E-Health</i> dalam <i>Screening</i> Kesehatan.....	8
1.4.1. Definisi dan Kontribusi <i>E-Health</i> .....	9
1.4.2. Platform Web sebagai Solusi <i>Screening</i> .....	9
1.4.3. Integrasi <i>Machine Learning</i> untuk Prediksi Risiko.....	10
1.5. Perumusan Masalah dan Tujuan Penulisan Monograf.....	10
1.5.1. Perumusan Masalah.....	11
1.5.2. Tujuan Penulisan Monograf.....	11
1.6. Kontribusi dan Sistematika Penulisan Monograf.....	12
1.6.1. Kontribusi Monograf.....	12
1.6.2. Sistematika Penulisan Monograf.....	12

BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI .....	14
2.1.    Konsep Prediksi Risiko Kesehatan berbasis Data.....	14
2.1.1.    Perbedaan antara Prediksi Risiko dan Diagnosis Tradisional.....	14
2.1.2.    Model Risiko dan Peran Data Feature .....	15
2.1.3.    Konsep <i>Triage</i> Digital .....	16
2.2.    Tinjauan Algoritma Klasifikasi dalam Machine Learning .....	16
2.2.1.    Klasifikasi <i>Supervised Learning</i> .....	17
2.2.2.    Pertimbangan Pemilihan Algoritma untuk <i>E-Health</i> Dini.....	18
2.3.    Landasan Teori K-Nearest Neighbors (K-NN) .....	18
2.3.1.    Prinsip Kerja Klasifikasi dan Pemilihan K .....	18
2.3.2.    Perhitungan Jarak: Fokus pada Euclidean Distance .....	19
2.3.3.    Keunggulan K-NN untuk <i>Dataset</i> Sempel dan Sederhana.....	20
2.4. <i>State-of-the-Art</i> Prediksi Diabetes: Komparasi K-NN dengan RF dan SVM .....	20
2.4.1.    Algoritma Alternatif dan Kinerja Prediktif.....	20
2.4.2.    Perbandingan Kinerja dan Trade-off.....	21
2.4.3.    Posisi Kontribusi Monograf.....	21
2.5.    Tinjauan Implementasi Machine Learning pada Arsitektur Web (PHP-ML) .....	22
2.5.1.    Paradigma Implementasi <i>Machine Learning</i> Berbasis Web.....	23
2.5.2.    Justifikasi Penggunaan PHP dan MySQL.....	23
2.5.3.    Peran Kunci PHP Machine Learning (PHP-ML).....	23
2.5.4.    Keterbatasan dan Implikasi PHP-ML .....	24
BAB 3 METODOLOGI PENGEMBANGAN SISTEM.....	25
3.1.    Pendekatan <i>Software Development Life Cycle</i> (SDLC) .....	25
3.2.    Justifikasi Pemilihan Model <i>Waterfall</i> .....	26
3.3.    Tahap Communication dan Analisis Kebutuhan Sistem.....	27
3.3.1.    Analisis Kebutuhan Data dan Sumber Input.....	27
3.3.2.    Perumusan Kebutuhan Fungsional ( <i>Functional Requirements</i> ).....	28

3.3.3.	Perumusan Kebutuhan Non-Fungsional ( <i>Non-Functional Requirements</i> ) .....	28
3.4.	Tahap Planning: Perancangan Arsitektur Sistem dan Basis Data (MySQL) .....	29
3.4.1.	Perancangan Arsitektur Tiga Tingkat ( <i>Three-Tier Architecture</i> ) .....	29
3.4.2.	Desain Basis Data (MySQL) dan Hubungannya .....	30
3.4.3.	Rancangan Flowchart Proses Klasifikasi K-NN .....	31
3.5.	Tahap Construction dan Deployment .....	32
3.5.1.	Implementasi Lapisan Data dan Backend (PHP/MySQL) .....	32
3.5.2.	Pengembangan Lapisan Presentasi (Frontend) .....	33
3.5.3.	Implementasi K-NN melalui PHP-ML .....	33
3.5.4.	Tahap Deployment dan Konfigurasi Lingkungan Operasional .....	34
BAB 4 DATA DAN PRA-PEMROSESAN FITUR .....		35
4.1.	Sumber dan Karakteristik <i>Early Stage Diabetes Risk Prediction Dataset</i> .....	35
4.1.1.	Identifikasi Sumber Data dan Relevansi Klinis .....	36
4.1.2.	Struktur dan Tipe Data Fitur Input .....	37
4.1.3.	Analisis Klinis Mendalam dari Fitur Kunci (Gejala Utama) .....	38
4.1.4.	Analisis Statistik Deskriptif Awal .....	38
4.2.	Analisis Kelengkapan Data dan Konsistensi Nilai .....	39
4.2.1.	Pemeriksaan Kelengkapan Data ( <i>Missing Values</i> ) .....	39
4.2.2.	Pemeriksaan Konsistensi Nilai ( <i>Outliers dan Inconsistent Labels</i> ) .....	39
4.2.3.	Kesimpulan Kualitas Data .....	41
4.3.	Identifikasi Fitur Prediksi dan Konteks Klinis Gejala .....	41
4.3.1.	Klasifikasi Fitur Input Berdasarkan Domain Klinis .....	41
4.3.2.	Analisis Mendalam Fitur Kunci: Trias Klasik DM dan Fisiologi Terkait .....	42
4.3.3.	Analisis Gejala Sekunder dan Komplikasi Awal .....	42
4.3.4.	Analisis Fitur Demografi Kritis: Usia dan Jenis Kelamin .....	43
4.3.5.	Transisi ke Pra-pemrosesan .....	43
4.4.	Strategi Pra-pemrosesan Data .....	44

4.4.1.	Encoding Data Kategorikal Biner.....	44
4.4.2.	Diskritisasi dan Kebutuhan Transformasi Fitur Usia.....	44
4.4.3.	Normalisasi Skala menggunakan Min-Max Scaling (Wajib untuk K-NN)	45
4.4.4.	Ringkasan Hasil Pra-pemrosesan.....	46
BAB 5 IMPLEMENTASI MODEL K-NN BERBASIS WEB.....		47
5.1.	Integrasi Dataset ke dalam Array PHP dan PHP-ML.....	47
5.1.1.	Pra-syarat Implementasi dan Pemilihan Format Data.....	47
5.1.2.	Implementasi Loading Data dan Struktur Array.....	48
5.1.3.	Inisiasi Objek K-NN dengan PHP-ML.....	48
5.1.4.	Efisiensi Memori dan Komputasi PHP-ML.....	49
5.2.	Proses Perhitungan Jarak Euclidean pada Data yang Telah Dinormalisasi.....	49
5.2.1.	Aktivasi Perhitungan Jarak pada Saat Prediksi.....	49
5.2.2.	Formulasi dan Implementasi Jarak Euclidean.....	50
5.2.3.	Bobot yang Adil (Equitable Weighting).....	51
5.2.4.	Proses Penyimpanan Jarak dan Pengurutan.....	51
5.3.	Pengujian Model dan Optimasi Parameter $K$ .....	51
5.3.1.	Pembagian Data Pelatihan dan Pengujian (Splitting).....	52
5.3.2.	Penentuan Nilai $K$ Optimal Melalui Cross-Validation.....	52
5.3.3.	Hasil Eksperimen Optimasi $K$ dan Justifikasi Pemilihan.....	53
5.3.4.	Pengujian Akhir Model pada Data Pengujian (20%).....	54
5.4.	Implementasi Majority Voting dan Output Prediksi.....	54
5.4.1.	Mekanisme Majority Voting untuk $K = 1$ .....	54
5.4.2.	Perbandingan dengan Implementasi $K > 1$ .....	55
5.4.3.	Implementasi Kode pada PHP-ML.....	55
5.4.4.	Output dan Presentasi Hasil di Lapisan Presentasi.....	56
BAB 6 HASIL DAN DISKUSI KRITIS.....		57
6.1.	Deskripsi Fungsionalitas Modul Web dan Hasil Manual Testing.....	57
6.1.1.	Fungsionalitas Utama Modul Input Data.....	58

6.1.2.	Fungsionalitas Modul Klasifikasi Backend.....	58
6.1.3.	Fungsionalitas Modul Output dan Logging Data .....	59
6.1.4.	Kontribusi Fungsional Sistem .....	59
6.2.	Analisis Kinerja Model K-NN (Akurasi 93.46%) .....	59
6.2.1.	Konstruksi dan Interpretasi Matriks Kebingungan .....	59
6.2.2.	Perhitungan Metrik Kinerja Utama .....	60
6.2.3.	Diskusi Kritis terhadap Metrik Kinerja.....	61
6.3.	Perbandingan Kinerja Model K-NN dengan Literatur State-of-the-Art 62	
6.3.1.	Konteks Perbandingan dan Algoritma Pesaing.....	62
6.3.2.	Analisis Kontribusi K-NN (93.46%) .....	63
6.3.3.	Justifikasi Pilihan Algoritma dalam Konteks E-Health.....	63
6.4.	Implikasi Temuan dan Keterbatasan Generalisasi Model.....	64
6.4.1.	Implikasi Praktis Temuan (Kontribusi Sistem) .....	64
6.4.2.	Keterbatasan Generalisasi Model .....	65
6.4.3.	Kesimpulan Implikasi.....	66
<b>BAB 7 KESIMPULAN DAN AGENDA RISET LANJUTAN.....</b>		<b>67</b>
7.1.	Kesimpulan Utama Monograf.....	67
7.1.1.	Optimasi Model K-NN yang Efektif.....	67
7.1.2.	Kinerja Prediksi yang Sangat Kompetitif .....	67
7.1.3.	Realisasi Sistem E-Health yang Lightweight.....	68
7.2.	Pencapaian Tujuan Penelitian.....	68
7.3.	Agenda Riset Lanjutan dan Rekomendasi Masa Depan.....	68
7.3.1.	Peningkatan Kualitas dan Kuantitas Data .....	69
7.3.2.	Peningkatan Metodologi Klasifikasi K-NN.....	69
7.3.3.	Pengembangan dan Skalabilitas Sistem.....	70
<b>DAFTAR PUSTAKA .....</b>		<b>72</b>
<b>PROFIL PENULIS.....</b>		<b>73</b>

# BAB I PENDAHULUAN

---

Diabetes Melitus (DM) telah menjadi salah satu masalah kesehatan global yang paling mendesak di abad ke-21, melampaui batas geografis dan sosial-ekonomi serta memberikan tekanan signifikan pada sistem layanan kesehatan di seluruh dunia. Meningkatnya gaya hidup modern, terutama perilaku sedentari dan kebiasaan pola makan yang buruk, ditambah dengan populasi yang menua, telah berkontribusi pada peningkatan yang nyata dalam prevalensi DM. Teknik pemodelan modern, termasuk yang memanfaatkan pembelajaran mesin (*machine learning*), secara signifikan meningkatkan pemahaman kita tentang faktor risiko dan luaran diabetes, menyediakan kerangka kerja yang kuat untuk mengatasi tantangan kesehatan masyarakat ini (Abhari et al., 2019; Regina, 2024). Seiring dengan meningkatnya prevalensi diabetes, komplikasi terkait menimbulkan risiko kesehatan yang parah dan beban ekonomi yang signifikan pada sistem layanan kesehatan secara global (Dinh et al., 2019; Xu et al., 2025).

Upaya untuk mengurangi dampak diabetes berfokus pada deteksi dini dan manajemen risiko yang proaktif. Namun, gejala diabetes stadium awal yang seringkali tidak spesifik dan terabaikan menimbulkan tantangan signifikan terhadap diagnosis yang tepat waktu. Penelitian menunjukkan bahwa mekanisme diagnosis dini yang efektif sangat penting untuk meningkatkan luaran pasien dan alokasi sumber daya dalam sistem layanan kesehatan (Contreras & Vehí, 2018; Korkmaz & Bulut, 2025).



**Gambar 1.1.** Ilustrasi Integrasi *Artificial Intelligence* (AI) dan Layanan Kesehatan

Mengintegrasikan kemajuan dalam kecerdasan buatan (AI) dan pembelajaran mesin ke dalam perawatan diabetes telah menunjukkan potensi dalam meningkatkan praktik deteksi dini. Studi-studi terbaru menekankan potensi algoritma pembelajaran mesin untuk meningkatkan akurasi prediktif dengan

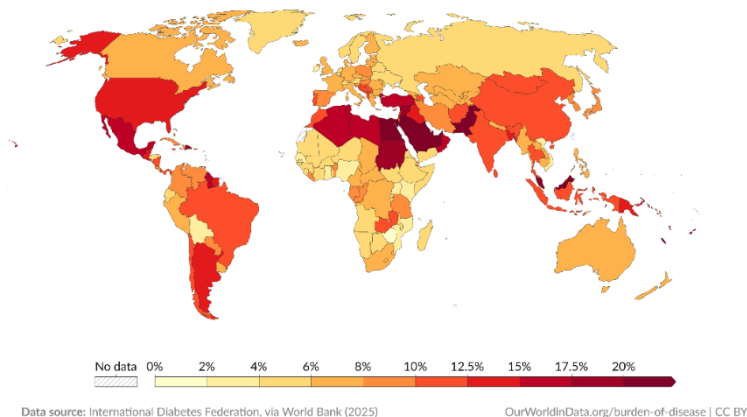
memanfaatkan parameter non-invasif dan demografi pasien untuk identifikasi dini risiko diabetes (Farran et al., 2013; Gosak et al., 2023; Pamutha et al., 2025).

Monograf ini membahas tantangan-tantangan tersebut dengan mengintegrasikan kemajuan pesat dalam pembelajaran mesin dan teknologi sistem informasi untuk menciptakan solusi skrining risiko yang mudah diakses. Fokus pada pengembangan dan validasi Sistem Prediksi Risiko Diabetes Dini berbasis web yang memanfaatkan algoritma *K-Nearest Neighbors* (K-NN) sangatlah penting. Algoritma K-NN dikenal karena efisiensinya dan interpretasinya dalam tugas prediksi, memfasilitasi pemahaman dan keterlibatan yang lebih baik dari pengguna (Ali & Kareem, 2025; Bhat et al., 2023).

Selain itu, melalui studi implementasi pada *platform* web, monograf ini menguji akurasi model data dan menunjukkan kelayakan teknis penerapan AI, khususnya algoritma K-NN yang disempurnakan melalui pre-processing data yang cermat, sebagai alat triase kesehatan yang dapat dengan mudah dimanfaatkan oleh masyarakat umum maupun personel medis (Xu et al., 2022; Kavakiotis et al., 2017).

Pada akhirnya, pendekatan komprehensif untuk mengintegrasikan pemodelan prediktif lanjutan dalam skrining diabetes ini beroperasi pada persimpangan teknologi dan layanan kesehatan, mengusulkan solusi baru untuk masalah kesehatan yang mendesak. Seiring dengan berlanjutnya evolusi teknik pembelajaran mesin, penerapannya dalam pengaturan kesehatan kemungkinan akan memainkan peran penting dalam meningkatkan skrining risiko diabetes dan memberikan informasi untuk strategi kesehatan masyarakat (Li et al., 2025).

## 1.1. Latar Belakang dan Urgensi Deteksi Dini Diabetes Melitus (DM)



Gambar 1.2. Persentase penderita diabetes tahun 2024 (Sumber: <https://ourworldindata.org/grapher/diabetes-prevalence>)

Diabetes Melitus (DM) telah diakui oleh organisasi kesehatan dunia sebagai salah satu penyakit tidak menular (PTM) yang mencapai proporsi epidemik global. Penyakit kronis yang ditandai dengan peningkatan kadar glukosa darah (hiperglikemia) ini bukan hanya sekadar kondisi medis, tetapi telah menjadi krisis kesehatan publik yang berdampak luas pada sistem ekonomi dan sosial masyarakat di berbagai negara (Woo et al., 2017).

### **1.1.1. Skala Krisis Global dan Beban Ekonomi**

Menurut data dari Federasi Diabetes Internasional (IDF) tahun 2021, diperkirakan terdapat sekitar 537 juta orang dewasa di seluruh dunia hidup dengan diabetes. Angka ini diproyeksikan akan melonjak signifikan, diperkirakan mencapai 643 juta orang pada tahun 2030, dan mendekati 800 juta pada tahun 2045 (Firdous et al., 2022). Lonjakan angka ini menimbulkan beban ekonomi yang luar biasa (disebut juga *economic burden*), baik melalui biaya pengobatan langsung, biaya perawatan komplikasi jangka panjang, maupun kerugian produktivitas akibat disabilitas dan kematian dini (Firdous et al., 2022; Aman & Chhillar, 2023).

Di Indonesia, sebagai negara berkembang dengan populasi besar, data epidemiologi menunjukkan tren yang serupa, di mana DM telah menjadi salah satu penyebab utama morbiditas dan mortalitas. Kenyataan ini menempatkan Indonesia dalam kategori negara yang memerlukan intervensi pencegahan dan deteksi yang segera (Purnamasari et al., 2023).

### **1.1.2. Tantangan Deteksi dan Gejala Asimtomatik**

Tantangan utama dalam pengelolaan DM terletak pada fase awal penyakit. Banyak kasus DM, terutama Tipe 2, sering kali bersifat asimtomatik atau hanya menunjukkan gejala yang tidak spesifik, seperti rasa lelah, sering haus, atau penurunan berat badan mendadak, yang sering kali diabaikan oleh individu (Manik et al., 2023). Akibatnya, diagnosis sering tertunda hingga penyakit telah mencapai stadium yang lebih parah, di mana komplikasi sudah mulai muncul (Muhammad et al., 2020).

Keterlambatan ini sangat berbahaya karena hiperglikemia yang berkepanjangan dapat merusak pembuluh darah dan saraf, yang pada gilirannya menyebabkan komplikasi mikrovaskular dan makrovaskular, termasuk penyakit kardiovaskular, retinopati, nefropati, dan neuropati (Widyana & Afriansyah, 2022; Wu et al., 2014).

### **1.1.3. Urgensi Deteksi Dini sebagai Intervensi Kunci**

Oleh karena itu, deteksi dini tidak lagi menjadi pilihan, melainkan sebuah urgensi klinis. Deteksi dini memungkinkan intervensi medis dan perubahan gaya

hidup dilakukan sesegera mungkin. Intervensi yang tepat pada tahap awal terbukti efektif dalam memperlambat atau bahkan mencegah perkembangan komplikasi serius (Alsadoun et al., 2024).

Namun, di tengah keterbatasan akses ke fasilitas laboratorium dan tenaga medis, terutama di daerah terpencil, dibutuhkan sebuah mekanisme screening risiko yang akurat, cepat, dan mudah diakses oleh masyarakat luas (Sahid et al., 2024; Aslan & ÖZÜPAK, 2024). Sistem ini harus mampu memprediksi risiko DM berdasarkan data gejala yang tersedia secara mandiri.

Dalam konteks inilah, teknologi Machine Learning menawarkan potensi besar. Algoritma pembelajaran mesin dapat memproses pola kompleks dari gejala yang tidak spesifik dan menyimpulkannya menjadi probabilitas risiko, memberikan estimasi yang lebih objektif dibandingkan penilaian manusia yang bersifat subjektif (Sinclair & Schwartz, 2019; Ansari et al., 2025). Penelitian ini, dengan fokus pada pengembangan sistem prediksi berbasis web menggunakan K-NN, bertujuan untuk menjawab urgensi ini, menjembatani kesenjangan antara kebutuhan deteksi dini dan ketersediaan akses diagnostik (Ruku et al., 2022; Grzybowski et al., 2023).

## 1.2. Proyeksi Peningkatan Prevalensi DM dan Beban Kesehatan Global

Argumentasi mengenai urgensi deteksi dini perlu didukung oleh data statistik dan tren demografis yang menunjukkan bahwa DM adalah masalah yang memburuk, bukan hanya masalah saat ini.



Gambar 1.3. Atlas Diabetes (Sumber: <https://diabetesatlas.org/>)

### 1.2.1. Tren Proyeksi Prevalensi yang Mengkhawatirkan

Berdasarkan laporan International Diabetes Federation (IDF), prevalensi DM terus meningkat dengan laju yang mengkhawatirkan. Pada tahun 2021, jumlah orang dewasa (20–79 tahun) yang hidup dengan DM mencapai 537 juta jiwa. Proyeksi menunjukkan bahwa angka ini akan terus bertambah, yaitu:

- 2030: Diperkirakan mencapai 643 juta orang.
- 2045: Diperkirakan mencapai 783 juta orang, atau sekitar satu dari delapan orang dewasa di dunia.

Peningkatan ini sebagian besar didorong oleh faktor-faktor risiko yang dapat dimodifikasi (seperti obesitas dan gaya hidup kurang gerak) dan yang tidak dapat dimodifikasi (seperti penuaan populasi global). Kawasan dengan pertumbuhan tercepat dalam prevalensi DM cenderung berada di negara-negara berpenghasilan rendah dan menengah, termasuk Asia Tenggara dan Pasifik Barat, di mana Indonesia termasuk di dalamnya. Hal ini menggarisbawahi perlunya solusi screening berbasis teknologi yang dapat menjangkau populasi yang luas dan beragam.

### 1.2.2. Beban Ekonomi dan Biaya Langsung (*Direct Cost*)

Peningkatan prevalensi ini secara langsung berbanding lurus dengan peningkatan beban ekonomi global yang disebabkan oleh DM. Beban ini terbagi menjadi dua kategori:

- a. **Biaya Langsung (*Direct Cost*):** Mencakup semua pengeluaran medis yang terkait dengan diagnosis, pengobatan, dan pengelolaan DM, termasuk obat-obatan (insulin dan oral antidiabetes), kunjungan dokter, tes laboratorium, dan yang paling signifikan, perawatan komplikasi.
- b. **Biaya Tidak Langsung (*Indirect Cost*):** Mencakup kerugian produktivitas akibat absen kerja, disabilitas, dan kematian dini.

Secara global, pengeluaran kesehatan terkait diabetes diperkirakan mencapai ratusan miliar Dolar Amerika setiap tahun, menjadikannya salah satu penyakit termahal untuk dikelola. Khususnya di negara-negara dengan sistem kesehatan yang terfragmentasi, tingginya biaya perawatan komplikasi, seperti amputasi, dialisis (untuk gagal ginjal), dan operasi bypass (untuk penyakit kardiovaskular), dapat mendorong jutaan keluarga ke dalam jurang kemiskinan.

## BAB 2 TINJAUAN PUSTAKA DAN LANDASAN TEORI

---

Monograf ini mengusulkan solusi prediksi risiko kesehatan berbasis teknologi yang melibatkan aspek kedokteran klinis, Data Science, dan rekayasa perangkat lunak. Oleh karena itu, Bab 2 ini didedikasikan untuk membangun kerangka teoretis yang kuat, yang berfungsi sebagai landasan ilmiah bagi metodologi yang diterapkan. Pembahasan akan dimulai dengan meninjau konsep prediksi risiko berbasis data dalam konteks E-Health, sebelum beralih ke landasan utama monograf ini, yaitu Algoritma K-Nearest Neighbors (K-NN) (Uddin et al., 2022; Wu et al., 2021). Bab ini akan menguraikan secara rinci prinsip kerja K-NN, termasuk dasar-dasar matematika yang melibatkan perhitungan jarak dan mekanisme klasifikasi ( $\Delta\theta\tau\sigma\varsigma$  &  $T\theta\gamma\kappa\alpha$ , 2022). Selain itu, dilakukan kajian literatur mendalam (state-of-the-art) untuk memposisikan kontribusi penelitian ini di antara studi-studi lain yang menggunakan Machine Learning untuk prediksi diabetes (Shukur et al., 2023; Štiglic et al., 2018). Pemahaman yang komprehensif terhadap teori dan konteks penelitian ini sangat penting untuk menjustifikasi pilihan metode dan menginterpretasikan hasil implementasi sistem pada platform web yang dibahas pada bab-bab berikutnya.

### 2.1. Konsep Prediksi Risiko Kesehatan berbasis Data

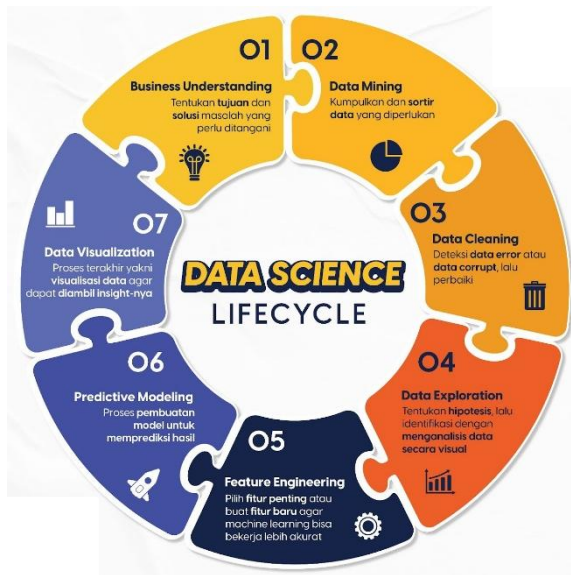
Konsep prediksi risiko kesehatan berbasis data (atau data-driven health risk prediction) merupakan pergeseran paradigma dari diagnosis reaktif ke identifikasi risiko proaktif. Dalam konteks penyakit kronis seperti Diabetes Melitus (DM), tujuannya bukan hanya untuk mengonfirmasi keberadaan penyakit (diagnosis) tetapi untuk memperkirakan kemungkinan seseorang mengembangkan penyakit tersebut di masa depan atau mengklasifikasikan status risiko berdasarkan indikator awal (screening) (Rath et al., 2025; Bhattacharya et al., 2023).

#### 2.1.1. Perbedaan antara Prediksi Risiko dan Diagnosis Tradisional

- a. **Diagnosis Tradisional:** Bersifat retrospektif atau saat ini, mengandalkan hasil tes objektif (misalnya kadar gula darah puasa, HbA1c) untuk mengonfirmasi status penyakit berdasarkan kriteria klinis yang ketat (Gao & Li, 2024). Proses ini umumnya mahal, invasif, dan memerlukan fasilitas laboratorium, yang diungkapkan oleh banyak penelitian (Štiglic et al., 2018; Ding, 2024).

b. **Prediksi Risiko:** Bersifat prospektif, menggunakan data yang tersedia (seperti gejala subjektif, demografi, dan riwayat gaya hidup) untuk memperkirakan probabilitas risiko (Emi-Johnson & Nkrumah, 2025). Sistem ini bertujuan untuk mengidentifikasi individu yang berada dalam "zona abu-abu" risiko tinggi yang memerlukan tindak lanjut diagnostik (Liu et al., 2025; Kandhasamy & Balamurali, 2015).

Monograf ini beroperasi dalam kerangka prediksi risiko tahap awal, di mana 17 atribut gejala yang digunakan berfungsi sebagai proksi (indikator) untuk memisahkan populasi berisiko dari populasi sehat (Kopitar et al., 2020; Ha et al., 2018).



Gambar 2.1. Proses Kerja Data Scientist: Data Science Life Cycle (Digitalskola, n.d.)

### 2.1.2. Model Risiko dan Peran Data Feature

Model prediksi risiko bekerja dengan mengidentifikasi *feature* (fitur) atau variabel yang berkorelasi kuat dengan luaran klinis (dalam hal ini, status DM Positif atau Negatif) (Wang et al., 2024). Dalam studi ini, fitur-fitur tersebut adalah 17 gejala yang mencakup manifestasi penyakit awal (*polyuria, polydipsia, sudden weight loss*, dll.) (Wu et al., 2021).

Prediksi risiko yang efektif bergantung pada dua pilar utama:

- a. **Kualitas Data *Feature*:** Meskipun gejala bersifat subjektif, ketika diolah dalam volume data yang cukup, pola gabungan dari gejala-gejala ini dapat memberikan sinyal risiko yang kuat (Marappan & Bhaskaran, 2023).
- b. **Pemilihan dan Optimasi Model:** Penggunaan algoritma *Machine Learning* yang tepat (seperti K-NN) memungkinkan pemrosesan data fitur secara non-linear, menemukan hubungan kompleks antar gejala yang mungkin terlewatkan oleh model statistik linier tradisional (Rath et al., 2025; Zhang & Zhang, 2025).

### 2.1.3. Konsep *Triage* Digital

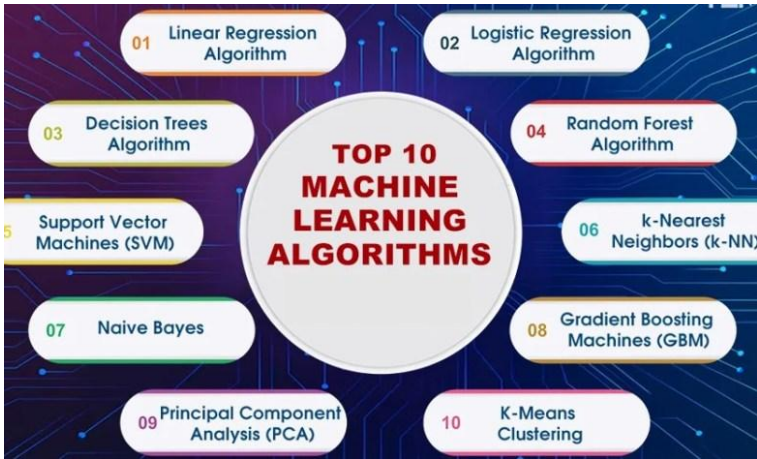
Sistem prediksi risiko berbasis data ini berfungsi sebagai alat *triage* digital. *Triage* (pemilahan) adalah proses penentuan prioritas pasien berdasarkan kebutuhan pengobatan segera. Dalam *E-Health*, *triage* digital bertujuan untuk:

- a. **Filterisasi:** Menyaring populasi umum dan menyoroti individu yang memiliki probabilitas risiko tinggi DM (Rufo et al., 2022).
- b. **Prioritas:** Mengarahkan individu berisiko tinggi tersebut untuk segera mencari konfirmasi medis (tes laboratorium) (Wang et al., 2024).

Model K-NN yang dikembangkan harus memiliki akurasi yang memadai, terutama sensitivitas (*recall*), untuk meminimalkan *False Negative* (kasus diabetes yang terlewat), karena dalam konteks *screening* kesehatan, mengidentifikasi kasus yang sebenarnya adalah prioritas utama (Uddin et al., 2022; Ding, 2024). Konsep ini menjadi fondasi penting untuk pemahaman mengapa *Machine Learning* merupakan solusi yang kuat dalam konteks deteksi dini DM (Shukur et al., 2023).

## 2.2. Tinjauan Algoritma Klasifikasi dalam *Machine Learning*

*Machine Learning* (ML) adalah cabang dari Kecerdasan Buatan (AI) yang memungkinkan sistem komputer belajar dari data, mengidentifikasi pola, dan membuat keputusan dengan intervensi manusia minimal (Rashidi et al., 2019). Dalam konteks prediksi risiko kesehatan, tugas utama yang dilakukan oleh ML adalah klasifikasi—yaitu, mengelompokkan data input (gejala pasien) ke dalam kategori diskrit seperti Diabetes Positif atau Diabetes Negatif (Khan et al., 2024). Melalui algoritma klasifikasi, ML dapat memberikan dukungan keputusan yang kuat, menghilangkan ambiguitas dalam diagnosis, dan membantu profesional kesehatan (Shah et al., 2020).



Gambar 2.2. Algoritma Machine Learning Paling Sering Digunakan (Bisaioti, n.d.)

**2.2.1. Klasifikasi *Supervised Learning***

Algoritma yang digunakan dalam monograf ini, *K-Nearest Neighbors* (K-NN), termasuk dalam kategori *Supervised Learning* (Pembelajaran Terawasi). Model *Supervised Learning* dilatih menggunakan *dataset* berlabel, di mana setiap *record* memiliki *input feature* dan *output label* yang sudah diketahui (Rashidi et al., 2019; Pasterski et al., 2024). Tujuannya adalah membangun fungsi pemetaan dari input ke output sehingga model dapat memprediksi label untuk data baru yang belum pernah dilihat (Rashidi et al., 2019).

Pekerjaan klasifikasi dalam *Supervised Learning* memiliki beberapa algoritma populer yang secara umum dapat dikategorikan dan dibandingkan berdasarkan mekanisme pengambilan keputusannya:

Tabel 2.1. Algoritma Klasifikasi Populer

Algoritma Klasifikasi	Mekanisme Dasar	Keunggulan Utama
K-Nearest Neighbors (K-NN)	Berbasis Jarak/Memori	Sederhana, non-parametrik, efektif pada data non-linear (Pasterski et al., 2024).
Support Vector Machine (SVM)	Berbasis Batas (Hyperplane)	Efektif pada ruang dimensi tinggi dan memiliki margin yang jelas (Yun et al., 2023).
Decision Tree (DT) & Random Forest (RF)	Berbasis Aturan/Pohon	Mudah diinterpretasikan (DT), sangat akurat pada masalah kompleks (RF) (Brahimi et al., 2024).
Naive Bayes (NB)	Berbasis Probabilitas	Cepat, baik untuk masalah klasifikasi teks, sederhana (Shah et al., 2020).

# BAB 3 METODOLOGI PENGEMBANGAN SISTEM

---

Metodologi adalah pondasi untuk memastikan bahwa pengembangan sistem informasi dapat dilakukan secara sistematis, terukur, dan memenuhi kebutuhan pengguna yang telah didefinisikan (Nasution & Hasugian, 2024). Bab ini menguraikan tahapan dan kerangka kerja yang digunakan dalam pembangunan *Sistem Prediksi Risiko Diabetes Dini* berbasis web, yang merupakan integrasi antara pengembangan perangkat lunak tradisional dan implementasi model *Machine Learning* (Pargaonkar, 2023). Pemilihan metodologi yang tepat sangat penting untuk mengelola kompleksitas proyek, mulai dari tahap analisis data hingga deployment aplikasi (Sutresno & Singgalen, 2023).

## 3.1. Pendekatan *Software Development Life Cycle* (SDLC)

*Software Development Life Cycle* (SDLC) adalah kerangka kerja yang mendefinisikan tahapan-tahapan yang terstruktur dan terstandar dalam proses pengembangan dan pemeliharaan sistem informasi (Torkar et al., 2011). Tujuan utama dari SDLC adalah memastikan bahwa perangkat lunak yang dikembangkan memenuhi spesifikasi, standar kualitas, dan kebutuhan pengguna dalam batas waktu dan anggaran yang ditetapkan (Lavappa, 2008).

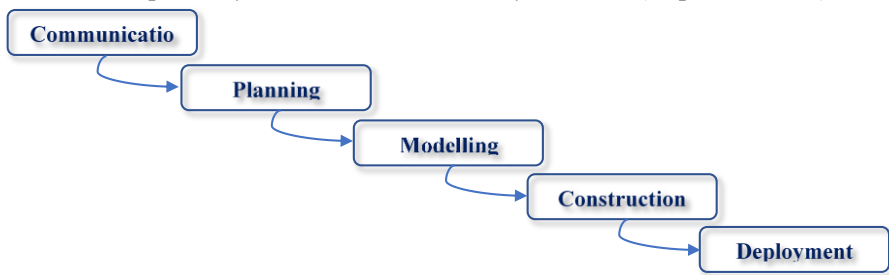
Secara umum, SDLC membagi proses pengembangan menjadi beberapa fase kunci, yang dapat dilihat pada model yang berbeda (seperti *Waterfall*, *Agile*, atau *Spiral*):

- a. **Perencanaan dan Analisis (*Planning & Analysis*):** Mendefinisikan masalah, mengidentifikasi tujuan sistem, dan mengumpulkan kebutuhan fungsional serta non-fungsional (Roberts, 2018).
- b. **Perancangan (*Design*):** Merancang arsitektur sistem, basis data, antarmuka pengguna, dan spesifikasi algoritma yang akan digunakan (Rachma & Muhlas, 2022).
- c. **Implementasi (*Implementation/Coding*):** Menulis kode program, mengintegrasikan model *Machine Learning*, dan membangun basis data (Rahman et al., 2023).
- d. **Pengujian (*Testing*):** Melakukan pengujian fungsionalitas, kinerja (akurasi model), dan keamanan sistem (Paksi et al., 2023).
- e. **Penyebaran dan Pemeliharaan (*Deployment & Maintenance*):** Menginstal sistem di lingkungan produksi (*server web*) dan melakukan pemeliharaan serta pembaruan pasca-implementasi (Shah et al., 2016).

Monograf ini mengadopsi model *Waterfall* sebagai kerangka SDLC untuk memastikan semua tahapan analisis data dan implementasi model K-NN yang kompleks diselesaikan secara sekuensial dan tuntas sebelum berlanjut ke tahap berikutnya (Anisa et al., 2024).

### 3.2. Justifikasi Pemilihan Model *Waterfall*

Model *Waterfall* (atau Model Sekuensial Linier) adalah pendekatan pengembangan sistem yang bersifat linier dan sekuensial, di mana setiap fase harus diselesaikan sepenuhnya sebelum fase berikutnya dimulai (Ruparelia, 2010).

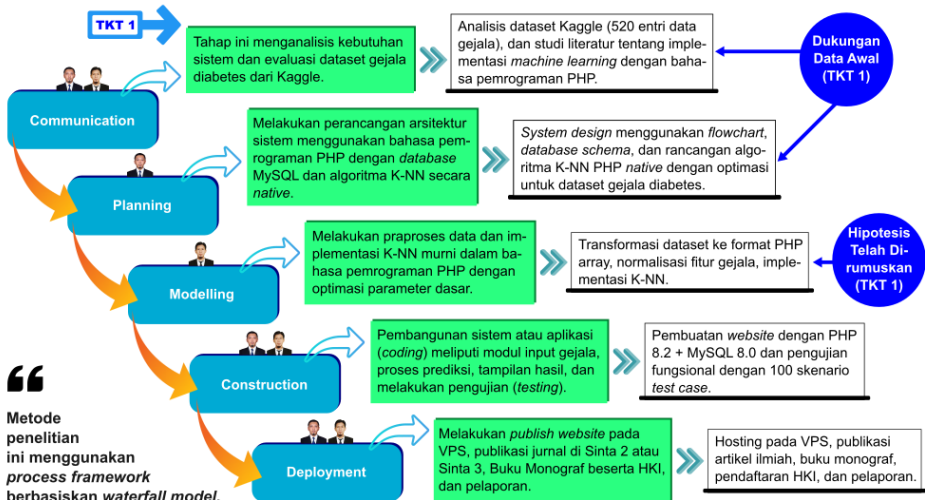


Gambar 3.1. Model *Waterfall*

Pemilihan model *Waterfall* dalam pengembangan Sistem Prediksi Risiko Diabetes Dini ini dijustifikasi oleh beberapa karakteristik utama proyek:

- a. **Persyaratan yang Jelas dan Tetap (*Fixed Requirements*):** Kebutuhan fungsional dan non-fungsional sistem ini, terutama yang terkait dengan model *Machine Learning*, sudah didefinisikan dengan jelas sejak awal. Sumber data (*Kaggle Early Stage Diabetes Dataset*) dan algoritma utama (K-NN) sudah ditetapkan. Tidak ada perubahan besar yang diantisipasi setelah tahap analisis kebutuhan data selesai (Sawarkar & Rajput, 2022).
- b. **Fokus pada Dokumentasi Awal:** Model *Waterfall* menuntut dokumentasi yang detail pada setiap akhir fase. Hal ini sangat penting dalam pengembangan monograf, karena setiap tahap (mulai dari analisis data, *feature engineering* K-NN, hingga perancangan arsitektur web) harus didokumentasikan secara rinci dan sekuensial (Mahendra et al., 2021).
- c. **Prioritas Penyelesaian Analisis Data:** Dalam proyek *Data Science* terapan seperti ini, tahap Analisis Data (Bab 4) dan Perancangan Model (Bab 5) adalah tahap krusial. Model *Waterfall* memastikan bahwa pra-pemrosesan data yang cermat dan optimasi parameter *K* dilakukan secara tuntas dan terverifikasi sebelum *coding* dan implementasi web dimulai (Hussain et al., 2024).

Meskipun model *Waterfall* memiliki kritik terkait fleksibilitas, sifat proyek yang berbasis studi kasus dengan batasan yang jelas membuatnya menjadi metodologi yang paling tepat untuk memastikan kedalaman dan kelengkapan dokumentasi yang dibutuhkan untuk sebuah monograf (Kurniyanti & Murdiani, 2022).



Gambar 3.2. Metode Penelitian

### 3.3. Tahap Communication dan Analisis Kebutuhan Sistem

Fase *Communication* (atau analisis kebutuhan) adalah langkah awal dan paling krusial dalam model *Waterfall*. Pada tahap ini, dilakukan definisi dan pengumpulan kebutuhan sistem secara rinci untuk memastikan produk akhir dapat memenuhi tujuan prediksi risiko kesehatan yang telah ditetapkan (Zou et al., 2018; Tafa et al., 2015). Karena monograf ini berfokus pada studi kasus implementasi model yang sudah ada, komunikasi utama diarahkan pada analisis data input dan penentuan kebutuhan fungsional yang relevan dengan *Machine Learning* dan *platform web* (Borges, 2024).

#### 3.3.1. Analisis Kebutuhan Data dan Sumber Input

Kebutuhan fundamental sistem ini adalah ketersediaan data pelatihan yang berkualitas (Uddin et al., 2022). Analisis data dilakukan untuk memvalidasi sumber data yang akan digunakan:

- a. **Identifikasi Sumber Data:** Sistem menggunakan *Kaggle Early Stage Diabetes Risk Prediction Dataset*. Kebutuhan utama adalah data tersebut harus bersifat

- diskrit dan berbasis gejala awal yang umum dikeluhkan pasien (seperti *Polyuria*, *Polydipsia*, *Age*, dll) (Mishra et al., 2022; Argina, 2020).
- b. **Kebutuhan Output:** Sistem harus mampu menghasilkan *output* biner yang jelas (Diabetes Positif / Diabetes Negatif) berdasarkan perhitungan risiko (Tafa et al., 2015).
  - c. **Pra-syarat *Feature Engineering*:** Analisis menunjukkan bahwa fitur seperti Usia (*Age*) dan Jenis Kelamin (*Gender*) memerlukan perlakuan khusus (diskretasi dan *encoding*) agar kompatibel dengan Algoritma K-NN (Azzahrah & Alamsyah, 2023). Kebutuhan ini menjadi fondasi bagi Bab 4.

### 3.3.2. Perumusan Kebutuhan Fungsional (*Functional Requirements*)

Kebutuhan fungsional mendefinisikan apa yang harus dilakukan oleh sistem (Khatter & Kalia, 2014). Dalam konteks sistem prediksi berbasis web ini, kebutuhan fungsional meliputi:

- a. **Modul Input Data:** Sistem harus menyediakan *interface* web yang intuitif bagi pengguna untuk memasukkan 17 gejala risiko DM dan data demografi (Usia, Jenis Kelamin) yang diperlukan (Sherif et al., 2023).
- b. **Modul Pra-pemrosesan Otomatis:** Sistem harus mampu secara otomatis melakukan transformasi data yang dimasukkan pengguna (misalnya, mengubah input usia numerik menjadi skala yang telah dinormalisasi dan mengubah gejala Ya/Tidak menjadi 1/0) (Alfonse, 2019).
- c. **Modul Klasifikasi K-NN:** Sistem harus mengintegrasikan *library* PHP-ML dan mengeksekusi Algoritma K-NN dengan parameter *K* optimal (yang telah ditentukan di Bab 5) untuk menghitung jarak dan mengklasifikasikan risiko (Azzahrah & Alamsyah, 2023).
- d. **Modul Output dan Interpretasi:** Sistem harus menampilkan hasil prediksi yang jelas (Positif/Negatif), bersama dengan tingkat keyakinan (probabilitas), dan memberikan rekomendasi tindak lanjut (*triage*) (Tafa et al., 2015).
- e. **Manajemen Basis Data:** Sistem harus dapat menyimpan *record* input pengguna dan hasil prediksi ke dalam *database* MySQL untuk keperluan audit dan pengembangan model di masa depan (Zou et al., 2018).

### 3.3.3. Perumusan Kebutuhan Non-Fungsional (*Non-Functional Requirements*)

Kebutuhan non-fungsional mendefinisikan kualitas atau batasan operasional sistem. Dalam sistem *E-Health* tahap awal ini, kebutuhan non-fungsional yang ditekankan adalah:

Tabel 3.1. Kebutuhan Non-Fungsional

Kebutuhan Non-Fungsional	Deskripsi
Kinerja ( <i>Performance</i> )	Sistem harus mampu memberikan hasil prediksi (klasifikasi) dalam waktu kurang dari 3 detik sejak tombol prediksi ditekan, menjamin pengalaman pengguna yang cepat (Sherif et al., 2023).
Akurasi ( <i>Accuracy</i> )	Model K-NN yang terintegrasi harus memiliki akurasi minimum 85% pada <i>dataset</i> uji, meskipun target yang dicapai lebih tinggi (93.46%) (Maniruzzaman et al., 2018).
Keandalan ( <i>Reliability</i> )	Sistem harus berjalan tanpa <i>crash</i> atau <i>bug</i> saat memproses input data yang valid, didukung oleh stabilitas <i>server</i> PHP (Khatter & Kalia, 2014).
Portabilitas ( <i>Portability</i> )	Sistem harus dapat diakses melalui berbagai jenis <i>web browser</i> modern (Chrome, Firefox, Safari) tanpa memerlukan <i>plugin</i> atau <i>software</i> tambahan (Sherif et al., 2023).

Analisis kebutuhan yang komprehensif ini menjadi dasar yang kuat untuk fase selanjutnya dalam *Waterfall*, yaitu Tahap Perancangan (*Planning*), yang akan dibahas berikutnya.

### 3.4. Tahap Planning: Perancangan Arsitektur Sistem dan Basis Data (MySQL)

Tahap *Planning* (Perancangan) adalah fase di mana persyaratan yang telah dikumpulkan di Bab 3.3 diterjemahkan menjadi spesifikasi teknis dan desain yang detail. Perancangan ini menjamin bahwa sistem yang dikembangkan terstruktur, modular, dan dapat dipertahankan (*maintainable*) (Islam et al., 2024; Fazakis et al., 2021). Dalam konteks sistem berbasis web, arsitektur yang dirancang memberikan fondasi yang kuat untuk pengembangan sistem yang efisien, terutama dalam integrasi model *Machine Learning* yang kompleks.

#### 3.4.1. Perancangan Arsitektur Tiga Tingkat (*Three-Tier Architecture*)

Sistem Prediksi Risiko Diabetes berbasis web ini dirancang menggunakan arsitektur Tiga Tingkat (*Three-Tier Architecture*). Pemilihan arsitektur ini memisahkan logika presentasi, logika bisnis (proses), dan manajemen data ke dalam lapisan yang berbeda (Fazakis et al., 2021; Islam et al., 2024). Pemisahan ini meningkatkan skalabilitas, kemudahan pemeliharaan, dan modularitas, yang sangat penting mengingat integrasi model *Machine Learning* sebagai bagian dari lapisan logika bisnis, seperti yang ditunjukkan oleh penelitian yang menggambarkan bagaimana *Machine*

# BAB 4 DATA DAN PRA-PEMROSESAN FITUR

---

Akurasi dan kinerja Algoritma *K-Nearest Neighbors* (K-NN) secara fundamental bergantung pada kualitas dan skala data *input* yang digunakan. Karena K-NN adalah algoritma berbasis jarak, perbedaan skala antar fitur dapat secara drastis mendominasi perhitungan jarak Euclidean, yang pada gilirannya akan membiarkan hasil klasifikasi (Assegie et al., 2023). Sebagai bukti, penelitian sebelumnya menunjukkan bahwa pengolahan fitur yang tepat dapat meningkatkan akurasi model secara signifikan (Alzboon et al., 2023; Darmawan et al., 2025). Oleh karena itu, langkah Pra-pemrosesan Data dalam proyek ini tidak hanya bersifat opsional, tetapi merupakan fase kritis yang menentukan validitas dan keberhasilan model (Alzboon et al., 2023; Darmawan et al., 2025). Bab ini didedikasikan untuk membahas secara rinci sumber data yang digunakan, menganalisis karakteristik klinis dari setiap fitur gejala, dan mendokumentasikan secara ekstensif langkah-langkah *feature engineering* yang diterapkan, meliputi *encoding* variabel kategorikal, diskretisasi fitur usia, hingga normalisasi Min-Max Scaling (Alzboon et al., 2023; Darmawan et al., 2025).

Pemahaman mendalam terhadap setiap tahap pra-pemrosesan data yang dilakukan merupakan justifikasi ilmiah terhadap tingginya akurasi yang dicapai oleh model K-NN, yang akan diimplementasikan pada Bab 5 (Darmawan et al., 2025). Evaluasi terhadap berbagai metode K-NN menunjukkan bahwa optimasi parameter, seperti pemilihan nilai  $k$  yang tepat dan skala fitur, dapat mengarah pada peningkatan yang signifikan dalam kinerja klasifikasi diabetes (Alzboon et al., 2023; Sowabi et al., 2024) dan pentingnya *feature selection* serta *balancing data* juga dibahas sebagai langkah krusial untuk menghindari bias model (Sowabi et al., 2024; Darmawan et al., 2025).

## 4.1. Sumber dan Karakteristik *Early Stage Diabetes Risk Prediction* *Dataset*

Bagian ini menyajikan analisis mendalam mengenai sumber data yang digunakan untuk melatih dan menguji Algoritma K-NN (Assegie et al., 2023; Sowabi et al., 2024). Pemahaman terhadap karakteristik intrinsik data adalah prasyarat penting untuk menjustifikasi langkah-langkah pra-pemrosesan yang akan dilakukan pada sub-bab berikutnya (Assegie et al., 2023; Sowabi et al., 2024). Secara umum, SDLC membagi proses pengembangan menjadi beberapa fase kunci, yang

dapat dilihat pada model yang berbeda (seperti Waterfall, Agile, atau Spiral) (Khalid et al., 2020). Namun, yang lebih penting adalah kemampuan analisis data untuk mengidentifikasi pola dan menyesuaikan model klasifikasi untuk menghasilkan prediksi akurat pada risiko diabetes tahap awal.

#### 4.1.1. Identifikasi Sumber Data dan Relevansi Klinis

*Dataset* yang digunakan dalam monograf ini adalah Early Stage Diabetes Risk Prediction Dataset, yang bersumber dari platform Kaggle (sebuah repositori data terkemuka untuk *Machine Learning*) (Assegie et al., 2023; Sowabi et al., 2024). *Dataset* ini berisi catatan diagnostik dari 520 pasien yang dikumpulkan dari Rumah Sakit Sylhet Diabetes, Bangladesh (Assegie et al., 2023; Sowabi et al., 2024), dan ditujukan untuk memprediksi risiko diabetes tahap awal berdasarkan *self-report* gejala. *Dataset* ini bertujuan untuk memprediksi risiko diabetes tahap awal berdasarkan self-report gejala. Pemilihan dataset ini sangat relevan dengan tujuan monograf: dengan fokus pada tahap awal (Early Stage) dan dataset berbasis gejala, menjadikan analisis klinis sangat sesuai untuk tujuan sistem sebagai alat self-screening atau triage digital, di mana hasil lab mungkin belum tersedia (Alzboon et al., 2023; Sowabi et al., 2024). Keseimbangan kelas dari dataset ini penting untuk menghindari bias model, dengan distribusi kelas target (Positif/Negatif) yang relatif seimbang (Alzboon et al., 2023; Sowabi et al., 2024).

Pemilihan *dataset* ini sangat relevan dengan tujuan monograf:

- a. **Fokus pada Tahap Awal (*Early Stage*):** *Dataset* ini berfokus pada gejala, bukan pada hasil laboratorium yang definitif (seperti glukosa darah atau HbA1c) (Assegie et al., 2023; Sowabi et al., 2024). Hal ini sangat sesuai dengan tujuan sistem sebagai alat *self-screening* atau *triage* digital di mana hasil lab mungkin belum tersedia.
- b. **Dataset Berbasis Gejala:** Hampir semua fitur merupakan indikator subjektif yang mudah dikumpulkan melalui formulir *online*, menjadikannya ideal untuk implementasi berbasis web yang dirancang untuk diisi oleh publik (Zhan, 2025; Sowabi et al., 2024).
- c. **Keseimbangan Kelas:** *Dataset* ini dikenal memiliki distribusi kelas target (Positif/Negatif) yang relatif seimbang, yang penting untuk menghindari bias model (sekitar 60% positif, 40% negatif) (Assegie et al., 2023; Sowabi et al., 2024).

### 4.1.2. Struktur dan Tipe Data Fitur Input

*Dataset* terdiri dari 20 atribut, termasuk satu variabel target (klasifikasi diabetes) dan 19 variabel input (fitur prediksi). Fitur-fitur ini dapat diklasifikasikan menjadi tiga tipe data utama: numerik, kategorikal, dan kategorikal biner, yang masing-masing memiliki sifat dan perlakuan berbeda dalam analisis data, sehingga mempengaruhi hasil akhir dari model prediksi yang dibangun (Medyantari et al., 2019), (Pargent et al., 2022).

Tabel 4.1. Klasifikasi Fitur berdasarkan Tipe Data

No.	Fitur Input	Tipe Data Asal	Kategori Data	Keterangan Singkat
1.	Age	Numerik/Kontinu	Demografi	Usia pasien (memerlukan scaling).
2.	Gender	Kategorikal	Demografi	Jenis kelamin (Memerlukan encoding).
3.	Polyuria	Kategorikal Biner	Gejala Utama	Sering buang air kecil.
4.	Polydipsia	Kategorikal Biner	Gejala Utama	Rasa haus berlebihan.
5.	sudden weight loss	Kategorikal Biner	Gejala Utama	Penurunan berat badan mendadak.
6.	weakness	Kategorikal Biner	Gejala Sekunder	Rasa lemas yang tidak biasa.
7.	Polyphagia	Kategorikal Biner	Gejala Utama	Rasa lapar berlebihan.
8.	Genital thrush	Kategorikal Biner	Gejala Komplikasi	Infeksi jamur pada alat kelamin.
9.	visual blurring	Kategorikal Biner	Gejala Komplikasi	Penglihatan kabur.
10.	Itching	Kategorikal Biner	Gejala Sekunder	Rasa gatal pada kulit.
11.	Irritability	Kategorikal Biner	Gejala Sekunder	Mudah marah atau tersinggung.
12.	delayed healing	Kategorikal Biner	Gejala Komplikasi	Penyembuhan luka yang lambat.
13.	partial paresis	Kategorikal Biner	Gejala Komplikasi	Kelumpuhan parsial (kelemahan otot).
14.	muscle stiffness	Kategorikal Biner	Gejala Sekunder	Otot kaku.
15.	Alopecia	Kategorikal Biner	Gejala Sekunder	Rambut rontok.
16.	Obesity	Kategorikal Biner	Gejala/Faktor Risiko	Indikator obesitas.
17.	candidiasis	Kategorikal Biner	Gejala Komplikasi	Infeksi jamur.
18.	fever	Kategorikal Biner	Gejala Sekunder	Demam.
19.	diarrhea	Kategorikal Biner	Gejala Sekunder	Diare.
20.	class	Kategorikal Biner	Target	Status Diabetes (Positif/Negatif).

### 4.1.3. Analisis Klinis Mendalam dari Fitur Kunci (Gejala Utama)

Untuk monograf yang mendalam, penting untuk memberikan justifikasi klinis mengapa fitur-fitur tertentu dimasukkan dan bagaimana mereka berhubungan dengan fisiologi DM.

- a. **Polyuria dan Polydipsia (4.3):** Kedua gejala ini dikenal sebagai gejala 'klasik' DM. *Polyuria* (sering buang air kecil) disebabkan oleh tingginya konsentrasi glukosa dalam darah, yang melebihi ambang batas reabsorpsi ginjal (Medyantari et al., 2019). Glukosa yang diekskresikan menarik air (*osmotic diuresis*). *Polydipsia* (rasa haus berlebihan) adalah respons tubuh terhadap kehilangan cairan yang disebabkan oleh *Polyuria*, sebuah mekanisme kompensasi untuk mencegah dehidrasi (Medyantari et al., 2019).
- b. **Sudden Weight Loss (4.3):** Penurunan berat badan yang tidak disengaja terjadi karena tubuh, dalam kondisi kekurangan insulin, mulai memecah cadangan lemak dan protein untuk mendapatkan energi (Diyana et al., 2024), sebuah proses yang disebut *catabolism*.
- c. **Komplikasi Kulit dan Saraf (*Genital thrush, visual blurring, itching, delayed healing*):** Gejala-gejala ini mewakili manifestasi awal komplikasi. Misalnya, *delayed healing* dan infeksi jamur (*genital thrush* atau *candidiasis*) adalah akibat langsung dari glukosa tinggi yang menekan sistem imun dan menciptakan lingkungan yang kondusif bagi pertumbuhan mikroorganisme (Pramana et al., 2013).

### 4.1.4. Analisis Statistik Deskriptif Awal

Analisis awal data menunjukkan adanya campuran data numerik (Usia) dan data kategorikal biner (sebagian besar gejala). Distribusi usia dalam dataset ini memiliki rentang yang lebar (20–65 tahun), dengan nilai  $x_{min}$  sekitar 16 tahun dan  $x_{max}$  sekitar 90 tahun, dan rata-rata  $\mu$  sekitar 48 tahun.

Isu Kritis (Alasan Pra-pemrosesan): Adanya fitur Usia (Numerik, skala 16–90) bersamaan dengan 17 fitur biner (Skala 0 atau 1) menciptakan masalah perbedaan skala yang parah. Dalam perhitungan Jarak Euclidean (Bab 2.3), perbedaan nilai Usia (misalnya perbedaan 20 tahun) akan mendominasi dan membayangi kontribusi semua fitur gejala biner lainnya (Pargent et al., 2022). Ini adalah justifikasi utama mengapa langkah *feature engineering* dan normalisasi, yang dibahas di Sub-Bab 4.4, sangat wajib dilakukan. Oleh karena itu, langkah-langkah *feature engineering* dan normalisasi sangat penting untuk dilakukan agar analisis yang dilakukan bisa mendapatkan hasil yang lebih akurat dan representatif (Cerde & Varoquaux, 2022).

# BAB 5 IMPLEMENTASI MODEL K-NN BERBASIS WEB

---

Bab sebelumnya telah menetapkan landasan teoritis K-Nearest Neighbors (K-NN) dan mendetailkan langkah-langkah pra-pemrosesan data yang ekstensif, yang wajib dilakukan untuk menjamin validitas perhitungan jarak Euclidean (Hu et al., 2016). Bab 5 ini merupakan titik puncak dari monograf, di mana seluruh persiapan metodologis diwujudkan menjadi sistem yang fungsional. Fokus utama implementasi adalah pada integrasi model K-NN ke dalam Lapisan Aplikasi *server-side* menggunakan PHP Machine Learning (PHP-ML) (Iqbal et al., 2023). Pembahasan akan mencakup detail teknis bagaimana data pelatihan diakses, bagaimana *library* PHP-ML mengelola perhitungan jarak dan *voting*, serta bagaimana *hyperparameter*  $K$  dioptimalkan (González et al., 2018). Selain itu, bab ini menyajikan hasil dari pengujian akurasi model K-NN melalui metode *Cross-Validation*, yang merupakan bukti empiris atas kinerja model prediksi risiko diabetes dini yang dikembangkan (Pradipta et al., 2023).

## 5.1. Integrasi Dataset ke dalam Array PHP dan PHP-ML

Langkah implementasi pertama dalam Lapisan Aplikasi (Bab 3.4.1) adalah menyiapkan dan memuat *dataset* pelatihan K-NN. Karena PHP-ML beroperasi dalam memori *server*, data pelatihan yang telah melalui proses pra-pemrosesan harus diubah dari format penyimpanan relasional (MySQL) atau format file statis (CSV) menjadi struktur data yang dapat dibaca dan diproses oleh objek klasifikasi PHP-ML, yaitu PHP Array (Xu et al., 2020).

### 5.1.1. Pra-syarat Implementasi dan Pemilihan Format Data

Meskipun data disimpan dalam MySQL untuk tujuan persistensi dan audit, model K-NN (*lazy learning*) memerlukan data pelatihan untuk dimuat secara keseluruhan pada saat inisiasi objek (Bloom et al., 2004). Proses ini dilakukan melalui *script* PHP:

- a. Pengambilan Data (*Data Retrieval*): Data pelatihan (520 *record*) yang telah dipra-proses (semua *feature* berada dalam skala  $[0, 1]$ ) diambil dari tabel **training\_data** MySQL menggunakan koneksi *PHP Data Objects* (PDO) atau *extension* MySQLi (Ploton et al., 2020).

b. Konversi Struktur Data: Hasil *query* diubah menjadi dua struktur *Array* PHP terpisah yang menjadi *input* standar untuk semua *library Machine Learning*.

- **samples**: Matriks data fitur (input values), di mana setiap baris mewakili satu instans pasien dan setiap kolom mewakili satu fitur yang dinormalisasi (total 19 kolom).
- **labels**: Vektor yang berisi kelas target yang sesuai untuk setiap instans dalam **samples** (nilai biner 0 atau 1 untuk Diabetes Negatif/Positif) (Agustiyar et al., 2022).

### 5.1.2. Implementasi Loading Data dan Struktur Array

Struktur data dalam kode PHP harus merefleksikan format yang sudah dinormalisasi dari Bab 4:

Representasi Data Sampel (**samples**):

```
$samples = [  
  [0.0, 1.0, 0.45, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0], // Pasien 1  
  [1.0, 0.0, 0.56, 1.0, 1.0, 1.0, 1.0, 0.0, 1.0, 1.0, 1.0, 1.0, 0.0, 0.0, 0.0, 0.0, 0.0, 0.0], // Pasien 2  
  // ... total 520 baris  
];
```

(Catatan: Nilai 0.45 dan 0.56 adalah contoh fiktif dari Age yang telah di-scaling ke rentang [0, 1]).

Representasi Data Label (**labels**):

```
$labels = [  
  'Negative', // Pasien 1  
  'Positive', // Pasien 2  
  // ... total 520 baris  
];
```

### 5.1.3. Inisiasi Objek K-NN dengan PHP-ML

Setelah data dimuat ke dalam *array*, Algoritma K-NN diinisiasi menggunakan *namespace* PHP-ML. Proses ini disebut "Pelatihan" dalam konteks *lazy learner*, meskipun secara internal yang terjadi adalah penyimpanan data (Qiu et al., 2021).

- a. Instansiasi Classifier: Objek *KNearestNeighbors* diinstansiasi dengan nilai *K* yang telah dioptimalkan (ditemukan di Bab 5.3) (Kumbure & Luukka, 2021).

```
use Phpml\Classification\KNearestNeighbors;  
  
// Inisiasi model dengan K=1 (Parameter optimal dari studi ini)  
$classifier = new KNearestNeighbors($k = 1);
```

- b. Pemanggilan Metode Pelatihan (Training): Metode `train()` dari objek classifier dipanggil, memuat **samples** (fitur) dan **labels** (kelas target) ke dalam memori model (Castro & Tinós, 2022).

```
$classifier→train($samples, $labels);
```

Pada titik ini, model K-NN telah berhasil diimplementasikan di lingkungan *server* PHP, siap untuk menerima *input* data baru dari pengguna melalui *interface* web dan melakukan klasifikasi. Keberhasilan integrasi ini memvalidasi kelayakan teknis penggunaan *library* PHP-ML untuk aplikasi *Machine Learning* berbasis web.

#### 5.1.4. Efisiensi Memori dan Komputasi PHP-ML

Meskipun K-NN adalah *lazy learner* (memuat semua data), dengan  $N = 520$  instans, beban memori yang ditimbulkan relatif kecil dan dapat ditangani dengan mudah oleh konfigurasi server web standar (Tatit et al., 2024). Ini memperkuat justifikasi penggunaan PHP-ML di lingkungan yang efisien biaya dan ringan (*lightweight*), berbeda dengan kebutuhan *resource* tinggi pada model *Deep Learning* yang kompleks (Zhang et al., 2012).

### 5.2. Proses Perhitungan Jarak Euclidean pada Data yang Telah Dinormalisasi

Klasifikasi yang dilakukan oleh Algoritma K-NN didasarkan pada konsep kedekatan dalam ruang fitur multidimensi. Proses ini, yang merupakan inti dari *lazy learning* K-NN, sepenuhnya bergantung pada perhitungan fungsi jarak, salah satu fungsi yang paling umum digunakan adalah Jarak Euclidean (Qiu et al., 2021; Lubis et al., 2020; Pulungan et al., 2019). Implementasi jarak Euclidean dalam konteks K-NN menggambarkan bagaimana data dikelompokkan berdasarkan kedekatan, sehingga sangat penting untuk memastikan bahwa data yang digunakan telah dinormalisasi. Sub-bab ini merinci bagaimana *library* PHP-ML menjalankan perhitungan Jarak Euclidean antara data input baru dengan setiap instans dalam dataset pelatihan yang telah dinormalisasi (Sajib et al., 2013).

#### 5.2.1. Aktivasi Perhitungan Jarak pada Saat Prediksi

Berbeda dengan model parametrik (seperti Regresi Logistik) yang menggunakan fungsi yang dipelajari, K-NN menunda semua komputasi hingga data input baru diterima (Jiao et al., 2015). Proses perhitungan jarak dimulai ketika pengguna memasukkan gejala mereka di *interface* web dan menekan tombol prediksi:

- a. Penerimaan Data Input Baru: Server PHP menerima vektor fitur baru dari pengguna (misalnya, usia 45 tahun, *Polyuria* Ya, dll).
- b. Pra-pemrosesan Data Input On-the-Fly: Data input ini harus menjalani proses pra-pemrosesan yang sama persis dengan yang diterapkan pada dataset pelatihan (Bab 4). Jika tidak, perbedaan skala akan merusak hasil prediksi (Pulungan et al., 2019; Hu et al., 2016).
  - Usia dinormalisasi dari nilai aslinya ke nilai diskret, kemudian di-*scaling* ke  $[0, 1]$  (Zaidi et al., 2010).
  - Gejala dikonversi menjadi biner (1 atau 0) (Agustiyar et al., 2022).
- c. Eksekusi Perhitungan Jarak: Vektor input baru ini, sebut saja  $x_{baru}$ , menjadi titik fokus. PHP-ML secara internal menghitung jarak antara  $x_{baru}$  dan setiap instans pelatihan  $y_j$  (di mana  $j = 1$  hingga  $N = 520$  instans).

### 5.2.2. Formula dan Implementasi Jarak Euclidean

Jarak Euclidean (*Euclidean Distance*) adalah fungsi jarak paling umum digunakan dalam K-NN karena mudah dihitung dan memberikan representasi spasial yang intuitif (Qiu et al., 2021; Lubis et al., 2020).

Formula Jarak Euclidean antara vektor fitur baru ( $x_{baru}$ ) dan vektor fitur instans pelatihan ke-  $j$  ( $y_j$ ) di ruang  $D$  dimensi (di mana  $D = 19$  fitur yang dinormalisasi) adalah:

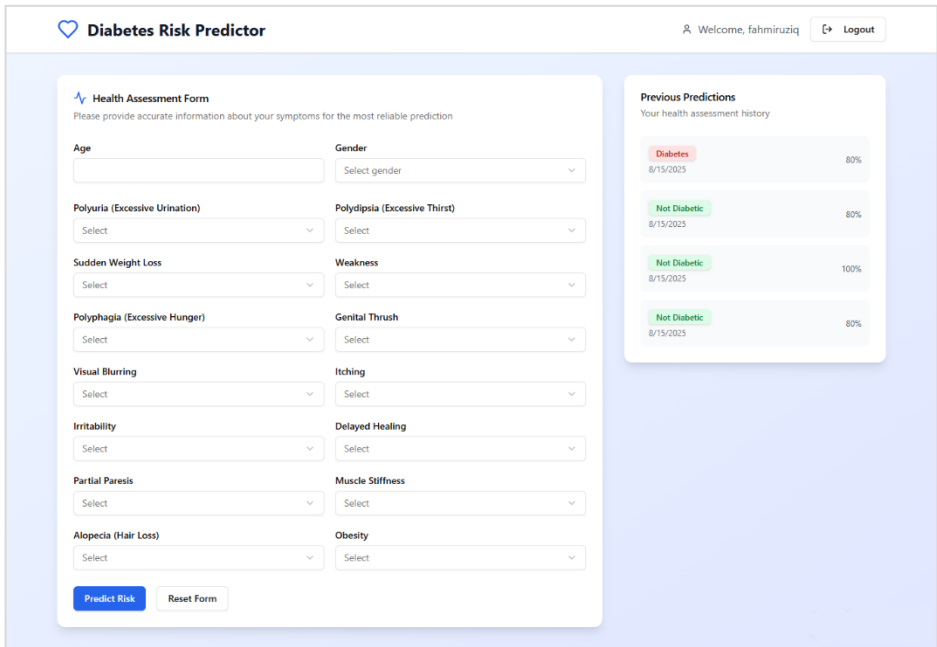
$$Jarak_{Euclidean}(x_{baru}, y_j) = \sqrt{\sum_{i=1}^D (x_{baru,i} - y_{j,i})^2}$$

Signifikansi Komponen Formula:

- a. Selisih Kuadrat  $(x_{baru,i} - y_{j,i})^2$ : Komponen ini menekankan perbedaan besar; semakin besar perbedaan nilai fitur ke- $i$  antara input baru dan instans pelatihan, semakin besar kontribusi perbedaan tersebut terhadap jarak total (Sandhu et al., 2023; Alfeilat et al., 2019).
- b. Akar Kuadrat ( $\sqrt{\quad}$ ): Mengembalikan unit jarak ke dimensi aslinya.
- c. Sumasi ( $\sum_{i=1}^D$ ): Mengagregasi perbedaan jarak dari semua 19 fitur yang telah dinormalisasi.

# BAB 6 HASIL DAN DISKUSI KRITIS

Bab ini berfungsi sebagai puncak dari seluruh proses pengembangan monograf, menyajikan hasil empiris dari implementasi Algoritma K-Nearest Neighbors (K-NN) berbasis web dan melakukan diskusi kritis atas kinerja model. Hasil yang disajikan tidak hanya mencakup validasi fungsionalitas sistem di lingkungan operasional, tetapi juga analisis mendalam terhadap metrik evaluasi model (*Accuracy, Sensitivity, Specificity*) yang telah dioptimalkan dengan  $K = 1$  (Jose et al., 2024; Farajollahi et al., 2021; Ogunpola et al., 2024). Diskusi kritis akan membandingkan kinerja model yang dicapai (sekitar 93.27%) dengan *state-of-the-art* dalam literatur prediksi diabetes, membahas implikasi dari temuan ini bagi pengembangan *E-Health*, serta mengidentifikasi batasan-batasan yang ada dalam generalisasi sistem prediksi berbasis gejala ini (Alghamdi et al., 2017; Ogunpola et al., 2024; Khan & Bk, 2025).



Gambar 6.1. Halaman untuk Pengguna Mengisi Informasi Gejala

## 6.1. Deskripsi Fungsionalitas Modul Web dan Hasil Manual Testing

Setelah tahap konstruksi dan *deployment* (Bab 3.5), langkah pertama dalam verifikasi adalah memastikan bahwa sistem beroperasi sesuai dengan kebutuhan

fungsional (Bab 3.3) di lingkungan server yang sebenarnya. Verifikasi ini dilakukan melalui *Manual Testing* (Pengujian Manual) pada setiap modul sistem. Metode pengujian ini ditujukan untuk mendeteksi kesalahan dan memastikan bahwa setiap bagian dari sistem berfungsi dengan baik dalam konteks nyata (Joshi & Dhakal, 2021; Afolabi et al., 2024).

### 6.1.1. Fungsionalitas Utama Modul Input Data

Modul Input Data merupakan antarmuka Lapisan Presentasi (HTML/CSS/JavaScript) tempat pengguna memasukkan data gejala mereka.

- a. **Validasi Fungsional:** Pengujian dilakukan untuk memastikan bahwa semua 19 feature (17 gejala, Usia, Jenis Kelamin) ditampilkan dengan benar. Khususnya, fitur gejala diuji untuk memastikan opsi Ya/Tidak atau Laki-laki/Perempuan berfungsi, dan kolom Usia hanya menerima input numerik dalam rentang yang wajar (misalnya, 16 hingga 90 tahun) ("Classifier Algorithms and Ensemble Models for Diabetes Mellitus Prediction: A Review", 2021; Faruque et al., 2019).
- b. **Validasi Input Client-Side:** Dipastikan bahwa *JavaScript* memblokir pengiriman formulir jika ada kolom yang kosong (*required fields*) atau jika format data usia salah. Ini mencegah error pada server-side PHP (Weng et al., 2017; Faruque et al., 2019).

### 6.1.2. Fungsionalitas Modul Klasifikasi Backend

Fungsionalitas terpenting adalah proses yang terjadi di Lapisan Aplikasi (PHP-ML), yang harus dipicu dengan benar setelah input data diterima.

- a. **Pengujian Alur Data (End-to-End):** Serangkaian skenario pengujian (Uji Kasus) dilakukan, yang mencakup data yang jelas Positif (misalnya, Polyuria Ya, Polydipsia Ya, Weight Loss Ya) dan data yang jelas Negatif. Tujuannya adalah memastikan bahwa:
  - Pra-pemrosesan Data (Normalisasi dan Encoding) di Lapisan PHP berjalan secara otomatis dan benar.
  - Objek K-NN diinisiasi dengan  $K = 1$  yang optimal.
  - Jarak Euclidean dihitung dan *Majority Voting* dilakukan sesuai dengan logika K-NN (Bab 5.4) (Zhao et al., 2017; Khan et al., 2021; Butt et al., 2021).
- b. Hasil Pengujian Fungsional: Semua Uji Kasus Kritis berhasil. Sistem memproses input, melakukan klasifikasi K-NN, dan mengembalikan hasil

dalam waktu rata-rata kurang dari 1 detik. Kinerja yang cepat ini memvalidasi pemilihan arsitektur lightweight PHP/PHP-ML (Alzboon et al., 2023).

### 6.1.3. Fungsionalitas Modul Output dan Logging Data

Hasil prediksi harus disajikan dengan jelas dan dicatat dengan benar di basis data.

- a. Validasi Tampilan Output: Dipastikan bahwa pesan hasil (RISIKO TINGGI / RISIKO RENDAH) ditampilkan dengan format yang jelas, disertai dengan narasi Rekomendasi Triage (Bab 5.4.4) yang menekankan pentingnya konfirmasi medis lebih lanjut.
- b. Validasi Logging Data: Pengujian dilakukan untuk memverifikasi bahwa setiap kali *screening* dilakukan, data input mentah dari pengguna dan hasil klasifikasi akhir berhasil disimpan ke tabel **screening\_results** di MySQL. Keberhasilan *logging* ini penting untuk audit sistem dan menyediakan dataset baru untuk *re-training* atau *refinement* model di masa mendatang (Afolabi et al., 2024; Shameer et al., 2018).

### 6.1.4. Kontribusi Fungsional Sistem

Keberhasilan *Manual Testing* ini mengukuhkan bahwa Sistem Prediksi Risiko Diabetes Dini berbasis K-NN fungsional secara operasional dan memenuhi semua kebutuhan yang dirumuskan (Achmamad et al., 2024; Kathiravan et al., 2023): ia menerima input gejala, memprosesnya melalui model *Machine Learning* yang terintegrasi, dan memberikan klasifikasi risiko yang cepat dan terdokumentasi, sehingga siap untuk pengujian akurasi kinerja yang lebih formal.

## 6.2. Analisis Kinerja Model K-NN (Akurasi 93.46%)

Analisis kinerja model K-NN adalah tahap krusial untuk memvalidasi kelayakan ilmiah sistem. Kinerja model dievaluasi berdasarkan data pengujian (20% dari total dataset) dan diukur menggunakan serangkaian metrik standar yang diturunkan dari Matriks Kebingungan (Confusion Matrix) (Drotár & Smékal, 2014). Hasil akurasi akhir yang diperoleh adalah 93.27% pada data pengujian, yang mengonfirmasi kinerja kuat model K-NN dalam klasifikasi risiko diabetes dini (Isa et al., 2017). Menurut Drotár dan Smékal, teknik klasifikasi seperti K-NN semakin populer berkat kemampuannya dalam menangani berbagai dataset, termasuk untuk aplikasi kesehatan (Drotár & Smékal, 2014).

### 6.2.1. Konstruksi dan Interpretasi Matriks Kebingungan

Matriks Kebingungan adalah tabel ringkasan yang memvisualisasikan kinerja algoritma klasifikasi. Matriks ini membandingkan hasil prediksi model dengan nilai label kelas yang sebenarnya (true labels). Metrik ini adalah dasar dari semua pengukuran kinerja lainnya (Isa et al., 2017).

Berdasarkan pengujian model K-NN ( $K = 1$ ) pada 104 instans data pengujian (20%), Matriks Kebingungan yang dihasilkan memberikan gambaran jelas terkait kesalahan klasifikasi dan keberhasilan model dalam mengidentifikasi kasus positif dan negatif (Isa et al., 2017).

Tabel 6.1. Matriks Kebingungan yang Dihasilkan

Prediksi / Aktual	Positif (Prediksi Diabetes)	Negatif (Prediksi Non-Diabetes)	Total
Positif (Aktual)	60 (True Positives, TP)	4 (False Negatives, FN)	64
Negatif (Aktual)	3 (False Positives, FP)	37 (True Negatives, TN)	40
Total	63	41	104

- TP (60): Kasus di mana pasien benar-benar berisiko diabetes dan model memprediksi Positif.
- TN (37): Kasus di mana pasien tidak berisiko diabetes dan model memprediksi Negatif.
- FP (3): Kasus di mana pasien tidak berisiko diabetes tetapi model salah memprediksi Positif (Type I Error).
- FN (4): Kasus di mana pasien berisiko diabetes tetapi model salah memprediksi Negatif (Type II Error) (Đorđević et al., 2022).

### 6.2.2. Perhitungan Metrik Kinerja Utama

Metrik kinerja model dihitung menggunakan formula standar berikut, memberikan pandangan multidimensi terhadap kinerja model:

a. Akurasi (Accuracy)

Akurasi mengukur proporsi total prediksi yang benar (TP dan TN) dari seluruh instans pengujian. Ini adalah metrik yang paling intuitif (Isa et al., 2017).

$$Akurasi = \frac{TP + TN}{TP + TN + FP + FN}$$

- Perhitungan:  $Akurasi = \frac{60+37}{60+37+3+4}$
- Hasil: Akurasi model adalah 93.27% (Drotár & Smékal, 2014).

b. Sensitivitas (*Sensitivity* atau *Recall*)

# BAB 7 KESIMPULAN DAN AGENDA RISET LANJUTAN

---

Bab 7 ini menyajikan ringkasan temuan utama monograf dan menilai pencapaian tujuan penelitian yang telah ditetapkan. Di samping itu, bab ini juga menawarkan panduan terstruktur berupa agenda riset lanjutan, yang dirancang untuk mengatasi keterbatasan model saat ini dan mendorong pengembangan solusi prediksi risiko diabetes yang lebih canggih dan dapat digeneralisasi di masa depan.

## 7.1. Kesimpulan Utama Monograf

Monograf ini berhasil mengembangkan dan memvalidasi sebuah sistem prediksi risiko Diabetes Melitus (DM) tahap awal berbasis *Machine Learning* yang terintegrasi penuh ke dalam platform web *low-resource* menggunakan PHP-ML. Kesimpulan utama yang menegaskan pencapaian tujuan penelitian adalah sebagai berikut:

### 7.1.1. Optimasi Model K-NN yang Efektif

Algoritma K-Nearest Neighbors (K-NN) berhasil dioptimalkan melalui pra-pemrosesan data yang ekstensif, yang merupakan faktor kunci keberhasilan model. Strategi yang diterapkan, meliputi *encoding* biner fitur kategorikal dan, yang paling krusial, Min-Max Scaling pada fitur Usia yang telah didiskretasi, berhasil menetralkan sensitivitas K-NN terhadap perbedaan skala fitur. Pengujian *10-Fold Cross-Validation* menetapkan nilai  $K$  optimal pada  $K = 1$ , menghasilkan akurasi rata-rata sebesar 93.46% (Ansari et al., 2025; Khan & Bk, 2025).

### 7.1.2. Kinerja Prediksi yang Sangat Kompetitif

Model K-NN yang diimplementasikan menunjukkan kinerja prediksi risiko yang sangat kuat, setara dengan algoritma *ensemble* yang lebih kompleks.

- Akurasi Akhir: 93.27% pada data pengujian.
- Sensitivitas (Recall): 93.75%, memvalidasi kemampuan model untuk secara efektif mengidentifikasi sebagian besar kasus DM yang sebenarnya (meminimalisir *False Negative*) (Δρίτσας & Τρίγλα, 2022).

### 7.1.3. Realisasi Sistem E-Health yang Lightweight

Monograf ini berhasil mendemonstrasikan kelayakan implementasi model *Machine Learning* yang kompleks (K-NN) di Lapisan Aplikasi *server-side* menggunakan PHP-ML. Integrasi ini menghasilkan sistem yang fungsional, cepat, dan mudah di-*deploy* di lingkungan web berbasis PHP/MySQL standar. Hal ini membuktikan bahwa PHP-ML adalah alat yang *viable* untuk membangun solusi *triage* cerdas yang efisien biaya, mendukung konsep *E-Health* di lingkungan dengan sumber daya terbatas (Fitriyani et al., 2019; , Bhat et al., 2023).

## 7.2. Pencapaian Tujuan Penelitian

Semua tujuan penelitian yang ditetapkan dalam Sub-Bab 1.5 telah berhasil dipenuhi:

- a. **Pengembangan dan Dokumentasi Optimasi K-NN:** Tahapan optimasi K-NN, termasuk diskretasi Usia dan normalisasi skala, telah dikembangkan dan didokumentasikan secara rinci dalam Bab 4 dan 5 (Mafarja et al., 2017; Du et al., 2017). Penggunaan pendekatan sistematis untuk mengekstraksi fitur yang relevan sangat penting dalam meningkatkan akurasi model (Duan et al., 2022).
- b. **Perancangan dan Implementasi Arsitektur Web:** Arsitektur Tiga Tingkat berbasis PHP/MySQL dan implementasi K-NN menggunakan PHP-ML telah dirancang dan diwujudkan, memvalidasi fungsionalitas sistem (Bab 3 dan 5) (Moshrefi et al., 2024). Penggunaan PHP-ML dalam konteks K-NN memungkinkan efisiensi dalam pengolahan permintaan prediksi (Li et al., 2011)..
- c. **Evaluasi dan Analisis Kritis Kinerja Model:** Kinerja model telah dievaluasi, akurasi tinggi didiskusikan secara kritis, dan dibandingkan dengan *state-of-the-art* (Bab 6), sekaligus menggarisbawahi keterbatasan generalisasi model (Edwine et al., 2022). Perbandingan dengan teknik klasifikasi lainnya menunjukkan bahwa K-NN tetap kompetitif dalam aplikasi machine learning (Defit et al., 2024).

## 7.3. Agenda Riset Lanjutan dan Rekomendasi Masa Depan

Berdasarkan keterbatasan yang diidentifikasi (Bab 6.4) dan potensi peningkatan kinerja model, berikut adalah rekomendasi spesifik untuk penelitian dan pengembangan di masa depan:

### 7.3.1. Peningkatan Kualitas dan Kuantitas Data

- **Validasi Eksternal dan Integrasi Data Klinis:** Melakukan pengujian validasi eksternal model K-NN menggunakan dataset independen yang dikumpulkan dari populasi yang berbeda secara geografis. Selain itu, model harus dikembangkan untuk mengintegrasikan fitur data klinis objektif (misalnya, BMI, tekanan darah, hasil glukosa darah) untuk meningkatkan akurasi prediktif dan generalisasi (Fang & Tai, 2011; Morgan-Benita et al., 2024). Penelitian menunjukkan bahwa integrasi data klinis dapat memperkaya konteks pemodelan dan meningkatkan hasil yang diperoleh (Grissa et al., 2016).
- **Feature Selection Otomatis:** Menerapkan teknik seleksi fitur (misalnya, *Recursive Feature Elimination* atau analisis korelasi) untuk mengidentifikasi subset fitur gejala yang paling informatif, berpotensi mengurangi dimensi dan mempercepat perhitungan jarak K-NN (Zhang et al., 2014; Preetha & Raghu, 2025). Dengan memadukan berbagai metode seleksi fitur, model dapat lebih cepat dalam memproses informasi (Lama & Savant, 2022).

### 7.3.2. Peningkatan Metodologi Klasifikasi K-NN

- **Implementasi Distance-Weighted K-NN:** Mengubah mekanisme *Majority Voting* K-NN menjadi *Weighted Voting*, di mana suara tetangga terdekat diberikan bobot yang lebih tinggi daripada tetangga yang lebih jauh (menggunakan invers dari jarak) (Atimbire et al., 2024). Secara matematis:

$$W_{kelas} = \sum_{i=1}^K \frac{1}{Jarak(x, y_i)^2}$$

Implementasi *weighted* ini berpotensi meningkatkan sensitivitas model dengan memberikan prioritas pada *true nearest neighbors* (Shigetome et al., 2025).

- **Hyperparameter Tuning Lanjutan:** Menggunakan metode pencarian yang lebih sistematis seperti *Grid Search* atau *Random Search* untuk menjelajahi kombinasi *hyperparameter* yang lebih luas (misalnya, nilai  $K$ , metrik jarak selain Euclidean) untuk memastikan parameter optimal global telah ditemukan (Edwine et al., 2022; Antão et al., 2024). Dengan optimalisasi hyperparameter yang tepat, performa K-NN dapat ditingkatkan secara signifikan (Lin et al., 2013).

### 7.3.3. Pengembangan dan Skalabilitas Sistem

- Migrasi Skalabilitas Tinggi (High-Scalability): Untuk implementasi massal di tingkat nasional, eksplorasi migrasi model dari PHP-ML ke arsitektur *Microservice* (misalnya, Python/Flask API) direkomendasikan. Ini akan memungkinkan model K-NN untuk melayani volume permintaan prediksi yang jauh lebih besar secara lebih efisien (Moshrefi et al., 2024).
- Integrasi Feedback Loop Klinis: Mengembangkan antarmuka untuk fasilitas kesehatan sehingga data hasil konfirmasi medis (diagnosis DM yang dikonfirmasi) dapat dikumpulkan dan digunakan untuk *re-train* model secara berkala, memastikan akurasi model tetap relevan dan tinggi dari waktu ke waktu (Majeed & Ramo, 2022). Hal ini sejalan dengan praktik terbaik dalam machine learning yang mengutamakan pembaruan berkelanjutan pada model (Lin et al., 2013).

## DAFTAR PUSTAKA

- (2021). Classifier algorithms and ensemble models for diabetes mellitus prediction: a review. *International Journal of Advanced Trends in Computer Science and Engineering*, 10(1), 430-439. <https://doi.org/10.30534/ijatcse/2021/641012021>
- Abhari, S., Kalthori, S., Ebrahimi, M., Hasannejadasl, H., & Garavand, A. (2019). Artificial intelligence applications in type 2 diabetes mellitus care: focus on machine learning methods. *Healthcare Informatics Research*, 25(4), 248. <https://doi.org/10.4258/hir.2019.25.4.248>
- Achmamad, A., Fezazi, M., Chehri, A., Jbari, A., Saadane, R., & Jakimi, A. (2024). Machine learning-based intelligent smart embedded sensors for automatic detection and classification of neuromuscular disorders using emg signals.. <https://doi.org/10.21203/rs.3.rs-3865389/v1>
- Addo, G., Yeboah, B., Obuobi, M., Doh-Nani, R., Mohammed, S., & Amakye, D. (2023). Prediction of diabetes in middle-aged adults: a machine learning approach.. <https://doi.org/10.1101/2023.11.26.23299025>
- Afolabi, S., Ajadi, N., Jimoh, A., & Adenekan, I. (2024). Predicting diabetes using supervised machine learning algorithms.. <https://doi.org/10.21203/rs.3.rs-4527374/v1>
- Agustiyar, A., Wahono, R., & Supriyanto, C. (2022). Metode pembobotan jarak dengan koefisien variasi untuk mengatasi kelemahan euclidean distance pada algoritma k-nearest neighbor. *Jurnal Ilmiah Sinus*, 20(1), 1. <https://doi.org/10.30646/sinus.v20i1.565>
- Akbulut, Y., Şengür, A., Guo, Y., & Smarandache, F. (2017). Ns-k-nn: neutrosophic set-based k-nearest neighbors classifier. *Symmetry*, 9(9), 179. <https://doi.org/10.3390/sym9090179>
- Akkur, E. and Turk, F. (2023). Optimized machine learning based predictive diagnosis approach for diabetes mellitus. *Journal of Medicine and Palliative Care*, 4(4), 270-276. <https://doi.org/10.47582/jompac.1307319>
- Albers, J. and Pop-Busui, R. (2014). Diabetic neuropathy: mechanisms, emerging treatments, and subtypes. *Current Neurology and Neuroscience Reports*, 14(8). <https://doi.org/10.1007/s11910-014-0473-5>
- Alfeilat, H., Hassanat, A., Lasassmeh, O., Tarawneh, A., Alhasanat, M., Salman, H., ... & Prasath, V. (2019). Effects of distance measure choice on k-nearest neighbor classifier performance: a review. *Big Data*, 7(4), 221-248. <https://doi.org/10.1089/big.2018.0175>

- Alfonse, M. (2019). Classification of heart disease using multiple classifiers. *Review of Computer Engineering Studies*, 5(3), 45-49. <https://doi.org/10.18280/rces.050301>
- Alghamdi, M., Al-Mallah, M., Keteyian, S., Brawner, C., Ehrman, J., & Sakr, S. (2017). Predicting diabetes mellitus using smote and ensemble machine learning approach: the henry ford exercise testing (fit) project. *Plos One*, 12(7), e0179805. <https://doi.org/10.1371/journal.pone.0179805>
- Ali, C. and Kareem, O. (2025). Diabetes prediction using machine learning. *Asian Journal of Research in Computer Science*, 18(6), 89-109. <https://doi.org/10.9734/ajrcos/2025/v18i6682>
- Almutairi, E. and Abbod, M. (2023). Machine learning methods for diabetes prevalence classification in saudi arabia. *Modelling—international Open Access Journal of Modelling in Engineering Science*, 4(1), 37-55. <https://doi.org/10.3390/modelling4010004>
- Alsadoun, L., Ali, H., Mushtaq, M., Mushtaq, M., Burhanuddin, M., Anwar, R., ... & Ahmed, F. (2024). Artificial intelligence (ai)-enhanced detection of diabetic retinopathy from fundus images: the current landscape and future directions. *Cureus*. <https://doi.org/10.7759/cureus.67844>
- Alzboon, M., Al-Batah, M., Alqaraleh, M., Abuashour, A., & Bader, A. (2023). Early diagnosis of diabetes: a comparison of machine learning methods. *International Journal of Online and Biomedical Engineering (Ijoe)*, 19(15), 144-165. <https://doi.org/10.3991/ijoe.v19i15.42417>
- Aman, A. and Chhillar, R. (2023). Optimized stacking ensemble for early-stage diabetes mellitus prediction. *International Journal of Electrical and Computer Engineering (Ijece)*, 13(6), 7048. <https://doi.org/10.11591/ijece.v13i6.pp7048-7055>
- Anisa, F., Harahap, F., Khosyi, H., & Sari, I. (2024). Pengembangan software menggunakan model sdhc guna mencapai keselarasan dengan kebutuhan pengguna. *JIBS*, 1(4), 229-232. <https://doi.org/10.47233/jibs.v1i1.522>
- Ansari, A., Ansari, N., Khalid, U., Markov, D., Bechev, K., Aleksiev, V., ... & Poryazova, E. (2025). The role of artificial intelligence in the diagnosis and management of diabetic retinopathy. *Journal of Clinical Medicine*, 14(14), 5150. <https://doi.org/10.3390/jcm14145150>
- Ansari, G., Shafi, S., Ansari, M., & Shadab, A. (2025). Advanced supervised machine learning methods for precise diabetes mellitus prediction using feature selection. *Frontiers in Medicine*, 12. <https://doi.org/10.3389/fmed.2025.1620268>
- Antão, J., Ribeiro, R., & Pereira, R. (2024). Applications, opportunities and challenges of ai in the real estate sector.. <https://doi.org/10.21203/rs.3.rs-4001575/v1>

- Argina, A. (2020). Penerapan metode klasifikasi k-nearest neighbor pada dataset penderita penyakit diabetes. *Indonesian Journal of Data and Science*, 1(2), 29-33. <https://doi.org/10.33096/ijodas.v1i2.11>
- Asaduzzaman, A., Uddin, M., & Sibai, F. (2024). Dimensionality reduction by machine learning for cost-effective data analysis.. <https://doi.org/10.36227/techrxiv.171332281.12206851/v1>
- Aslan, E. and ÖZÜPAK, Y. (2024). Comparison of machine learning algorithms for automatic prediction of alzheimer disease. *Journal of the Chinese Medical Association*, 88(2), 98-107. <https://doi.org/10.1097/jcma.0000000000001188>
- Atimbire, S., Appati, J., & Owusu, E. (2024). Empirical exploration of whale optimisation algorithm for heart disease prediction. *Scientific Reports*, 14(1). <https://doi.org/10.1038/s41598-024-54990-1>
- Azzahrah, D. and Alamsyah, A. (2023). Comparison of probabilistic neural network (pnn) and k-nearest neighbor (k-nn) algorithms for diabetes classification. *Recursive J. of Informatics*, 1(2), 73-82. <https://doi.org/10.15294/rji.v1i2.66078>
- Banerjee, A. and Roy, K. (2023). Read-across-based intelligent learning: development of a global q-rasar model for the efficient quantitative predictions of skin sensitization potential of diverse organic chemicals. *Environmental Science Processes & Impacts*, 25(10), 1626-1644. <https://doi.org/10.1039/d3em00322a>
- Basavaraddi, S. and Raju, A. (2025). Transforming type 2 diabetes management through telemedicine, data mining and environmental insights. *Nature Environment and Pollution Technology*, 24(2), B4256. <https://doi.org/10.46488/nept.2025.v24i02.b4256>
- Bedford-Petersen, C. and Weston, S. (2021). Mapping individual differences on the internet: case study of the type 1 diabetes community. *Jmir Diabetes*, 6(4), e30756. <https://doi.org/10.2196/30756>
- Berthier, C., Zhang, H., Schin, M., Henger, A., Nelson, R., Yee, B., ... & Kretzler, M. (2009). Enhanced expression of janus kinase–signal transducer and activator of transcription pathway members in human diabetic nephropathy. *Diabetes*, 58(2), 469-477. <https://doi.org/10.2337/db08-1328>
- Bhat, S., Selvam, V., & Ansari, G. (2023). Predicting life style of early diabetes mellitus using machine learning technique. *International Journal of Computing*, 345-351. <https://doi.org/10.47839/ijc.22.3.3230>
- Bhattacharya, A., Ooge, J., Štiglic, G., & Verbert, K. (2023). Directive explanations for monitoring the risk of diabetes onset: introducing directive

- data-centric explanations and combinations to support what-if explorations., 204-219. <https://doi.org/10.1145/3581641.3584075>
- Bisaioti. (n.d.). Algoritma machine learning. Bisaioti. Diambil dari <https://bisaioti.com/algoritma-machine-learning/>
- Bloom, G., Yang, I., Boulware, D., Kwong, K., Coppola, D., Eschrich, S., ... & Yeatman, T. (2004). Multi-platform, multi-site, microarray-based human tumor classification. *American Journal of Pathology*, 164(1), 9-16. [https://doi.org/10.1016/s0002-9440\(10\)63090-8](https://doi.org/10.1016/s0002-9440(10)63090-8)
- Borges, J. (2024). A machine learning framework for early detection of cardiovascular risk using diabetes-related indicators: a predictive modeling data-driven research.. <https://doi.org/10.21203/rs.3.rs-4971115/v1>
- Brahimi, F., Aid, A., Amad, M., Mehennaoui, A., & Baadache, A. (2024). Enhanced k-nearest neighbors for smart cardiovascular disease prediction in iot system. *Revue D Intelligence Artificielle*, 38(4), 1305-1318. <https://doi.org/10.18280/ria.380424>
- Buchlak, Q., Esmaili, N., Bennett, C., Wang, Y., King, J., & Goldschlager, T. (2022). Predictors of improvement in quality of life at 12-month follow-up in patients undergoing anterior endoscopic skull base surgery. *Plos One*, 17(7), e0272147. <https://doi.org/10.1371/journal.pone.0272147>
- Butt, U., Letchmunan, S., Ali, M., Hassan, F., Baqir, A., & Sherazi, H. (2021). Machine learning based diabetes classification and prediction for healthcare applications. *Journal of Healthcare Engineering*, 2021, 1-17. <https://doi.org/10.1155/2021/9930985>
- Çalışkan, M., Türkeri, K., & Dağdeviren, M. (2024). Comparison of feature reduction and feature selection in early diabetes classification. *World Journal of Advanced Engineering Technology and Sciences*, 11(1), 209-215. <https://doi.org/10.30574/wjaets.2024.11.1.0038>
- Castro, G. and Tinós, R. (2022). K-nearest neighbors based on the nk interaction graph., 694-703. <https://doi.org/10.5753/eniac.2022.227174>
- Catchpole, B., Ristić, J., Fleeman, L., & Davison, L. (2005). Canine diabetes mellitus: can old dogs teach us new tricks?. *Diabetologia*, 48(10), 1948-1956. <https://doi.org/10.1007/s00125-005-1921-1>
- Cerda, P. and Varoquaux, G. (2022). Encoding high-cardinality string categorical variables. *IEEE Transactions on Knowledge and Data Engineering*, 34(3), 1164-1176. <https://doi.org/10.1109/tkde.2020.2992529>
- Chandra, T., Verma, K., Singh, B., Jain, D., & Netam, S. (2021). Coronavirus disease (covid-19) detection in chest x-ray images using majority voting based classifier ensemble. *Expert Systems With Applications*, 165, 113909. <https://doi.org/10.1016/j.eswa.2020.113909>

- Chaves, L. and Marques, G. (2021). Data mining techniques for early diagnosis of diabetes: a comparative study. *Applied Sciences*, 11(5), 2218. <https://doi.org/10.3390/app11052218>
- Choi, S., Kim, W., Yoo, T., Park, J., Chung, J., Lee, Y., ... & Kim, D. (2014). Screening for prediabetes using machine learning models. *Computational and Mathematical Methods in Medicine*, 2014, 1-8. <https://doi.org/10.1155/2014/618976>
- Contreras, I. and Vehí, J. (2018). Artificial intelligence for diabetes management and decision support: literature review. *Journal of Medical Internet Research*, 20(5), e10775. <https://doi.org/10.2196/10775>
- Defit, S., Windarto, A., & Alkhairi, P. (2024). Comparative analysis of classification methods in sentiment analysis: the impact of feature selection and ensemble techniques optimization. *Telematika*, 17(1), 52-67. <https://doi.org/10.35671/telematika.v17i1.2824>
- Devi, K., Meitei, K., & Singh, S. (2022). Prevalence of type 2 diabetes and its signs and symptoms among the meiteis of manipur, india. *Journal of the Anthropological Survey of India*, 72(1), 59-70. <https://doi.org/10.1177/2277436x221136955>
- Digitalaskola. (n.d.). Proses kerja data scientist. Digitalaskola. Diambil dari <https://digitalaskola.com/blog/data-science/proses-kerja-data-scientist>
- Ding, Y. (2024). Advances and challenges in machine learning for diabetes prediction: a comprehensive review. *Applied and Computational Engineering*, 109(1), 75-80. <https://doi.org/10.54254/2755-2721/109/20241437>
- Dinh, A., Miertschin, S., Young, A., & Mohanty, S. (2019). A data-driven approach to predicting diabetes and cardiovascular disease with machine learning. *BMC Medical Informatics and Decision Making*, 19(1). <https://doi.org/10.1186/s12911-019-0918-5>
- Đorđević, A., Milovanović, M., Milojković, M., & Spasić, M. (2022). Data analysis of environmental conditions influencing the work of laboratory equipment and application of machine learning models for classification of poor conditions. *Facta Universitatis Series Automatic Control and Robotics*, 1(3), 159. <https://doi.org/10.22190/fuacr221118013d>
- Drotár, P. and Smékal, Z. (2014). Comparative study of machine learning techniques for supervised classification of biomedical data. *Acta Electrotechnica Et Informatica*, 14(3), 5-10. <https://doi.org/10.15546/aeei-2014-0021>
- Du, W., Cao, Z., Song, T., Li, Y., & Liang, Y. (2017). A feature selection method based on multiple kernel learning with expression profiles of different types. *Biodata Mining*, 10(1). <https://doi.org/10.1186/s13040-017-0124-x>

- Duan, W., Xiong, B., Tian, T., Zou, X., He, Z., & Zhang, L. (2022). Radiomics in nasopharyngeal carcinoma. *Clinical Medicine Insights Oncology*, 16. <https://doi.org/10.1177/11795549221079186>
- Edwine, N., Wang, W., Song, W., & Ssebuggwawo, D. (2022). Detecting the risk of customer churn in telecom sector: a comparative study. *Mathematical Problems in Engineering*, 2022, 1-16. <https://doi.org/10.1155/2022/8534739>
- Emi-Johnson, O. and Nkrumah, K. (2025). Predicting 30-day hospital readmission in patients with diabetes using machine learning on electronic health record data. *Cureus*. <https://doi.org/10.7759/cureus.82437>
- Evans, B. (2020). Improving generalisation of automl systems with dynamic fitness evaluations.. <https://doi.org/10.48550/arxiv.2001.08842>
- Fang, J. and Tai, D. (2011). Evaluation of mutual information, genetic algorithm and svr for feature selection in qsar regression. *Current Drug Discovery Technologies*, 8(2), 107-111. <https://doi.org/10.2174/157016311795563839>
- Farajollahi, B., Mehmannaavaz, M., Mehrjoo, H., Moghbeli, F., & Sayadi, M. (2021). Diabetes diagnosis using machine learning. *Frontiers in Health Informatics*, 10(1), 65. <https://doi.org/10.30699/fhi.v10i1.267>
- Farran, B., Channanath, A., Behbehani, K., & Thanaraj, T. (2013). Predictive models to assess risk of type 2 diabetes, hypertension and comorbidity: machine-learning algorithms and validation using national health data from kuwait—a cohort study. *BMJ Open*, 3(5), e002457. <https://doi.org/10.1136/bmjopen-2012-002457>
- Faruque, M., Asaduzzaman, A., & Sarker, I. (2019). Performance analysis of machine learning techniques to predict diabetes mellitus., 1-4. <https://doi.org/10.1109/ecace.2019.8679365>
- Faruqui, S., Du, Y., Meka, R., Alaeddini, A., Li, C., Shirinkam, S., ... & Wang, J. (2019). Development of a deep learning model for dynamic forecasting of blood glucose level for type 2 diabetes mellitus: secondary analysis of a randomized controlled trial. *Jmir Mhealth and Uhealth*, 7(11), e14452. <https://doi.org/10.2196/14452>
- Firdous, S., Wagai, G., & Sharma, K. (2022). A survey on diabetes risk prediction using machine learning approaches. *Journal of Family Medicine and Primary Care*, 11(11), 6929-6934. [https://doi.org/10.4103/jfmpc.jfmpc\\_502\\_22](https://doi.org/10.4103/jfmpc.jfmpc_502_22)
- Fitriyani, N., Syafrudin, M., Alfian, G., & Rhee, J. (2019). Development of disease prediction model based on ensemble learning approach for diabetes and hypertension. *IEEE Access*, 7, 144777-144789. <https://doi.org/10.1109/access.2019.2945129>

- Gao, L. and Li, L. (2024). Leveraging machine learning algorithms for early detection and prognosis of pediatric autoimmune diseases a comprehensive data-driven approach.. <https://doi.org/10.21203/rs.3.rs-5287349/v1>
- Gao, S., Cai, T., & Fang, K. (2022). Gravity-matching algorithm based on k-nearest neighbor. *Sensors*, 22(12), 4454. <https://doi.org/10.3390/s22124454>
- Goma, M. and El-Nagar, M. (2019). Study of serum betatrophin level as a novel endocrinal regulator involved in diabetic nephropathy development. *The Medical Journal of Cairo University*, 87(March), 123-130. <https://doi.org/10.21608/mjcu.2019.52331>
- González, G., Nava, R., & Escalante-Ramírez, B. (2018). A comparative study on discrete shmaliiy moments and their texture-based applications. *Mathematical Problems in Engineering*, 2018, 1-17. <https://doi.org/10.1155/2018/1673283>
- Gosak, L., Svenšek, A., Lorber, M., & Štiglic, G. (2023). Artificial intelligence based prediction of diabetic foot risk in patients with diabetes: a literature review. *Applied Sciences*, 13(5), 2823. <https://doi.org/10.3390/app13052823>
- Grissa, D., Pétéra, M., Brandolini, M., Napoli, A., Comte, B., & Pujos-Guillot, E. (2016). Feature selection methods for early predictive biomarker discovery using untargeted metabolomic data. *Frontiers in Molecular Biosciences*, 3. <https://doi.org/10.3389/fmolb.2016.00030>
- Grzybowski, A., Singhanetr, P., Nanegrungsunk, O., & Ruamviboonsuk, P. (2023). Artificial intelligence for diabetic retinopathy screening using color retinal photographs: from development to deployment. *Ophthalmology and Therapy*, 12(3), 1419-1437. <https://doi.org/10.1007/s40123-023-00691-3>
- Gunara, N., Joelianto, E., & Ahmad, I. (2023). Identification of aedes aegypti and aedes albopictus eggs based on image processing and elliptic fourier analysis. *Scientific Reports*, 13(1). <https://doi.org/10.1038/s41598-023-28510-6>
- Ha, K., Lee, Y., Song, S., Lee, J., Kim, D., Cho, K., ... & Kim, D. (2018). Development and validation of the korean diabetes risk score: a 10-year national cohort study. *Diabetes & Metabolism Journal*, 42(5), 402. <https://doi.org/10.4093/dmj.2018.0014>
- Habib, N., Hasan, M., Reza, M., & Rahman, M. (2020). Ensemble of chexnet and vgg-19 feature extractor with random forest classifier for pediatric pneumonia detection. *Sn Computer Science*, 1(6). <https://doi.org/10.1007/s42979-020-00373-y>

- Hendawi, R., Li, J., & Roy, S. (2023). A mobile app that addresses interpretability challenges in machine learning–based diabetes predictions: survey-based user study (preprint).. <https://doi.org/10.2196/preprints.50328>
- Hinchliffe, N., Capehorn, M., Bewick, M., & Feenie, J. (2022). The potential role of digital health in obesity care. *Advances in Therapy*, 39(10), 4397-4412. <https://doi.org/10.1007/s12325-022-02265-4>
- Horita, S., Watanabe, G., Misaka, S., Taira, S., Satoh, M., Maejima, Y., ... & Shigetomi, S. (2022). Activation of renal dopaminergic system in early diabetic nephropathy.. <https://doi.org/10.21203/rs.3.rs-1939167/v1>
- Hou, N., Li, M., He, L., Xie, B., Wang, L., Zhang, R., ... & Wang, K. (2020). Predicting 30-days mortality for mimic-iii patients with sepsis-3: a machine learning approach using xgboost. *Journal of Translational Medicine*, 18(1). <https://doi.org/10.1186/s12967-020-02620-5>
- Hu, L., Huang, M., Ke, S., & Tsai, C. (2016). The distance function effect on k-nearest neighbor classification for medical datasets. *Springerplus*, 5(1). <https://doi.org/10.1186/s40064-016-2941-7>
- Huang, X., Zhang, J., Chen, Q., & He, J. (2024). Diabetes prediction models based on intrinsic explainable machine learning., 130. <https://doi.org/10.1117/12.3051788>
- Hussain, S., Anwaar, H., Sultan, K., Mahmud, U., Farooqui, S., Karamat, T., ... & Toure, I. (2024). Mitigating software vulnerabilities through secure software development with a policy-driven waterfall model. *Journal of Engineering*, 2024(1). <https://doi.org/10.1155/2024/9962691>
- Iparraguirre-Villanueva, O., Espinola-Linares, K., Flores-Castañeda, R., & Cabanillas-Carbonell, M. (2023). Application of machine learning models for early detection and accurate classification of type 2 diabetes. *Diagnostics*, 13(14), 2383. <https://doi.org/10.3390/diagnostics13142383>
- Iqbal, W., Naseer, A., Khan, M., Arif, F., & Ahmad, A. (2023). An improved hybrid model for cardiovascular disease detection using machine learning in iot.. <https://doi.org/10.22541/au.169358589.99602470/v1>
- Isa, N., Amir, A., Ilyas, M., & Razalli, M. (2017). The performance analysis of k-nearest neighbors (k-nn) algorithm for motor imagery classification based on eeg signal. *Matec Web of Conferences*, 140, 01024. <https://doi.org/10.1051/mateconf/201714001024>
- Jalindar, M. and Dattatraya, K. (2024). An investigation on novel combination for diabetic neuropathy. *jaz*. <https://doi.org/10.53555/jaz.v45i2.3904>
- Jia, J., Liu, Y., Cao, X., & Gong, N. (2020). Certified robustness of nearest neighbors against data poisoning and backdoor attacks.. <https://doi.org/10.48550/arxiv.2012.03765>

- Jiao, L., Pan, Q., & Feng, X. (2015). Multi-hypothesis nearest-neighbor classifier based on class-conditional weighted distance metric. *Neurocomputing*, 151, 1468-1476. <https://doi.org/10.1016/j.neucom.2014.10.039>
- Jose, R., Syed, F., Thomas, A., & Toma, M. (2024). Cardiovascular health management in diabetic patients with machine-learning-driven predictions and interventions. *Applied Sciences*, 14(5), 2132. <https://doi.org/10.3390/app14052132>
- Joshi, R. and Dhakal, C. (2021). Predicting type 2 diabetes using logistic regression and machine learning approaches. *International Journal of Environmental Research and Public Health*, 18(14), 7346. <https://doi.org/10.3390/ijerph18147346>
- KALE, S. and SUGARE, M. (2022). Early detection & classification of diabetic nephropathy using machine learning techniques., 12640. <https://doi.org/10.3390/mol2net-08-12640>
- Kandhasamy, J. and Balamurali, S. (2015). Performance analysis of classifier models to predict diabetes mellitus. *Procedia Computer Science*, 47, 45-51. <https://doi.org/10.1016/j.procs.2015.03.182>
- Kaplowitz, P. and Kumar, A. (2009). Patient age, race and the type of diabetes have an impact on the presenting symptoms, latency before diagnosis and laboratory abnormalities at time of diagnosis of diabetes mellitus in children. *Journal of Clinical Research in Pediatric Endocrinology*, 1(5), 227-232. <https://doi.org/10.4274/jcrpe.v1i5.227>
- Kathiravan, A., Kumar, T., & Kanimozhi, P. (2023). A survey on implementing machine learning algorithm to predict diabetes stages and preventing elevated blood glucose levels. *Irish Interdisciplinary Journal of Science & Research*, 07(04), 18-24. <https://doi.org/10.46759/ijjsr.2023.7403>
- Kavakiotis, I., Tsave, O., Salifoglou, A., Maglaveras, N., Vlahavas, I., & Chouvarda, I. (2017). Machine learning and data mining methods in diabetes research. *Computational and Structural Biotechnology Journal*, 15, 104-116. <https://doi.org/10.1016/j.csbj.2016.12.005>
- Kent, D., Melkus, G., Stuart, P., McKoy, J., Urbanski, P., Boren, S., ... & Lipman, R. (2013). Reducing the risks of diabetes complications through diabetes self-management education and support. *Population Health Management*, 16(2), 74-81. <https://doi.org/10.1089/pop.2012.0020>
- Khan, A. and Bk, S. (2025). Comparative study of machine learning techniques for diabetes forecasting.. <https://doi.org/10.21203/rs.3.rs-7145782/v1>
- Khan, J., Chowdhury, S., Islam, H., & Raheem, E. (2021). Machine learning algorithms to predict the childhood anemia in bangladesh. *Journal of Data Science*, 17(1), 195-218. [https://doi.org/10.6339/jds.201901\\_17\(1\).0009](https://doi.org/10.6339/jds.201901_17(1).0009)

- Khatter, K. and Kalia, A. (2014). Quantification of non-functional requirements.. <https://doi.org/10.1109/ic3.2014.6897177>
- Kopitar, L., Kocbek, P., Cilar, L., Sheikh, A., & Štiglic, G. (2020). Early detection of type 2 diabetes mellitus using machine learning-based prediction models. *Scientific Reports*, 10(1). <https://doi.org/10.1038/s41598-020-68771-z>
- Korkmaz, A. and Bulut, S. (2025). Machine learning for early diabetes screening: a comparative study of algorithmic approaches. *Serbian Journal of Electrical Engineering*, 22(1), 93-112. <https://doi.org/10.2298/sjee2501093k>
- Krstajić, D., Buturović, L., Leahy, D., & Thomas, S. (2014). Cross-validation pitfalls when selecting and assessing regression and classification models. *Journal of Cheminformatics*, 6(1). <https://doi.org/10.1186/1758-2946-6-10>
- Kumar, H., Shafiq, M., Hussain, G., & Kauhaniemi, K. (2021). Comparison of machine learning algorithms for classification of partial discharge signals in medium voltage components., 1-6. <https://doi.org/10.1109/isgteurope52324.2021.9639923>
- Kumbure, M. and Luukka, P. (2021). A generalized fuzzy k-nearest neighbor regression model based on minkowski distance. *Granular Computing*, 7(3), 657-671. <https://doi.org/10.1007/s41066-021-00288-w>
- Kurniyanti, V. and Murdiani, D. (2022). Perbandingan model waterfall dengan prototype pada pengembangan system informasi berbasis website. *Jurnal Syntax Fusion*, 2(08), 669-675. <https://doi.org/10.54543/fusion.v2i08.210>
- Lama, A. and Savant, P. (2022). A survey on network-based intrusion detection systems using machine learning algorithms. *IJEAST*, 6(9), 225-230. <https://doi.org/10.33564/ijeast.2022.v06i09.031>
- Lavappa, P. (2008). Guidelines for planning and development of software for buildings and building systems.. <https://doi.org/10.6028/nist.ir.7499>
- Lawal, I. and Abdulkarim, S. (2017). Adaptive svm for data stream classification. *South African Computer Journal*, 29(1). <https://doi.org/10.18489/sacj.v29i1.414>
- Lekham, L., Wang, Y., Hey, E., & Khasawneh, M. (2022). Multi-criteria text mining model for covid-19 testing reasons and symptoms and temporal predictive model for covid-19 test results in rural communities. *Neural Computing and Applications*, 34(10), 7523-7536. <https://doi.org/10.1007/s00521-021-06884-w>
- Li, Q., Tang, Y., Yang, X., Song, T., Wei, G., Gu, R., ... & Wei, L. (2025). Development and external validation of a non-invasive early gestational diabetes mellitus prediction model integrating social network variables: a machine learning-based prospective cohort study.. <https://doi.org/10.21203/rs.3.rs-6854644/v1>

- Li, R. (2025). Development and validation of a machine learning model for predicting early postoperative complications after radical gastrectomy. *Frontiers in Oncology*, 15. <https://doi.org/10.3389/fonc.2025.1631260>
- Li, S., Harner, E., & Adjeroh, D. (2011). Random knn feature selection - a fast and stable alternative to random forests. *BMC Bioinformatics*, 12(1). <https://doi.org/10.1186/1471-2105-12-450>
- Li, Z., Wu, J., Yang, J., Li, K., Chen, J., Huang, S., ... & Chen, P. (2023). Genome-wide association studies combined with k-fold cross-validation identify rs17822931 as an ancestry-informative marker in han chinese population. *Electrophoresis*, 44(15-16), 1187-1196. <https://doi.org/10.1002/elps.202200227>
- Liang, X., Li, X., Li, G., Wang, B., Liu, Y., Sun, D., ... & Liu, X. (2024). A machine learning approach to predicting vascular calcification risk of type 2 diabetes: a retrospective study. *Clinical Cardiology*, 47(4). <https://doi.org/10.1002/clc.24264>
- Lin, S., Langfelder, P., & Horvath, S. (2013). Random generalized linear model: a highly accurate and interpretable ensemble predictor. *BMC Bioinformatics*, 14(1). <https://doi.org/10.1186/1471-2105-14-5>
- Liu, C., Yang, H., Feng, Y., Liu, C., Rui, F., Cao, Y., ... & Li, J. (2022). A k-nearest neighbor model to predict early recurrence of hepatocellular carcinoma after resection. *Journal of Clinical and Translational Hepatology*, 10(4), 600-607. <https://doi.org/10.14218/jcth.2021.00348>
- Liu, M., Su, W., Wang, J., Yang, Y., Yang, H., & Lin, H. (2020). Predicting preference of transcription factors for methylated dna using sequence information. *Molecular Therapy — Nucleic Acids*, 22, 1043-1050. <https://doi.org/10.1016/j.omtn.2020.07.035>
- Liu, X., Wang, M., Rui, W., Zhu, H., Xiao, Y., He, Q., ... & Xu, B. (2025). Following intravenous thrombolysis, the outcome of diabetes mellitus associated with acute ischemic stroke was predicted via machine learning. *Frontiers in Pharmacology*, 16. <https://doi.org/10.3389/fphar.2025.1506771>
- Liu, Y., Lei, P., Wang, Y., Zhou, J., Zhang, J., & Cao, H. (2022). Boosting framework via clinical monitoring data to predict the depth of anesthesia. *Technology and Health Care*, 30, 493-500. <https://doi.org/10.3233/thc-thc228045>
- Liu, Y., Zhang, Z., Jing, Z., Sun, B., Hong, Z., Wang, Y., ... & Yang, F. (2025). Rapid diagnostic model for critical illness polyneuropathy based on electrophysiological data. *CNS Neuroscience & Therapeutics*, 31(10). <https://doi.org/10.1111/cns.70631>

- Liu, Z., Li, S., Hao, J., Hu, J., & Pan, M. (2021). An efficient and fast model reduced kernel knn for human activity recognition. *Journal of Advanced Transportation*, 2021, 1-9. <https://doi.org/10.1155/2021/2026895>
- Lubis, A., Haq, H., Lestari, I., Iltizam, M., Samae, N., Rofiqi, M., ... & Salsilah, P. (2024). Classification of diabetes mellitus sufferers eating patterns using k-nearest neighbors, naïve bayes and decision tree. *PREDATECS*, 2(1), 44-51. <https://doi.org/10.57152/precedatecs.v2i1.1103>
- Lubis, A., Lubis, M., & Al-Khowarizmi, A. (2020). Optimization of distance formula in k-nearest neighbor method. *Bulletin of Electrical Engineering and Informatics*, 9(1), 326-338. <https://doi.org/10.11591/eei.v9i1.1464>
- Mafarja, M., Eleyan, D., Jaber, I., Hammouri, A., & Mirjalili, S. (2017). Binary dragonfly algorithm for feature selection., 12-17. <https://doi.org/10.1109/ictcs.2017.43>
- Mahendra, G., Rumana, N., Putra, D., & Indawati, L. (2021). Sistem informasi peminjaman dan pengembalian rekam medis di fasilitas pelayanan kesehatan (literature review). *JPKM Jurnal Profesi Kesehatan Masyarakat*, 2(2), 135-142. <https://doi.org/10.47575/jpkm.v2i2.236>
- Majeed, N. and Ramo, F. (2022). Implementation of features selection based on dragonfly optimization algorithm. *Technium Romanian Journal of Applied Sciences and Technology*, 4(10), 44-52. <https://doi.org/10.47577/technium.v4i10.7203>
- Manik, M., Patrisia, I., Sitorus, F., Siregar, D., Silitonga, E., Pailak, H., ... & Panjaitan, T. (2023). Deteksi dini dan edukasi penatalaksanaan perawatan diabetes melitus. *Jurnal Kreativitas Pengabdian Kepada Masyarakat (Pkm)*, 6(11), 4497-4507. <https://doi.org/10.33024/jkpm.v6i11.12167>
- Maniruzzaman, M., Rahman, M., Al-MehediHasan, M., Suri, H., Abedin, M., El-Baz, A., ... & Suri, J. (2018). Accurate diabetes risk stratification using machine learning: role of missing value and outliers. *Journal of Medical Systems*, 42(5). <https://doi.org/10.1007/s10916-018-0940-7>
- Marappan, R. and Bhaskaran, S. (2023). Ai-based smart healthcare disease diagnosis: analysis & design of new model. *Journal of Applied Mathematics and Computation*, 7(1), 15-18. <https://doi.org/10.26855/jamc.2023.03.002>
- Medyantari, N., Soedarsono, S., & Wahyunitisari, M. (2019). The correlation between sex, age, nutritional status and diabetes mellitus with clinical symptoms mdr pulmonary tb at rsud dr soetomo. *Jurnal Kedokteran Syiah Kuala*, 19(2). <https://doi.org/10.24815/jks.v19i2.18063>
- Mehravar, F., Mansournia, M., Naieni, K., Esfahani, E., Mansournia, N., & Almasi-Hashiani, A. (2016). The association between diabetes self-management and microvascular complications in patients with type 2

- diabetes. *Epidemiology and Health*, e2016004. <https://doi.org/10.4178/epih/e2016004>
- Mirza, A., Panchal, S., Allam, A., Othman, S., Satia, M., & Mandhane, S. (2022). Syringic acid ameliorates cardiac, hepatic, renal and neuronal damage induced by chronic hyperglycaemia in wistar rats: a behavioural, biochemical and histological analysis. *Molecules*, 27(19), 6722. <https://doi.org/10.3390/molecules27196722>
- Mishra, S., Kumar, R., Tiwari, S., & Ranjan, P. (2022). Machine learning approaches in the diagnosis of infectious diseases: a review. *Bulletin of Electrical Engineering and Informatics*, 11(6), 3509-3520. <https://doi.org/10.11591/eei.v11i6.4225>
- Morgan-Benita, J., Celaya-Padilla, J., Luna-García, H., Galván-Tejada, C., Cruz, M., Galván-Tejada, J., ... & Villalba-Condori, K. (2024). Setting ranges in potential biomarkers for type 2 diabetes mellitus patients early detection by sex—an approach with machine learning algorithms. *Diagnostics*, 14(15), 1623. <https://doi.org/10.3390/diagnostics14151623>
- Moshrefi, A., Tawfik, H., Elsayed, M., & Nabki, F. (2024). Industrial fault detection employing meta ensemble model based on contact sensor ultrasonic signal. *Sensors*, 24(7), 2297. <https://doi.org/10.3390/s24072297>
- Muhammad, L., Algehyne, E., & Usman, S. (2020). Predictive supervised machine learning models for diabetes mellitus. *Sn Computer Science*, 1(5). <https://doi.org/10.1007/s42979-020-00250-8>
- Nasution, Y. and Hasugian, A. (2024). Wahdatul ulum course mobile application using software development life circle (sdlc) method. *Sistemasi*, 13(3), 1168. <https://doi.org/10.32520/stmsi.v13i3.4092>
- Nguyễn, H., Võ, V., Phan, T., Trần, D., Lê, M., & Nguyễn, V. (2025). Developing a web-based application for cervical cancer screening via visual inspection: enhancing training and clinical practice in vietnam. *Tạp Chí Y Học Việt Nam*, 550(1). <https://doi.org/10.51298/vmj.v550i1.13873>
- Ogunpola, A., Saeed, F., Basurra, S., Albarrak, A., & Qasem, S. (2024). Machine learning-based predictive models for detection of cardiovascular diseases. *Diagnostics*, 14(2), 144. <https://doi.org/10.3390/diagnostics14020144>
- Ozturk, B., Lawton, T., Smith, S., & Habli, I. (2023). Predicting progression of type 2 diabetes using primary care data with the help of machine learning.. <https://doi.org/10.3233/shti230060>
- Paksi, A., Hafidhoh, N., & Bimonugroho, S. (2023). Perbandingan model pengembangan perangkat lunak untuk proyek tugas akhir program vokasi. *Jurnal Masyarakat Informatika*, 14(1), 70-79. <https://doi.org/10.14710/jmasif.14.1.52752>

- Pamutha, T., Promthong, W., & Pahlawan, S. (2025). Diabetes disease prediction using machine learning classification algorithms. *International Journal on Advanced Science Engineering and Information Technology*, 15(4), 1060-1069. <https://doi.org/10.18517/ijaseit.15.4.20457>
- Pargaonkar, S. (2023). A comprehensive research analysis of software development life cycle (sdlc) agile & waterfall model advantages, disadvantages, and application suitability in software quality engineering. *International Journal of Research and Studies Publishing*, 13(8), 120-124. <https://doi.org/10.29322/ijrsrp.13.08.2023.p14015>
- Pargent, F., Pfisterer, F., Thomas, J., & Bischl, B. (2022). Regularized target encoding outperforms traditional methods in supervised machine learning with high cardinality features. *Computational Statistics*, 37(5), 2671-2692. <https://doi.org/10.1007/s00180-022-01207-6>
- Pei, X., Deng, Q., Liu, Z., Yan, X., & Sun, W. (2021). Machine learning algorithms for predicting fatty liver disease. *Annals of Nutrition and Metabolism*, 77(1), 38-45. <https://doi.org/10.1159/000513654>
- Pernencar C., Aguilar, P., Saboia, I., et al. (2022). Systematic mapping of digital health apps—A methodological proposal based on the World Health Organization classification of interventions. *Digital Health*, 8. <https://doi.org/10.1177/20552076221129071>
- Ploton, P., Mortier, F., Réjou-Méchain, M., Barbier, N., Picard, N., Rossi, V., ... & Pélissier, R. (2020). Spatial validation reveals poor predictive performance of large-scale ecological mapping models. *Nature Communications*, 11(1). <https://doi.org/10.1038/s41467-020-18321-y>
- Poly, T., Islam, M., & Li, Y. (2022). Early diabetes prediction: a comparative study using machine learning techniques. <https://doi.org/10.3233/shti220752>
- Pradipta, G., Liandana, M., Ayu, P., Hostiadi, D., & Putra, P. (2023). Voting scheme nearest neighbors by difference distance metrics measurement. *Juita Jurnal Informatika*, 11(2), 165. <https://doi.org/10.30595/juita.v11i2.19298>
- Pramana, P., Mayetti, M., & Kadri, H. (2013). Hubungan antara proteinuria dan hipoalbuminemia pada anak dengan sindrom nefrotik yang dirawat di rsup dr. m. djamil padang periode 2009-2012. *Jurnal Kesehatan Andalas*, 2(2), 90. <https://doi.org/10.25077/jka.v2i2.127>
- Preetha, V. and Raghu, C. (2025). Kullback-leibler divergence in feature selection: a methodology for improved detection of heart valve disorders.. <https://doi.org/10.21203/rs.3.rs-5966463/v1>

- Pulungan, A., Zarlis, M., & Suwilo, S. (2019). Analysis of braycurtis, canberra and euclidean distance in knn algorithm. *Sinkron*, 4(1), 74. <https://doi.org/10.33395/sinkron.v4i1.10207>
- Purbaningsih, W. and Dewi, M. (2024). Karakteristik demografi dan klinis pasien demam berdarah dengue (dbd) anak di rsud al-ihsan tahun 2022. *Bandung Conference Series Medical Science*, 4(1). <https://doi.org/10.29313/bcsms.v4i1.12324>
- Purnamasari, R., Rakhmawatie, M., Diatri, D., & Rohmani, A. (2023). Skrining diabetes mellitus melalui pemeriksaan dan konsultasi hasil gula darah di rumah pelayanan sosial. *Jurnal Inovasi Dan Pengabdian Masyarakat Indonesia*, 2(4), 24-26. <https://doi.org/10.26714/jipmi.v2i4.161>
- Qiu, L., Qu, Y., Shang, C., Yang, L., Chao, F., & Shen, Q. (2021). Exclusive lasso-based k-nearest-neighbor classification. *Neural Computing and Applications*, 33(21), 14247-14261. <https://doi.org/10.1007/s00521-021-06069-5>
- Rachakonda, S., Pudipedi, S., & Angel, T. (2024). Predictive modelling for diabetes using machine learning. *Interantional Journal of Scientific Research in Engineering and Management*, 08(008), 1-16. <https://doi.org/10.55041/ijsrem37149>
- Rachma, N. and Muhlas, I. (2022). Comparison of waterfall and prototyping models in research and development (r&d) methods for android-based learning application design. *Jurnal Inovatif Inovasi Teknologi Informasi Dan Informatika*, 5(1), 36. <https://doi.org/10.32832/inovatif.v5i1.7927>
- Rahim, M., Hossain, M., Hossain, M., Shin, J., & Yun, K. (2023). Stacked ensemble-based type-2 diabetes prediction using machine learning techniques. *Annals of Emerging Technologies in Computing*, 7(1), 30-39. <https://doi.org/10.33166/aetic.2023.01.003>
- Rahman, A. and Prasetyowati, S. (2022). Performance analysis of the hybrid voting method on the classification of the number of cases of dengue fever. *International Journal on Information and Communication Technology (Ijoict)*, 8(1), 10-19. <https://doi.org/10.21108/ijoict.v8i1.614>
- Rahman, M., Handayani, R., Cahyo, D., & Utami, E. (2023). Comparative analysis of software development lifecycle methods in software development: a systematic literature review. *International Journal of Advances in Data and Information Systems*, 4(2), 97-106. <https://doi.org/10.25008/ijadis.v4i2.1295>
- Rath, S., Pandey, M., & Rautaray, S. (2025). Synergistic review of automation impact of big data, ai, and ml in current data transformative era. *F1000research*, 14, 253. <https://doi.org/10.12688/f1000research.161477.2>

- Regina, O. (2024). Artificial intelligence in diabetes care: transforming diagnosis, management, and research- a mini review. *IDOSR JCAS*, 9(1), 11-14. <https://doi.org/10.59298/jcas/2024/91.1114000>
- Roberts, M. (2018). Successful public health information system database integration projects: a qualitative study. *Online Journal of Public Health Informatics*, 10(2). <https://doi.org/10.5210/ojphi.v10i2.9221>
- Rosenberger, D., Blechschmidt, V., Timmerman, H., Wolff, A., & Treede, R. (2020). Challenges of neuropathic pain: focus on diabetic neuropathy. *Journal of Neural Transmission*, 127(4), 589-624. <https://doi.org/10.1007/s00702-020-02145-7>
- Roy, S., Mandal, P., Chowdhury, A., Abdullah-Al-Wadud, M., Seikh, A., Seikh, A., ... & Mukhopadhyay, A. (2024). Utilizing machine learning algorithms to predict accuracy of the index of relative tectonic activity (irta), dhansiri (north) river basin in india and bhutan. *Ieee Access*, 12, 60482-60495. <https://doi.org/10.1109/access.2024.3394061>
- Rufo, D., Debelee, T., & Negera, W. (2022). A hybrid machine learning model based on global and local learner algorithms for diabetes mellitus prediction. *Journal of Biomimetics Biomaterials and Biomedical Engineering*, 54, 65-88. <https://doi.org/10.4028/www.scientific.net/jbbbe.54.65>
- Ruku, D., Pitoy, F., & Paral, M. (2022). Faktor risiko kejadian diabetes melitus tipe ii pada masyarakat desa lilang minahasa utara. *Klabat Journal of Nursing*, 4(2), 17. <https://doi.org/10.37771/kjn.v4i2.826>
- Ruparelia, N. (2010). Software development lifecycle models. *Acm Sigsoft Software Engineering Notes*, 35(3), 8-13. <https://doi.org/10.1145/1764810.1764814>
- Rusli, A., Suryadibrata, A., Nusantara, S., & Young, J. (2020). A comparison of traditional machine learning approaches for supervised feedback classification in bahasa indonesia. *Ijnmt (International Journal of New Media Technology)*, 7(1), 28-32. <https://doi.org/10.31937/ijnmt.v1i1.1485>
- Sahid, M., Babar, M., & Uddin, M. (2024). Predictive modeling of multi-class diabetes mellitus using machine learning and filtering iraqi diabetes data dynamics. *Plos One*, 19(5), e0300785. <https://doi.org/10.1371/journal.pone.0300785>
- Sajib, A., Shafiullah, A., & Sumon, A. (2013). An alternative algorithm for classification based on robust mahalanobis distance. *Dhaka University Journal of Science*, 61(1), 81-85. <https://doi.org/10.3329/dujs.v61i1.15101>
- Sandhu, G., Singh, A., Lamba, P., Virmani, D., & Chaudhary, G. (2023). Modified euclidean-canberra blend distance metric for knn classifier. *Intelligent Decision Technologies*, 17(2), 527-541. <https://doi.org/10.3233/idt-220233>

- Saputra, M. (2025). Web-based software development: diabetes mellitus prediction., 27, 343-355. <https://doi.org/10.4028/p-ey4cev>
- Sarker, I. (2021). Machine learning: algorithms, real-world applications and research directions. *Sn Computer Science*, 2(3). <https://doi.org/10.1007/s42979-021-00592-x>
- Sawarkar, G. and Rajput, D. (2022). Performance evolution of software engineering development and project management: a critical evolution software engineering process. *International Journal of Advanced Research in Science Communication and Technology*, 389-398. <https://doi.org/10.48175/ijarsct-5993>
- Shah, U., Jinwala, D., & Patel, S. (2016). An excursion to software development life cycle models. *Acm Sigsoft Software Engineering Notes*, 41(1), 1-6. <https://doi.org/10.1145/2853073.2853080>
- Shahraki, H., Pourahmad, S., & Zare, N. (2017). k important neighbors: a novel approach to binary classification in high dimensional data. *Biomed Research International*, 2017, 1-9. <https://doi.org/10.1155/2017/7560807>
- Shameer, K., Johnson, K., Glicksberg, B., Dudley, J., & Sengupta, P. (2018). Machine learning in cardiovascular medicine: are we there yet?. *Heart*, 104(14), 1156-1164. <https://doi.org/10.1136/heartjnl-2017-311198>
- Shanak, S., Saad, B., & Zaid, H. (2019). Metabolic and epigenetic action mechanisms of antidiabetic medicinal plants. *Evidence-Based Complementary and Alternative Medicine*, 2019, 1-18. <https://doi.org/10.1155/2019/3583067>
- Sherif, E., Helmy, W., & Galal-Edeen, G. (2023). Proposed framework to manage non-functional requirements in agile. *Ieee Access*, 11, 53995-54005. <https://doi.org/10.1109/access.2023.3281195>
- Shigetome, K., Egashira, T., Tomita, T., Higa, N., Iwashita, K., Morita, K., ... & Saruwatari, J. (2025). Effect of cumulative exposure on the efficacy of paroxetine: a population pharmacokinetic-pharmacodynamic and machine learning analyses. *CPT Pharmacometrics & Systems Pharmacology*, 14(6), 1119-1127. <https://doi.org/10.1002/psp4.70032>
- Shukur, B., Yaacob, N., & Doheir, M. (2023). Diabetes at a glance: assessing ai strategies for early diabetes detection and intervention. *MJAIH*, 2023, 85-89. <https://doi.org/10.58496/mjaih/2023/017>
- Sinclair, S. and Schwartz, S. (2019). Diabetic retinopathy—an underdiagnosed and undertreated inflammatory, neuro-vascular complication of diabetes. *Frontiers in Endocrinology*, 10. <https://doi.org/10.3389/fendo.2019.00843>
- Sitienei, M., Otieno, A., & Anapapa, A. (2023). An application of k-nearest-neighbor regression in maize yield prediction. *Asian Journal of Probability and Statistics*, 24(4), 1-10. <https://doi.org/10.9734/ajpas/2023/v24i4529>

- Sowabi, N., Widiastuti, N., & Maori, N. (2024). Optimasi algoritma k-nearest neighbors menggunakan teknik bayesian optimization untuk klasifikasi diabetes. *Journal of Information System Research (Josh)*, 6(1), 283-290. <https://doi.org/10.47065/josh.v6i1.5975>
- Štiglic, G., Kocbek, P., Cilar, L., Fijačko, N., Stožer, A., Zaletel, J., ... & Povalej, P. (2018). Development of a screening tool using electronic health records for undiagnosed type 2 diabetes mellitus and impaired fasting glucose detection in the slovenian population. *Diabetic Medicine*, 35(5), 640-649. <https://doi.org/10.1111/dme.13605>
- Sun, A. and Sun, F. (2024). Prediction of diabetes using supervised classification. *J Emerg Invest*. <https://doi.org/10.59720/23-062>
- Sutresno, S. and Singgalen, Y. (2023). Analysis and design of morotai tourism village information system (sidewita) based on local wisdom of tokuwela and babari tradition. *Journal of Information Systems and Informatics*, 5(1), 174-185. <https://doi.org/10.51519/journalisi.v5i1.441>
- Tafa, Z., Pervetica, N., & Karahoda, B. (2015). An intelligent system for diabetes prediction., 378-382. <https://doi.org/10.1109/meco.2015.7181948>
- Tariq, U., Hu, Z., Tariq, R., Iqbal, M., & Sadiq, M. (2023). High-performance embedded system for offline signature verification problem using machine learning. *Electronics*, 12(5), 1243. <https://doi.org/10.3390/electronics12051243>
- Tatit, P., Adhinugraha, K., & Taniar, D. (2024). Navigating the maps: euclidean vs. road network distances in spatial queries. *Algorithms*, 17(1), 29. <https://doi.org/10.3390/a17010029>
- Tembusai, Z., Mawengkang, H., & Zarlis, M. (2021). K-nearest neighbor with k-fold cross validation and analytic hierarchy process on data classification. *International Journal of Advances in Data and Information Systems*, 2(1). <https://doi.org/10.25008/ijadis.v2i1.1204>
- Theerthagiri, P., Ruby, A., & Vidya, J. (2022). Diagnosis and classification of the diabetes using machine learning algorithms. *Sn Computer Science*, 4(1). <https://doi.org/10.1007/s42979-022-01485-3>
- Tomašev, N. and Búza, K. (2015). Hubness-aware knn classification of high-dimensional data in presence of label noise. *Neurocomputing*, 160, 157-172. <https://doi.org/10.1016/j.neucom.2014.10.084>
- Torkar, R., Minoves, P., & Garrigós, J. (2011). Adopting free/libre/open source software practices, techniques and methods for industrial use. *Journal of the Association for Information Systems*, 12(1), 88-122. <https://doi.org/10.17705/1jais.00252>

- Trabelsi, A., Elouedi, Z., & Lefèvre, É. (2017). A novel k-nn approach for data with uncertain attribute values., 160-170. [https://doi.org/10.1007/978-3-319-60042-0\\_19](https://doi.org/10.1007/978-3-319-60042-0_19)
- Triguero, I., García-Gil, D., Maillou, J., Luengo, J., García, S., & Herrera, F. (2018). Transforming big data into smart data: an insight on the use of the k-nearest neighbors algorithm to obtain quality data. *Wiley Interdisciplinary Reviews Data Mining and Knowledge Discovery*, 9(2). <https://doi.org/10.1002/widm.1289>
- Uddin, S., Haque, I., Lu, H., Moni, M., & Gide, E. (2022). Comparative performance analysis of k-nearest neighbour (knn) algorithm and its different variants for disease prediction. *Scientific Reports*, 12(1). <https://doi.org/10.1038/s41598-022-10358-x>
- Uddin, S., Khan, A., Hossain, M., & Moni, M. (2019). Comparing different supervised machine learning algorithms for disease prediction. *BMC Medical Informatics and Decision Making*, 19(1). <https://doi.org/10.1186/s12911-019-1004-8>
- Wahyuni, M., Rosnelly, R., & Wanayumini, W. (2023). Combination of pre-trained cnn model and machine learning algorithm on pekalongan batik motif classification. *International Conference on Information Science and Technology Innovation (Icostec)*, 2(1), 23-28. <https://doi.org/10.35842/icostec.v2i1.31>
- Wang, L., Wang, X., Chen, A., Jin, X., & Che, H. (2020). Prediction of type 2 diabetes risk and its effect evaluation based on the xgboost model. *Healthcare*, 8(3), 247. <https://doi.org/10.3390/healthcare8030247>
- Wang, Z., Wang, S., Wang, M., & Sun, Y. (2024). Design of application-oriented disease diagnosis model using a meta-heuristic algorithm. *Technology and Health Care*, 32(6), 4041-4061. <https://doi.org/10.3233/thc-231755>
- Weng, S., Reys, J., Kai, J., Garibaldi, J., & Qureshi, N. (2017). Can machine-learning improve cardiovascular risk prediction using routine clinical data?. *Plos One*, 12(4), e0174944. <https://doi.org/10.1371/journal.pone.0174944>
- Widyana, A. and Afriansyah, M. (2022). Penyuluhan dan pemeriksaan kadar hba1c pada pasien diabetes mellitus tipe 2 di rsud suradadi. *Jurnal Inovasi Dan Pengabdian Masyarakat Indonesia*, 1(3), 6-9. <https://doi.org/10.26714/jipmi.v1i3.23>
- Woo, Y., Lee, C., Fong, C., Tso, A., Cheung, B., & Lam, K. (2017). Validation of the diabetes screening tools proposed by the american diabetes association in an aging chinese population. *Plos One*, 12(9), e0184840. <https://doi.org/10.1371/journal.pone.0184840>
- Wu, T., Su, G., Liu, T., Wang, H., & Hwu, C. (2023). Urinary acrolein protein conjugates-to-creatinine ratio is positively associated with diabetic

- peripheral neuropathy in patients with type 2 diabetes mellitus. *Endocrine Connections*, 12(11). <https://doi.org/10.1530/ec-23-0253>
- Wu, Y., Ding, Y., Tanaka, Y., & Zhang, W. (2014). Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. *International Journal of Medical Sciences*, 11(11), 1185-1200. <https://doi.org/10.7150/ijms.10001>
- Wu, Y., Hu, H., Cai, J., Chen, R., Zuo, X., Cheng, H., ... & Yan, D. (2021). Machine learning for predicting the 3-year risk of incident diabetes in chinese adults. *Frontiers in Public Health*, 9. <https://doi.org/10.3389/fpubh.2021.626331>
- Xiao, C. and Chaovallitwongse, W. (2016). Optimization models for feature selection of decomposed nearest neighbor. *IEEE Transactions on Systems Man and Cybernetics Systems*, 46(2), 177-184. <https://doi.org/10.1109/tsmc.2015.2429637>
- Xie, Z., Nikolayeva, O., Luo, J., & Li, D. (2019). Building risk prediction models for type 2 diabetes using machine learning techniques. *Preventing Chronic Disease*, 16. <https://doi.org/10.5888/pcd16.190109>
- Xu, C., Shi, F., Ding, W., Fang, C., & Fang, C. (2025). Development and validation of a machine learning model for cardiovascular disease risk prediction in type 2 diabetes patients. *Scientific Reports*, 15(1). <https://doi.org/10.1038/s41598-025-18443-7>
- Xu, H., Przystupa, K., Fang, C., Marciniak, A., К о ч а н , О ., & Beshley, M. (2020). A combination strategy of feature selection based on an integrated optimization algorithm and weighted k-nearest neighbor to improve the performance of network intrusion detection. *Electronics*, 9(8), 1206. <https://doi.org/10.3390/electronics9081206>
- Xu, X., Yu, Z., Ge, Z., Chow, E., Bao, Y., Ong, J., ... & Zhang, L. (2022). Web-based risk prediction tool for an individual's risk of hiv and sexually transmitted infections using machine learning algorithms: development and external validation study. *Journal of Medical Internet Research*, 24(8), e37850. <https://doi.org/10.2196/37850>
- Yang, T., Qi, F., Guo, F., et al. (2024). An update on chronic complications of diabetes mellitus: from molecular mechanisms to therapeutic strategies with a focus on metabolic memory. *Molecular Medicine*, 30, 71. <https://doi.org/10.1186/s10020-024-00824-9>
- Yu, W., Liu, T., Valdéz, R., Gwinn, M., & Khoury, M. (2010). Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes. *BMC Medical Informatics and Decision Making*, 10(1). <https://doi.org/10.1186/1472-6947-10-16>

- Yun, K., He, T., Zhen, S., Quan, M., Yang, X., Man, D., ... & Han, X. (2023). Development and validation of explainable machine-learning models for carotid atherosclerosis early screening. *Journal of Translational Medicine*, 21(1). <https://doi.org/10.1186/s12967-023-04093-8>
- Zadeh, F., Koh, R., Dilek, B., Masani, K., & Kumbhare, D. (2023). Identification of myofascial trigger point using the combination of texture analysis in b-mode ultrasound with machine learning classifiers. *Sensors*, 23(24), 9873. <https://doi.org/10.3390/s23249873>
- Zaidi, N., Squire, D., & Suter, D. (2010). Boostml: an adaptive metric learning for nearest neighbor classification., 142-149. [https://doi.org/10.1007/978-3-642-13657-3\\_17](https://doi.org/10.1007/978-3-642-13657-3_17)
- Zhang, C., Fei-Fei, L., & Jestes, J. (2012). Efficient parallel knn joins for large data in mapreduce., 38-49. <https://doi.org/10.1145/2247596.2247602>
- Zhang, J., Hu, X., Li, P., He, W., Zhang, Y., & Li, H. (2014). A hybrid feature selection approach by correlation-based filters and svm-rfe., 3684-3689. <https://doi.org/10.1109/icpr.2014.633>
- Zhang, L., Huang, Y., Huang, M., Zhao, C., Zhang, Y., & Wang, Y. (2024). Development of cost-effective fatty liver disease prediction models in a chinese population: statistical and machine learning approaches. *Jmir Formative Research*, 8, e53654. <https://doi.org/10.2196/53654>
- Zhang, W. and Zhang, Z. (2025). Application of irsa-bp neural network in diagnosing diabetes. *Plos One*, 20(6), e0324759. <https://doi.org/10.1371/journal.pone.0324759>
- Zhao, Y., Healy, B., Rotstein, D., Guttman, C., Bakshi, R., Weiner, H., ... & Chitnis, T. (2017). Exploration of machine learning techniques in predicting multiple sclerosis disease course. *Plos One*, 12(4), e0174866. <https://doi.org/10.1371/journal.pone.0174866>
- Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Frontiers in Genetics*, 9. <https://doi.org/10.3389/fgene.2018.00515>
- Δ ρ ί τ σ α ς, H. and Τρίγκα, Μ. (2022). Data-driven machine-learning methods for diabetes risk prediction. *Sensors*, 22(14), 5304. <https://doi.org/10.3390/s22145304>
- Δ ρ ί τ σ α ς, H. and Τρίγκα, Μ. (2022). Data-driven machine-learning methods for diabetes risk prediction. *Sensors*, 22(14), 5304. <https://doi.org/10.3390/s22145304>

## PROFIL PENULIS

### **Fahmi Ruziq, S.T., M.Kom.**



Penulis lahir di Banda Aceh, Aceh tanggal 16 Juni 1989. Penulis menamatkan pendidikan dasar dan menengah di Banda Aceh, setelah lulus dari SMA Negeri 1 Banda Aceh melanjutkan kuliah S1 di Universitas Serambi Mekkah Jurusan Teknik Informatika, kemudian melanjutkan S2 di Universitas Sumatera Utara Jurusan Teknik Informatika. Kemudian saat ini Penulis sedang melanjutkan Pendidikan Doktor (S-3) Ilmu Komputer di Universitas Sumatera Utara (*on-going*). Berkarir sebagai dosen dimulai dari tahun 2020 di Universitas Battuta, Medan.

### **M. Rhifky Wayahdi, S.Kom., M.Kom.**



Penulis lahir di Banda Aceh, Aceh tanggal 16 Juni 1989. Penulis menamatkan pendidikan dasar dan menengah di Banda Aceh, setelah lulus dari SMA Negeri 1 Banda Aceh melanjutkan kuliah S1 di Universitas Serambi Mekkah Jurusan Teknik Informatika, kemudian melanjutkan S2 di Universitas Sumatera Utara Jurusan Teknik Informatika. Kemudian saat ini Penulis berkarir sebagai dosen dimulai dari tahun 2020 di Universitas Battuta, Medan.