JURNAL ILMIAH SISTEM INFORMASI (JUSI)
VOL 5, No. 2, Mei 2026, pp. 429 - 441
p-ISSN : 2803-1507 (media cetak) | e-ISSN : 2803-1531 (media online)
Link Publikasi: https://ejurnal.provisi.ac.id/index.php/JUISI/issue/view/171

Decision-Support AI Health Dashboard Using Synthetic Data for Daily Health Monitoring

Humaidi Fikri*[1], Muchlis[2]
[1,2] STMIK Antar Bangsa
Jl. HOS Cokroaminoto No. 29-35 Blok A5, 15157, Tangerang, Banten, Indonesia
e-mail: humaidifikri17@gmail.com*[1], Muchlis.re@gmail.com[2]

ARTICLE INFO

Article History:

Received: 18-02-2026
Revised: 16-03-2026
Accepted: 12-04-2026
Published: 21-05-2026

ABSTRACT

Digital health dashboards are increasingly being developed to support healthy lifestyles through data and Artificial Intelligence (AI) technology. However, most existing dashboards focus too much on visualizations and do not provide users with insight, so users often struggle to understand the meaning of the visualizations and do not know what steps to take next. This study aims to develop an AI-based health dashboard for decision-making by integrating the predictive capabilities of Machine Learning (ML) models with the interpretive capabilities of Large Language Models (LLMs). Synthetic data are used to represent a user's daily health data, such as body mass index trends, calorie and nutrient intake, sleep duration, and stress levels. A regression model was selected to identify patterns, trends, and the user's health condition based on this data. The ML model's predictions can serve as early warning signals for users, but in this study, the ML model was integrated into the LLM for context-based interpretation. The LLM model then generates descriptive and prescriptive health explanations and insights that are easier for users to understand and act upon. The results of the study show that integrating ML and LLM models can enhance dashboard functionality, transforming dashboards from mere graphical visualization tools into decision-making systems. Although this study still uses synthetic data, the methods used have the potential to be further developed using real health data.

Keywords: *Artificial Intelligence, Health Dashboard, Large Language Model, Synthetic Data, Decision Support.*

Abstrak

Pengembangan *dashboard* kesehatan berbasis digital semakin banyak dikembangkan sebagai sarana untuk mendukung gaya hidup sehat melalui pemanfaatan data dan teknologi *Artificial Intelligence* (AI). Namun, sebagian besar *dashboard* yang ada terlalu berfokus pada grafik visualisasi dan tidak memberikan *insight* kepada pengguna, sehingga pengguna sering kali kesulitan memahami makna dari grafik visualisasi yang ada di dashboard serta pengguna tidak tahu langkah apa yang harus dilakukan selanjutnya. Penelitian ini bertujuan untuk mengembangkan *dashboard* kesehatan berbasis AI untuk pengambilan sebuah keputusan dengan mengintegrasikan hasil prediksi model *Machine Learning* (ML) dan kemampuan interpretasi *Large Language Model* (LLM). Data sintetis digunakan untuk merepresetasikan data kesehatan harian seorang pengguna, seperti tren indeks massa tubuh, asupan kalori, asupan nutrisi, durasi tidur dan tingkat stres. Model regresi dipilih untuk mengidentifikasi pola, tren, dan kondisi kesehatan pengguna berdasarkan data tersebut. Hasil prediksi dari model ML bisa menjadi sebuah sinyal awal untuk pengguna, namun dalam penelitian ini model ML diintegrasikan ke dalam model LLM untuk dilakukan interpretasi berbasis konteks. Model LLM

DOI: https://doi.org/10.51903/mp66jy72

p-ISSN : 2803-1507 e-ISSN : 2803-1531

kemudian mengeluarkan penjelasan dan *insight* kesehatan yang bersifat deskriptif dan preskriptif sehingga lebih mudah dipahami dan ditindaklanjuti oleh pengguna. Hasil penelitian menunjukkan bahwa integrasi model ML dan model LLM mampu meningkatkan fungsi dashboard dari sekadar alat visualisasi grafik menjadi sistem pengambilan keputusan. Meskipun penelitian ini masih menggunakan data sintetis, metode yang digunakan memiliki potensi untuk dikembangkan lebih lanjut dengan menggunakan data kesehatan yang nyata.

Kata Kunci: *Artificial Intelligence, Dashboard Kesehatan, Large Language Model, Data Sintetis, Pengambilan Keputusan.*

1. PENDAHULUAN

Seiring dengan perkembangan industri teknologi dan perubahan gaya hidup masyarakat modern, *dashboard* kesehatan menjadi kebutuhan yang semakin penting. Perkembangan *Artificial Intelligence* (AI) dan *Machine Learning* (ML) telah membuka peluang besar dalam pengolahan data kesehatan untuk mendukung pengambilan keputusan berbasis data secara personal pada waktu yang bersamaan [1], [2], [3]. Banyak sistem pemantauan kesehatan *digital* yang sudah dikembangkan dalam bentuk *dashboard* interaktif yang mampu menampilkan indikator kesehatan seperti berat badan, *Body Mass Index* (BMI), asupan kalori, dan aktivitas harian. Namun, sebagian besar *dashboard* tersebut masih berfokus pada penyajian data numerik dan visualisasi statistik tanpa memberikan penjelasan mengenai makna dan implikasi dari visualisasi pada dashboard tersebut. [4], [5].

Fenomena di lapangan menunjukkan bahwa rendahnya literasi data kesehatan dan kompleksitas *output* model ML menjadi hambatan utama dalam pemanfaatan sistem kesehatan berbasis AI. Studi terkini menunjukkan bahwa pengguna non-teknis sering kali kesulitan memahami hasil prediksi atau skor yang dihasilkan oleh model prediktif, sehingga mengurangi kepercayaan dan efektivitas sistem tersebut dalam mendukung perubahan perilaku kesehatan [6]. Selain itu, isu privasi data pribadi dan keterbatasan akses terhadap data kesehatan pribadi mendorong banyak penelitian untuk menggunakan data sintetis sebagai alternatif yang aman dan etis dalam pengembangan sistem kesehatan *digital* [7], [8]. Meskipun data sintetis dinilai mampu merepresentasikan pola statistik data nyata, tantangan interpretasi hasil model tetap menjadi persoalan yang belum terpecahkan secara memadai.

Sudah banyak penelitian yang membahas penerapan ML dalam prediksi indikator kesehatan, seperti prediksi berat badan, risiko obesitas, atau pola nutrisi, dengan menggunakan berbagai algoritma regresi dan klasifikasi [9], [10], [11]. Di sisi lain, kajian terbaru juga mulai mengembangkan potensi *Large Language Model* (LLM) dalam ruang lingkup kesehatan, terutama sebagai alat bantu *clinical decision support* dan *Explainable AI* (XAI) [12], [13]. Namun, sebagian besar penelitian tersebut menjadikan LLM hanya sebatas fitur tambahan atau antarmuka percakapan, tanpa integrasi langsung dengan *output* model ML yang digunakan dalam sistem analitik kesehatan. Akibatnya, potensi LLM sebagai jembatan kognitif antara hasil komputasi dan pemahaman pengguna tidak bisa dimanfaatkan secara maksimal.

Berdasarkan observasi tersebut, dapat diidentifikasi adanya *research gap* pada aspek integrasi antara model prediktif ML dan LLM dalam sistem *dashboard* kesehatan yang berorientasi pada pemahaman pengguna. Penelitian sebelumnya lebih menekankan pada peningkatan akurasi model atau perbandingan algoritma, sementara aspek interpretasi kontekstual hasil prediksi masih kurang mendapat perhatian. Selain itu, belum banyak penelitian yang secara terang-terangan merancang model LLM sebagai komponen inti yang bertugas menjelaskan, menafsirkan, dan mengontekstualisasikan hasil prediksi ML berdasarkan data kesehatan pengguna. *Gap* ini menunjukkan perlunya pendekatan baru yang tidak hanya akurat secara teknis, tetapi juga komunikatif dan mudah dipahami oleh pengguna awam.

Oleh karena itu, penelitian ini bertujuan untuk mengembangkan *dashboard* kesehatan berbasis AI yang mengintegrasikan model regresi ML dengan LLM sebagai modul penjelasan otomatis terhadap hasil prediksi. Kontribusi utama penelitian ini terletak pada perancangan arsitektur sistem yang memosisikan LLM sebagai solusi interpretasi, bukan sekadar fitur tambahan, dalam menjelaskan output numerik model ML. Pendekatan ini diharapkan dapat meningkatkan pemahaman pengguna terhadap kondisi kesehatannya serta mendukung pengambilan keputusan yang lebih berbasis informasi. Selain itu, penelitian juga memberikan kontribusi metodologis dalam pemanfaatan data sintetis sebagai saran pengembangan dan evaluasi sistem kesehatan digital yang aman dan beretika. Penelitian ini membantu pengembangan sistem-sistem pendukung keputusan berbasis AI untuk kesehatan pribadi. Berbeda dengan penelitian sebelumnya yang fokus pada peningkatan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

akurasi model prediksi atau visualisasi data, penelitian ini mengusulkan kombinasi model ML dan LLM sebagai mekanisme interpretasi berbasis bahasa alami. Pendekatan ini bertujuan membuat *dashboard* kesehatan lebih dari sekadar alat visualisasi. Tujuannya adalah menciptakan sistem pendukung keputusan yang dapat memberikan penjelasan yang relevan dan berguna bagi pengguna.

2. TINJAUAN PUSTAKA

2.1. *Sistem Dashboard Kesehatan dan Peran Machine Learning*

Perkembangan sistem *dashboard* kesehatan *digital* merupakan respons terhadap meningkatnya kebutuhan pemantauan kesehatan secara berkelanjutan dan personal. Implementasi *dashboard* kesehatan berperan dalam mengintegrasikan berbagai indikator kesehatan individu ke dalam visualisasi yang ringkas dan mudah diakses. Penelitian terbaru menunjukkan bahwa integrasi ML dalam *dashboard* kesehatan mampu berdasarkan data historis pengguna [14]. Algoritma ML memungkinkan sistem tidak hanya bersifat deskriptif, tetapi juga prediktif, sehingga berpotensi memberikan peringatan dini terhadap perubahan kondisi kesehatan.

Berbagai pendekatan ML telah diterapkan dalam sistem kesehatan *digital*, termasuk regresi, klasifikasi, dan *ensemble learning*, untuk memprediksi indikator seperti indeks massa tubuh, tingkat aktivitas fisik, dan risiko penyakit kronis [15]. Meskipun demikian, sebagian besar penelitian memfokuskan evaluasi pada aspek performa model, seperti akurasi dan kesalahan prediksi. Penyajian hasil analisis ini sering kali terbatas pada angka atau grafik tanpa penjelasan yang memadai mengenai implikasi kesehatan dari hasil tersebut. Kondisi ini menyebabkan pengguna awam kesulitan memahami dan memanfaatkan informasi yang disediakan oleh sistem [16]. Keterbatasan tersebut menunjukkan bahwa keberhasilan *dashboard* kesehatan tidak hanya ditentukan oleh kecanggihan algoritma ML, tetapi juga oleh kemampuan sistem dalam menyampaikan makna hasil analisis secara jelas dan kontekstual. Oleh karena itu, diperlukan metode yang melampaui visualisasi data dan berfokus pada pengambilan keputusan berbasis pemahaman pengguna.

2.2. *Explainable Artificial Intelligence dalam Sistem Pendukung Keputusan Kesehatan*

XAI muncul sebagai metode untuk mengatasi sifat *black-box* dari banyak model ML, khususnya dalam ruang lingkup kesehatan yang menuntut transparansi dan kredibilitas tinggi. XAI bertujuan memberikan penjelasan yang dapat dipahami manusia mengenai bagaimana dan mengapa suatu sistem menghasilkan keputusan tertentu. Penelitian dalam beberapa tahun terakhir menunjukkan bahwa sistem pendukung keputusan kesehatan yang dilengkapi dengan mekanisme XAI lebih mudah diterima oleh pengguna dan praktisi kesehatan [17]. Berbagai teknik XAI telah dikembangkan, seperti *feature importance*, *local explanation*, dan visualisasi kontribusi variabel, untuk meningkatkan interpretabilitas model ML [18]. Namun, teknik-teknik tersebut umumnya dirancang untuk pengguna dengan latar belakang teknis dan kurang ramah bagi pengguna awam. Dalam konteks *dashboard* kesehatan personal, penjelasan berbasis grafik atau skor teknis sering kali belum cukup untuk membantu pengguna memahami kondisi kesehatannya secara menyeluruh. Studi terbaru menekankan bahwa penjelasan berbasis bahasa alami lebih efektif dalam meningkatkan pemahaman dan kepercayaan pengguna terhadap sistem AI dibandingkan dengan pendekatan visual atau numerik semata [19]. Hal ini membuka peluang integrasi teknologi AI yang mampu menghasilkan penjelasan naratif sebagai bagian dari sistem pendukung keputusan kesehatan.

2.3. *Large Language Model sebagai Pendekatan Interpretasi Berbasis Bahasa Alami*

Representasi LLM menandai kemajuan signifikan dalam bidang *Natural Language Processing* (NLP) melalui kemampuannya dalam memahami konteks dan menghasilkan teks yang koheren. Dalam ruang lingkup kesehatan, LLM mulai dieksplorasi untuk berbagai aplikasi, termasuk analisis teks medis, dokumentasi klinis, dan sistem pendukung keputusan [20]. Keunggulan utama LLM terletak pada kemampuannya menerjemahkan informasi kompleks menjadi penjelasan berbasis bahasa alami yang mudah dipahami. Beberapa penelitian menunjukkan bahwa LLM berpotensi digunakan sebagai lapisan interpretasi untuk menjelaskan hasil analisis ML secara kontekstual [21]. Namun, pemanfaatan LLM dalam *dashboard* kesehatan masih terbatas dan sering kali diposisikan hanya sebagai antarmuka percakapan, bukan sebagai komponen inti dalam alur analitik sistem. Selain itu, integrasi langsung antara *output* numerik model ML dan proses penjelasan berbasis LLM masih jarang dibahas secara sistematis. Keterbatasan ini menunjukkan adanya peluang penelitian untuk memanfaatkan LLM sebagai mekanisme interpretasi keputusan (*decision interpretation mechanism*) yang menjembatani hasil prediksi ML dan pemahaman pengguna. Dengan

p-ISSN : 2803-1507 e-ISSN : 2803-1531

pendekatan ini, LLM tidak hanya berfungsi sebagai alat komunikasi, tetapi juga sebagai solusi awal untuk meningkatkan kegunaan dan nilai praktis sistem *dashboard* kesehatan.

2.4. *Penggunaan Data Sintetis dalam Pengembangan Sistem Kesehatan Digital*

Penggunaan data sintetis menjadi solusi yang semakin relevan dalam pengembangan sistem kesehatan *digital*, terutama terkait isu privasi data dan keterbatasan akses data nyata. Data sintetis memungkinkan peneliti mengembangkan dan menguji sistem AI tanpa melanggar regulasi perlindungan data pribadi. Penelitian menunjukkan bahwa data sintetis dapat merepresentasikan pola statistik data kesehatan nyata secara memadai untuk tujuan pengembangan sistem berbasis ML [22]. Meskipun demikian, penggunaan data sintetis juga menimbulkan tantangan dalam interpretasi hasil analisis. Karena data tidak berasal dari pengukuran langsung, hasil prediksi ML memerlukan penjelasan kontekstual agar tetap relevan bagi pengguna [23]. Oleh karena itu, integrasi LLM sebagai lapisan interpretasi menjadi semakin penting dalam sistem berbasis data sintetis, guna memastikan hasil analisis dapat dipahami dan dimanfaatkan secara optimal.

2.5. *Gap Penelitian dan Posisi Studi*

Berdasarkan kajian pustaka yang telah dipaparkan, dapat disimpulkan bahwa penelitian sebelumnya masih berfokus pada peningkatan performa model ML atau penyajian visualisasi data dalam *dashboard* kesehatan. Aspek interpretasi hasil prediksi berbasis bahasa alami dan berorientasi pada pengguna masih relatif kurang dieksplorasi. Selain itu, pemanfaatan LLM sebagai komponen inti yang terintegrasi langsung dengan *output* analitik ML dalam *dashboard* kesehatan personal masih jarang ditemukan. Penelitian ini memposisikan diri untuk mengisi celah tersebut dengan mengusulkan pendekatan integratif yang menghubungkan model regresi ML dan LLM sebagai lapisan interpretasi keputusan. Keunikan pendekatan ini terletak pada pemanfaatan LLM untuk menjelaskan hasil prediksi numerik secara kontekstual, sehingga sistem dashboard kesehatan tidak hanya informatif, tetapi juga mendukung pemahaman dan pengambilan keputusan pengguna.

3. METODOLOGI PENELITIAN

3.1. *Jenis dan Metode Penelitian*

Penelitian ini menggunakan pendekatan *Research and Development* (R&D) dengan metode kuantitatif eksperimental. Pendekatan R&D dipilih karena penelitian berfokus pada perancangan, pengembangan, dan evaluasi sistem *dashboard* kesehatan berbasis AI. Model pengembangan sistem *dashboard* mengacu pada kerangka kerja rekayasa sistem iteratif yang sudah banyak digunakan dalam pengembangan sistem berbasis ML [24]. Penelitian ini menggunakan pendekatan kuantitatif untuk mengevaluasi seberapa baik model regresi dapat memprediksi perubahan berat badan berdasarkan variabel kesehatan pengguna. Evaluasi dilakukan menggunakan metrik statistik seperti *Mean Absolute Error* (MAE), *Mean Squared Error* (MSE), dan koefisien determinasi (R^2), yang sudah umum digunakan dalam evaluasi model regresi kesehatan [25]. Penelitian ini juga mengintegrasikan LLM sebagai mekanisme untuk memahami hasil prediksi numerik. Integrasi model prediktif dan model LLM mengikuti pendekatan sistem hibrida yang mulai berkembang dalam literatur *AI-enabled decision support systems* [26].

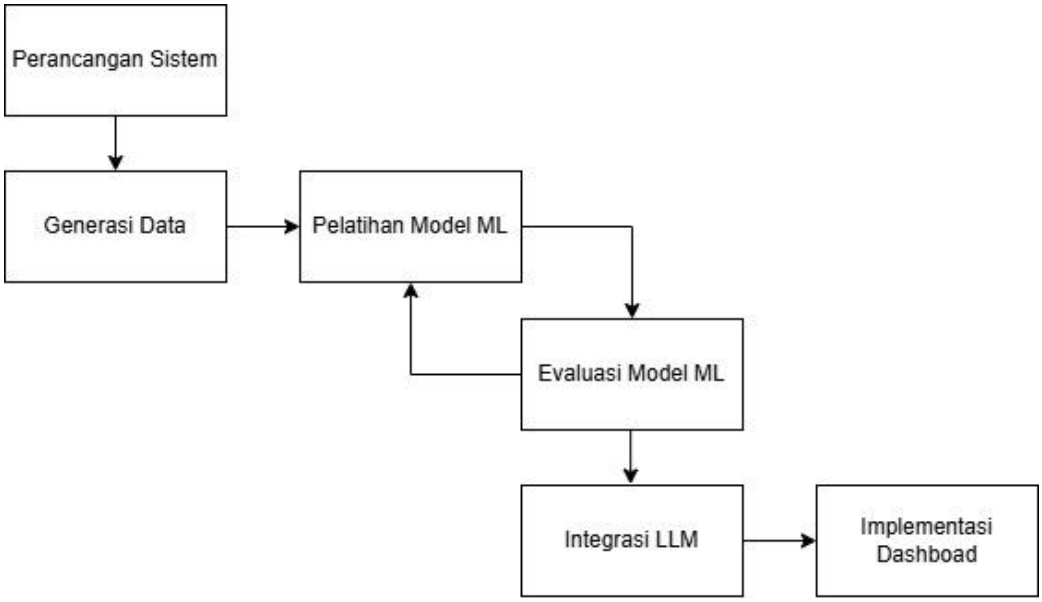
3.2. *Tahapan Penelitian*

Penelitian ini disusun dengan sangat rapi agar bisa diikuti oleh peneliti lain. Secara umum, ada lima langkah utama dalam penelitian ini:

1. Perancangan arsitektur sistem
2. Pembangunan dataset sintetis
3. Pengembangan dan pelatihan model regresi
4. Integrasi model regresi dengan LLM
5. Implementasi dan evaluasi *dashboard*

Alur penelitian dapat dilihat dalam Gambar 1.

p-ISSN : 2803-1507 e-ISSN : 2803-1531



Gambar 1. Alur Tahapan Penelitian

Setiap langkah dalam diagram tersebut akan dijelaskan satu per satu.

3.2.1. *Perancangan Arsitektur Sistem*

Tahap ini adalah tentang merancang struktur dasar sistem. Sistem ini terdiri dari tiga lapisan utama, yaitu:

1. Lapisan data
2. Lapisan analitik yang menggunakan ML
3. Lapisan interpretasi yang menggunakan LLM

Arsitektur sistem dirancang dengan pendekatan modular. Artinya, setiap komponen dapat diuji secara terpisah [27]. Hal ini memungkinkan pengujian yang lebih efektif dan efisien.

3.2.2. *Pembanguan Data Sintetis*

Dataset yang digunakan dalam penelitian ini adalah dataset sintetis yang dibuat untuk mensimulasikan kondisi kesehatan seseorang. Dataset ini terdiri dari 500 contoh data yang dihasilkan dengan menggunakan pendekatan distribusi statistik yang terkontrol, dengan mempertimbangkan hubungan logis antara variabel kesehatan.

Variabel yang digunakan dalam dataset ini meliputi berat badan, tinggi badan, BMI, asupan kalori harian, asupan nutrisi (protein, karbohidrat, lemak, dan gula), asupan air, tingkat stres, kualitas tidur, dan total waktu tidur. Rentang nilai untuk masing-masing variabel ini ditentukan berdasarkan literatur kesehatan yang ada, yaitu berat badan antara 50 hingga 60kg, tinggi badan ditetapkan 160cm, asupan kalori antara 2000 hingga 3500kal per hari, asupan protein antara 50 hingga 150gr per hari, asupan lemak antara 40 hingga 100gr per hari, asupan gula antara 20 hingga 100gr per hari, asupan karbohidrat antara 150 hingga 350gr per hari, asupan air antara 1 hingga 4l per hari, tingkat stres antara 1 hingga 10, tingkat kualitas tidur antara 1 hingga 10, dan total waktu tidur antara 4 hingga 8 jam.

Dataset sintetis dalam penelitian ini dikembangkan menggunakan pendekatan stokastik dengan batasan logis (*constrained randomness*). Meskipun variabel asupan nutrisi dan aktivitas harian memiliki unsur variabilitas untuk mensimulasikan fluktuasi perilaku manusia yang dinamis, distribusi data tetap dibatasi dalam rentang parameter klinis yang wajar. Hal ini dilakukan untuk memastikan bahwa hubungan antar variabel utama (seperti korelasi asupan kalori terhadap fluktuasi berat badan) tetap memiliki pola yang konsisten dan dapat dipelajari secara sistematis oleh model *machine learning*.

Dataset ini kemudian dibagi menjadi dua bagian menggunakan metode *train-test split* dengan rasio 80:20, yaitu 80% untuk proses pelatihan model dan 20% untuk proses pengujian model.

3.2.3. *Pengembangan Model Regresi*

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Model regresi linier dipakai untuk memahami hubungan antara variabel yang tidak bergantung dan perubahan berat badan. Cara ini dipilih karena hasilnya bisa dengan mudah dipahami dan sangat cocok untuk data numerik yang berkelanjutan [28].

Persamaan regresi linier dapat diterangkan pada Persamaan (1).

$$Y = \beta_0 + \beta_1 \mathrm{X}_1 + \beta_2 \mathrm{X}_2 + \cdots + \beta_n \mathrm{X}_n \quad (1)$$

Keterangan:

Y = Prediksi perubahan berat badan

$\mathrm{X}_n = Variabel\ independen$

$\beta_n = Koefisien\ regresi$

Evaluasi model dilakukan menggunakan Persamaan (1), Persamaan (2), dan Persamaan (3).

$$\mathrm{MAE} = \frac{1}{n}\sum|y_i - \hat{y}_i| \quad (2)$$

$$\mathrm{MAE} = \frac{1}{n}\sum(y_i - \hat{y}_i)^2 \quad (3)$$

$$\mathcal{R}^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2} \quad (4)$$

Metode validasi menggunakan teknik *train-test split* dengan rasio 80:20 untuk menghindari *overfitting* [29].

3.2.4. *Integrasi Large Language Model*

Model regresi menghasilkan *output* numerik yang kemudian dikirim ke LLM dalam bentuk *structured prompt* mengikuti metode yang sudah ada sebelumnya, yaitu *prompt-based explanation generation*, yang digunakan untuk menggabungkan model prediktif dan model LLM [21]. LLM memiliki beberapa fungsi, antara lain:

1. Menjelaskan hubungan antar variabel yang ada dalam data
2. Memberikan interpretasi kontekstual dari hasil prediksi
3. Menyederhanakan hasil numerik menjadi bahasa alami yang lebih mudah dipahami

3.2.5. *Implementasi Dashboard*

Dashboard ini dibuat dengan menggunakan bahasa pemrograman Python dan beberapa pustaka penting. Pustaka yang digunakan antara lain *Scikit-learn* yang berguna untuk membuat model regresi, *Pandas* untuk mengelola data, *Matplotlib* atau *Plotly* untuk membuat visualisasi data. Selain itu, sistem ini juga menggunakan *Application Programming Interface* (API) LLM yang berfungsi untuk mentafsirkan teks. Sistem ini dikembangkan dengan memperhatikan kebutuhan pengguna, sehingga disebut *human-centered* AI. Dengan pendekatan ini, diharapkan hasil analitik yang dihasilkan dapat dengan mudah dipahami oleh pengguna [30].

3.3. *Validasi dan Evaluasi Sistem*

Evaluasi dilakukan pada dua aspek:

1. Evaluasi Model Prediktif
   Berdasarkan beberapa ukuran, seperti MAE, MSE, dan R^2, kita dapat mengetahui seberapa baik model tersebut bekerja.
2. Evaluasi Interpretasi LLM
   Evaluasi interpretasi LLM dalam penelitian ini dilakukan menggunakan pendekatan LLM-*as-a-Judge*, yaitu dengan memanfaatkan model LLM sebagai evaluator untuk menilai kualitas *output* yang dihasilkan oleh model LLM lainnya. Pendekatan ini semakin banyak digunakan dalam penelitian terbaru karena mampu memberikan evaluasi yang konsisten dan skalabel terhadap kualitas teks yang dihasilkan oleh LLM [31]

3.4. *Replikasi dan Transparansi*

p-ISSN : 2803-1507 e-ISSN : 2803-1531

Proses pengembangan dilakukan secara menyeluruh dengan menggunakan bahasa pemrograman Python 3.x. Dataset sintetis dan kode untuk melatih model dapat direplikasi dengan menggunakan parameter yang telah dijelaskan sebelumnya. Untuk memudahkan pemahaman, dokumentasi tentang alur sistem disediakan dalam bentuk diagram arsitektur yang menjelaskan integrasi model LLM. Penelitian ini menggunakan data sintetis dan tidak melibatkan partisipan manusia secara langsung. Oleh karena itu, penelitian ini tidak memerlukan persetujuan etik. Meskipun demikian, pendekatan pengembangan tetap memperhatikan prinsip perlindungan data dan privasi yang sesuai dengan praktik penelitian kesehatan *digital* [23].

4. HASIL DAN PEMBAHASAN

4.1. *Implementasi Sistem Dashboard Kesehatan Berbasis AI*

Penelitian ini menghasilkan sistem *dashboard* kesehatan yang menggunakan AI untuk membantu orang memahami kondisi kesehatan mereka. Sistem ini dapat menganalisis data kesehatan, memprediksi perubahan berat badan, dan memberikan penjelasan yang mudah dipahami tentang hasil prediksi tersebut. Untuk membuat sistem ini, penulis menggunakan bahasa pemrograman *Python* dan beberapa pustaka penting seperti *Pandas* untuk mengolah data, *Scikit-learn* untuk membuat model prediksi, dan *Matplotlib* serta *Plotly* untuk menampilkan grafik kesehatan. Sistem ini dirancang untuk membantu pengguna membuat keputusan yang lebih baik tentang kesehatan mereka dengan menggabungkan analisis prediktif dan penjelasan [32].

Hasil dari sistem ini menunjukkan bahwa *dashboard* dapat menampilkan beberapa komponen penting, seperti grafik kesehatan, tren perubahan berat badan, hasil prediksi, dan penjelasan, Pengguna dapat memperoleh pemahaman yang lebih lengkap tentang kondisi kesehatan mereka. Ini sejalan dengan penelitian yang menyatakan bahwa sistem kesehatan digital yang dapat memberikan penjelasan yang mudah dipahami dapat meningkatkan keterlibatan pengguna dalam memantau kesehatan mereka [33].

4.2. *Hasil Evaluasi Model Prediksi Machine Learning*

Untuk mengetahui seberapa baik sistem dapat memprediksi perubahan berat badan, dilakukan evaluasi model regresi. Evaluasi ini menggunakan beberapa variabel kesehatan pengguna, seperti BMI, asupan kalori harian, dan aktivitas harian. Data yang digunakan merupakan data sintetis yang telah dikalibrasi menggunakan pendekatan stokastik terkontrol. Hal ini dilakukan untuk mensimulasikan dinamika variabel harian (seperti asupan nutrisi) dengan tetap mempertahankan korelasi logis terhadap perubahan berat badan, sehingga performa model regresi yang dihasilkan mencerminkan kemampuan sistem dalam menangkap pola kesehatan yang realistis. Hasil dari evaluasi model ini dapat dilihat pada Tabel 1.

Tabel 1. Hasil Evaluasi Model Regresi

No	Model	R2	MAE	MSE
1	Linear Regression	0.98	1.75	4.48
2	Random Forest Regressor	0.95	2.63	10.25
3	Support Vector Regressor	0.20	11.37	204.28
4	Decision Tree Regressor	0.89	3.73	25.66
5	Extreme Gradient Regressor	0.97	1.97	6.26

Berdasarkan hasil evaluasi pada Tabel 1, model *Linear Regression* menunjukkan nilai error yang sangat rendah serta koefisien determinasi yang tinggi. Nilai R^2 sebesar 0.98 menunjukkan bahwa sebagian besar variasi perubahan berat badan dapat dijelaskan oleh variabel yang digunakan dalam model. Hasil ini menunjukkan bahwa model mampu menangkap hubungan antara variabel kesehatan dengan cukup baik.

Meskipun nilai R^2 yang didapat sangat tinggi, kita harus berhati-hati dalam mengartikannya. Tingginya nilai R^2 dalam penelitian ini mungkin karena penggunaan data sintetis yang memiliki pola hubungan yang lebih terkontrol daripada data sebenarnya [23]. Oleh karena itu, hasil ini tidak dimaksudkan untuk menunjukkan keunggulan model secara absolut, melainkan sebagai validasi awal terhadap mekanisme prediksi dalam sistem yang dikembangkan. Selain itu, hasil perbandingan dengan model lain seperti *Random Forest* dan *XGBoost* menunjukkan bahwa model yang lebih kompleks tidak memberikan peningkatan performa yang signifikan. Hal ini mengindikasikan bahwa hubungan antarvariabel dalam dataset cenderung linear, sehingga model *Linear Regression* menjadi lebih sesuai untuk digunakan dalam penelitian ini.

Dalam sistem pendukung keputusan, menggunakan model yang sederhana, stabil, dan mudah diinterpretasikan memiliki keuntungan. Apalagi, hasil prediksi akan diintegrasikan dengan LLM untuk memberikan penjelasan yang jelas dan mudah dipahami pengguna.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

4.3. *Hasil Evaluasi Model LLM*

Untuk memperkuat evaluasi secara kuantitatif, penelitian ini menggunakan pendekatan LLM-*as-a-Judge* untuk menilai kualitas informasi kesehatan yang dihasilkan oleh lima model LLM yang diuji. Tujuannya adalah untuk melakukan penilaian sistematis terhadap hasil model berdasarkan kriteria yang sudah ditetapkan, sehingga memungkinkan perbandingan kinerja antar model secara lebih objektif dan terukur [31]. Pendekatan ini dipilih karena tidak ada standar tunggal yang mendefinisikan informasi kesehatan yang "benar". Oleh karena itu, metrik seperti BLEU atau ROUGE yang bergantung pada referensi tidak bisa digunakan [34].

Framework evaluasi dirancang mengikuti pendekatan *G-Eval* yang diusulkan oleh Liu et al. [34], dengan menggunakan *chain-of-thought* (CoT) dan paradigma *form-filling* untuk menilai kualitas *output* secara multidimensi. Empat dimensi evaluasi ditetapkan, yaitu: (1) *Relevance* — kesesuaian *insight* dengan data pengguna; (2) *Accuracy* — ketepatan klaim medis dan gizi; (3) *Clarity* — kemudahan pemahaman bagi pengguna awam; dan (4) *Actionability* — ketersediaan langkah konkret yang dapat langsung diimplementasikan. Setiap dimensi dinilai pada skala *Likert* 1 hingga 5 dengan *behavioral anchor* eksplisit untuk memitigasi *leniency bias* yang umum ditemukan pada sistem LLM-*as-a-judge* [35]. Model *judge* yang digunakan adalah Moonshotasi Kimi-K2, dipilih dari keluarga model yang berbeda dari kelima model generator untuk memitigasi *preference leakage* dan *self-enhancement bias* [36]. Evaluasi dilakukan terhadap 5 sampel yang diambil secara acak dari dataset sintetis, menghasilkan total 25 penilaian independen (5 sampel × 5 model).

Hasil evaluasi LLM-*as-a-Judge* disajikan pada Tabel 2. Lima model LLM dievaluasi sebagai generator *insight* kesehatan, mencakup model dari tiga keluarga berbeda: OpenAI OSS (gpt-oss-120b, gpt-oss-20b), Meta Llama (llama-4-scout-17b, llama-3.1-8b), dan Alibaba Qwen (qwen3-32b).

Tabel 2. Hasil Evaluasi LLM-as-a-Judge (Skor Rata-rata, Normalisasi 0–100)

No	Model	Relevance	Accuracy	Clarity	Actionability	Composite
1	llama-4-scout-17b	90.0	55.0	95.0	35.0	68.7
2	gpt-oss-120b	90.0	55.0	85.0	20.0	62.5
3	llama-3.1-8b	65.0	35.0	75.0	35.0	52.5
4	qwen3-32b	31.2	50.0	37.5	25.0	26.0
5	gpt-oss-20b	25.0	15.0	30.0	10.0	20.0

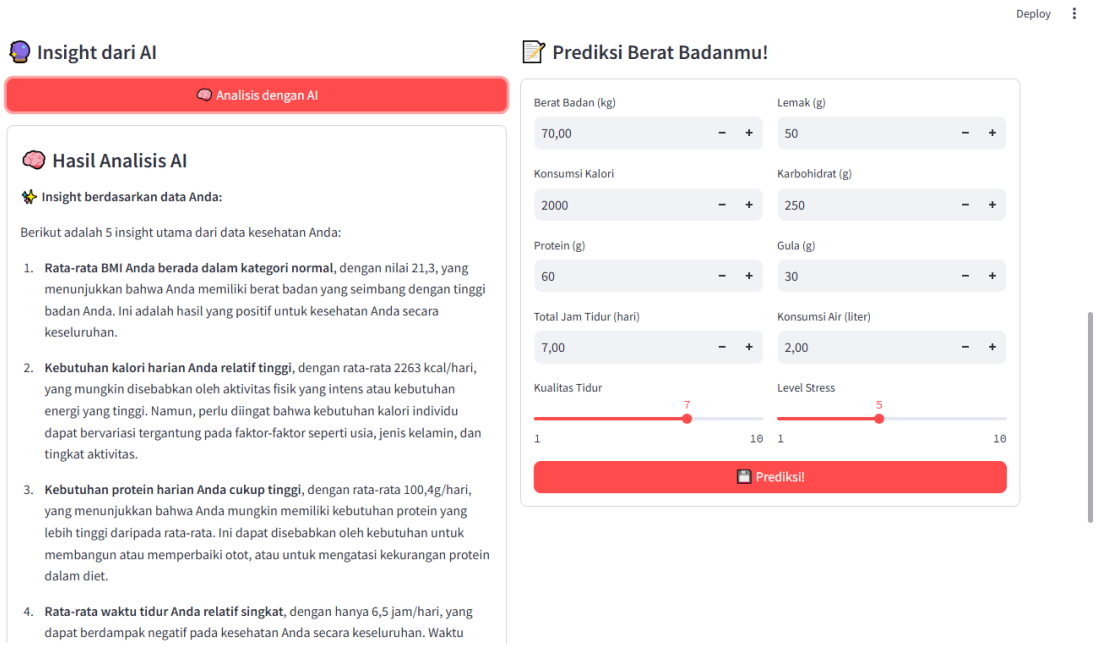
Berdasarkan Tabel 2, model llama-4-scout-17b mendapatkan skor tertinggi sebesar 68.7, diikuti oleh gpt-oss-120b dengan skor 62.5, lalu llama-3.1-8b dengan skor 52.5, qwen3-32b dengan skor 26.0, dan gpt-oss-20b dengan skor 20.0. Hal ini menunjukkan bahwa model yang menggunakan arsitektur *Mixture-of-Experts* (MoE) dengan parameter yang lebih efisien tidak selalu kalah dari model dengan jumlah parameter yang lebih besar. Ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa arsitektur dan kualitas pelatihan lebih memengaruhi performa daripada jumlah parameter [37].

Analisis lebih lanjut menunjukkan pola yang sama pada semua model: *actionability* merupakan dimensi dengan skor terendah di semua model (rata-rata 25.0), jauh di bawah *clarity* dan *relevance*. Temuan ini mengindikasikan bahwa model-model LLM yang dievaluasi secara konsisten kurang mampu memberikan panduan tindakan yang spesifik, realistis, dan terukur dalam konteks *health monitoring*. Model yang digunakan dalam sistem ini (llama-3.1-8b) memperoleh *composite score* sebesar 52.5, di bawah llama-4-scout-17b (68.7) dan gpt-oss-120b (62.5). Pemilihan model Llama-3.1-8b dalam implementasi sistem didasarkan pada pertimbangan efisiensi komputasi, waktu respon, dan kemudahan penerapan. Perbedaan skor sebesar 16.2 poin dianggap sebagai *trade-off* yang dapat diterima antara kualitas output dan biaya operasional, sesuai dengan prinsip pemilihan model dalam sistem AI yang memiliki sumber terbatas [38].

4.4. *Integrasi Large Language Model untuk Interpretasi Prediksi*

Salah satu hal terpenting dari penelitian ini adalah cara penulis menggabungkan model LLM sebagai lapisan untuk menjelaskan hasil prediksi. Angka-angka yang dihasilkan oleh model regresi dikirim ke LLM dalam bentuk *structured prompt*, dan kemudian LLM membuat penjelasan yang jelas tentang hubungan antara kesehatan pengguna dan hasil prediksi. Hasil dari implementasi LLM dapat dilihat pada Gambar 2.

p-ISSN : 2803-1507 e-ISSN : 2803-1531



Gambar 2. Implementasi LLM

Cara ini memungkinkan sistem menjelaskan hasil analisis dengan lebih baik dibandingkan cara visualisasi yang biasa digunakan. Penelitian sebelumnya menujukkan bahwa penjelasan yang menggunakan bahasa sehari-hari dapat membuat pengguna lebih percaya terhadap sistenAI dan membantu mereka memahami hasil analisis dengan lebih baik [21]. Selain itu, integrasi LLM juga membantu menjelaskan bagaimana sistem AI bekerja, sehingga pengguna dapat lebih mudah memahami hasil analisis. Hal ini sejalan dengan penelitian terbaru yang menggabungkan model ML dan LLM untuk membuat sistem pendukung keputusan berbasis AI lebih transparan [39].

5. DISKUSI

5.1. *Analisis Perbandingan dengan Penelitian Sebelumnya*

Jika dibandingkan dengan penelitian sebelumnya, sebagian besar studi tentang *dashboard* kesehatan hanya fokus pada dua hal: menampilkan data dengan cara yang menarik dan membuat model prediksi yang lebih akurat. Misalnya, penelitian oleh Gomez dan timnya menunjukkan bahwa dengan menggunakan model ML dalam sistem kesehatan, kita bisa lebih baik dalam mendeteksi pola kesehatan pengguna [14]. Namun, penelitian itu tidak menjelaskan bagaimana hasil analisisnya disampaikan kepada orang yang tidak ahli teknologi. Penelitian lain oleh Allgaier dan timnya juga menyebutkan bahwa interpretasi dalam sistem AI kesehatan sangat penting. Hal ini bisa meningkatkan kepercayaan pengguna terhadap sistem yang membantu pengambilan keputusan [40]. Sayangnya, cara yang mereka gunakan masih terbatas pada cara menampilkan data atau metode untuk mengetahui fitur mana yang penting. Studi ini berbeda dari penelitian tersebut. Penulis mengusulkan penggunaan LLM sebagai cara untuk menjelaskan hasil prediksi model regresi dengan bahasa sederhana. Dengan begitu, kami menggabungkan prediksi dan penjelasan dalam satu sistem *dashboard* kesehatan.

5.2. *Implikasi Penelitian*

Penelitian ini memiliki beberapa hasil yang sangat penting, baik dari sisi teori maupun praktik. Dari sisi teori, penelitian ini membuktikan bahwa menggabungkan model ML dan LLM dapat menjadi cara yang efektif untuk membuat sistem kesehatan berbasis AI lebih mudah dipahami. Dengan menggunakan kemampuan LLM untuk membuat penjelasan, kita dapat memperluas konsep AI yang dapat dijelaskan dan membuatnya lebih bermanfaat. Dari sisi praktik, sistem yang dikembangkan dalam penelitian ini dapat digunakan sebagai alat untuk memantau kesehatan pribadi. Dengan adanya penjelasan yang mudah dipahami, orang dapat lebih mudah memahami kondisi kesehatan mereka dan membuat keputusan yang lebih baik tentang bagaimana menjaga kesehatan mereka.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

5.3. *Keterbatasan Penelitian*

Penelitian ini menunjukkan bahwa integrasi ML dan LLM dalam sistem *dashboard* kesehatan memiliki potensi yang besar. Namun, ada beberapa hal yang perlu diperhatikan karena masih terdapat beberapa keterbatasan. Pertama, penelitian ini menggunakan dataset sintetis yang berfungsi sebagai validasi fungsional dan pembuktian konsep (*proof of concept*). Meskipun distribusi data telah dirancang menyerupai karakteristik klinis, hasil prediksi dan interpretasi dalam studi ini lebih ditekankan untuk menunjukkan kapabilitas integrasi sistem *decision-support*, sehingga diperlukan pengujian lebih lanjut dengan dataset klinis otentik untuk penerapan pada kondisi kesehatan nyata. Kedua, penilaian terhadap LLM masih bersifat kualitatif dan belum melibatkan pendapat langsung dari orang yang menggunakan sistem atau tenaga kerja kesehatan. Oleh karena itu, penelitian selanjutnya sebaiknya menggunakan data kesehatan yang sebenarnya dan melakukan penilaian berbasis pengguna untuk mengetahui seberapa efektif interpretasi yang dihasilkan oleh LLM dalam memantau kesehatan personal.

Dari sudut pandang sistem informasi, penelitian ini termasuk dalam pengembangan sistem keputusan yang berbasis AI. Hal yang paling berdampak dari penelitian ini adalah cara menggabungkan komponen analitik dan interpretasi menjadi satu sistem yang utuh. Dengan menggunakan penjelasan berbasis model LLM, sistem ini tidak hanya menampilkan data dan hasil ramalan, tetapi juga memberikan pemahaman yang lebih dalam tentang konteksnya, sehingga membantu pengguna dalam membuat keputusan. Pendekatan ini membuat tampilan utama sistem menjadi lebih penting sebagai alat untuk membantu pengambilan keputusan, bukan hanya sebagai cara untuk memvisualisasikan data.

6. KESIMPULAN DAN SARAN

6.1. *Kesimpulan*

Penelitian ini bertujuan untuk membuat sistem *dashboard* kesehatan yang cerdas dengan bantuan AI. Sistem ini bisa menggabungkan model prediksi ML dan LLM untuk memberikan penjelasan yang lebih baik tentang hasil analisis kesehatan. Dari hasil uji coba dan evaluasi sistem, kita bisa menarik beberapa kesimpulan penting. Pertama, sistem *dashboard* yang kita kembangkan bisa menggabungkan proses analisis prediksi dengan model regresi dan penjelasan dalam bahasa alami. Jadi, hasil prediksi tidak hanya ditampilkan dalam bentuk grafik atau angka statistik, tetapi juga dijelaskan dengan cara yang lebih mudah dipahami.

Kedua, model regresi yang kita gunakan bisa memodelkan hubungan antara variabel kesehatan seperti BMI, asupan kalori, dan aktivitas fisik dengan perubahan berat badan dengan baik. Hasil evaluasi model menunjukkan kesalahan prediksi yang kecil dan koefisien determinasi yang tinggi. Jadi, model ini bisa digunakan sebagai bagian analitik dalam sistem pemantauan kesehatan berbasis data. Ketiga, integrasi LLM dalam sistem *dashboard* membantu meningkatkan kemampuan sistem AI dan memberikan penjelasan. Dengan kemampuan LLM dalam menghasilkan penjelajahan dalam bahasa alami, sistem bisa menyajikan informasi kesehatan yang lebih kontekstual dan mudah dipahami oleh pengguna awam. Pendekatan ini sejalan dengan konsep XAI yang menekankan pentingnya transparansi dan interpretasi dalam sistem berbasis AI [40].

6.2. *Saran*

Berdasarkan penelitian yang sudah dilakukan, ada beberapa saran yang bisa dipertimbangkan untuk pengembangan penelitian dan implementasi sistem di masa depan. Pertama, penelitian selanjutnya sebaiknya menggunakan data kesehatan yang nyata. Dengan begitu, hasil evaluasi model prediktif bisa lebih akurat dan mewakili kondisi pengguna sebenarnya. Data nyata juga memungkinkan analisis yang lebih lengkap tentang hubungan antara variabel kesehatan dan perubahan kondisi tubuh.

Kedua, evaluasi interpretasi yang dihasilkan oleh model LLM bisa ditingkatkan dengan studi pengguna. Caranya adalah dengan melibatkan pengguna atau tenaga kesehatan sebagai responden. Pendekatan ini memberikan gambaran yang lebih objektif tentang seberapa baik pengguna memahami dan mempercayai penjelasan yang diberikan sistem.

Ketiga, pengembangan sistem di masa depan bisa menambahkan variabel kesehatan lain. Misalnya, kualitas tidur, tingkat stres, atau pola aktivitas harian. Dengan begitu, sistem bisa memberikan analisis kesehatan yang lebih komprehensif. Keempat, penelitian lanjutan bisa mencari metode XAI lain yang bisa dikombinasikan dengan model LLM. Tujuannya adalah meningkatkan transparansi model prediktif dalam sistem kesehatan digital. Dengan pengembangan lebih lanjut pada data, metode analisis dan evaluasi pengguna, sistem

p-ISSN : 2803-1507 e-ISSN : 2803-1531

dashboard kesehatan berbasis AI diharapkan bisa menjadi alat yang lebih efektif. Tujuanannya adalah untuk mendukung pemantauan kesehatan pribadi dan pengambilan keputusan tentang gaya hidup sehat.

DAFTAR PUSTAKA

[1] J. Karo-Karo *et al.*, "Prediksi Penyakit Jantung Menggunakan Algoritma Machine Learning Berdasarkan Indikator Kesehatan," *Jurnal Ilmiah Sistem Informasi*, vol. 5, no. 1, pp. 84–95, Dec. 2026, doi: 10.51903/1tgz3234.

[2] Z. Zahra, M. Melyani, F. Yusuf, M. T. Herawati, S. Royanti, and A. D. Rosihana, "Implementation of Expert System Applications using Forward Chaining to Detect Dental Health," *Journal of Management and Informatics*, vol. 4, no. 3, pp. 1105–1126, Dec. 2025, doi: 10.51903/jmi.v4i3.212.

[3] U. Udonna, I. Umoren, and G. Ansa, "Contextual Framework for Remote Intelligent Monitoring and Detection System for Prediction of Pregnancy Complications," *Journal of Technology Informatics and Engineering*, vol. 4, no. 2, pp. 225–251, Aug. 2025, doi: 10.51903/jtie.v4i2.244.

[4] Z. Zuo, M. Watson, D. Budgen, R. Hall, C. Kennelly, and N. Al Moubayed, "Data Anonymization for Pervasive Health Care: Systematic Literature Mapping Study," *JMIR Med. Inform.*, vol. 9, no. 10, p. e29871, 2021, doi: 10.2196/29871.

[5] M. Giuffrè and D. L. Shung, "Harnessing the Power of Synthetic Data in Healthcare: Innovation, Application, and Privacy," *NPJ Digit. Med.*, vol. 6, no. 1, p. 186, Dec. 2023, doi: 10.1038/s41746-023-00927-3.

[6] B. N. Jacobsen, "Machine Learning and the Politics of Synthetic Data," *Big Data Soc.*, vol. 10, no. 1, pp. 1–12, 2023, doi: 10.1177/20539517221145372.

[7] G. Santangelo, G. Nicora, R. Bellazzi, and A. Dagliati, "How Good is Your Synthetic Data? SynthRO, a Dashboard to Evaluate and Benchmark Synthetic Tabular Data," *BMC Med. Inform. Decis. Mak.*, vol. 25, no. 1, p. 12, Dec. 2025, doi: 10.1186/s12911-024-02731-9.

[8] Y. Liu, U. R. Acharya, and J. H. Tan, "Preserving Privacy in Healthcare: A Systematic Review of Deep Learning Approaches for Synthetic Data Generation," *Comput. Biol. Med.*, vol. 185, p. 108571, 2025, doi: 10.1016/j.cmpb.2024.108571.

[9] M. Hernandez, G. Epelde, A. Alberdi, R. Cilla, and D. Rankin, "Synthetic Data Generation for Tabular Health Records: A Systematic Review," *Neurocomputing*, vol. 493, pp. 28–45, 2022, doi: 10.1016/j.neucom.2022.04.053.

[10] K. Ramudu, V. M. Mohan, D. Jyothirmai, D. V. S. S. V. Prasad, R. Agrawal, and S. Boopathi, "Machine Learning and Artificial Intelligence in Disease Prediction: Applications, Challenges, Limitations, Case Studies, and Future Directions," in *Contemporary Applications of Data Fusion for Advanced Healthcare Informatics*, IGI Global, 2023, pp. 248–277, doi: 10.4018/978-1-6684-8913-0.ch013.

[11] A. Ahmed, R. Xi, M. Hou, S. A. Shah, and S. Hameed, "Harnessing Big Data Analytics for Healthcare: A Comprehensive Review of Frameworks, Implications, Applications, and Impacts," *IEEE Access*, vol. 11, pp. 111861–111889, 2023, doi: 10.1109/access.2023.3323574.

[12] G. O. Ghosheh, J. Li, and T. Zhu, "A Survey of Generative Adversarial Networks for Synthesizing Structured Electronic Health Records," *ACM Comput. Surv.*, vol. 56, no. 6, pp. 1–35, 2024, doi: 10.1145/3636424.

[13] S. Jenko *et al.*, "Artificial Intelligence in Healthcare: How to Develop and Implement Safe, Ethical and Trustworthy AI Systems," *AI (Switzerland)*, vol. 6, no. 6, pp. 1110–1135, 2025, doi: 10.3390/ai6060116.

[14] A. Gomez-Cabrera and P. J. Escamilla-Ambrosio, "Review of Machine-Learning Techniques Applied to Structural Health Monitoring Systems for Building and Bridge Structures," *Appl. Sci.*, vol. 12, no. 21, p. 10754, 2022, doi: 10.3390/app122110754.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[15] F. Iqbal, M. I. Satti, A. Irshad, and M. A. Shah, "Predictive Analytics in Smart Healthcare for Child Mortality Prediction using a Machine Learning Approach," *Open Life Sci.*, vol. 18, no. 1, p. 20220609, 2023, doi: 10.1515/biol-2022-0609.

[16] F. Boge and A. Mosig, "Causality and Scientific Explanation of Artificial Intelligence Systems in Biomedicine," *Pflugers Arch.*, vol. 477, no. 4, pp. 589–602, 2025, doi: 10.1007/s00424-024-03033-9.

[17] A. Holzinger, "Explainable AI and Multi-Modal Causability in Medicine," *i-com*, vol. 19, no. 3, pp. 171–179, 2021, doi: 10.1515/icom-2020-0024.

[18] F. Bodria, F. Giannotti, R. Guidotti, F. Naretto, D. Pedreschi, and S. Rinzivillo, "Benchmarking and Survey of Explanation Methods for Black Box Models," *Data Min. Knowl. Discov.*, vol. 37, no. 5, pp. 1719–1778, 2023, doi: 10.1007/s10618-023-00933-9.

[19] N. Bienefeld *et al.*, "Solving the Explainable AI Conundrum by Bridging Clinicians' Needs and Developers' Goals," *NPJ Digit. Med.*, vol. 6, no. 1, p. 94, 2023, doi: 10.1038/s41746-023-00837-4.

[20] A. A. Noor, A. Manzoor, M. D. Mazhar Qureshi, M. A. Qureshi, and W. Rashwan, "Unveiling Explainable AI in Healthcare: Current Trends, Challenges, and Future Directions," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 15, no. 2, p. e70018, 2025, doi: 10.1002/widm.70018.

[21] K. Singhal *et al.*, "Large Language Models Encode Clinical Knowledge," *Nature*, vol. 620, no. 7972, pp. 172–180, 2023, doi: 10.1038/s41586-023-06291-2.

[22] J. Li, Y. Bao, Y. Yang, and C. Mao, "ChatGPT Promotes Healthcare: Current Applications and Potential Challenges - Correspondence," *Int. J. Surg.*, vol. 110, no. 5, pp. 3138–3139, 2024, doi: 10.1097/js9.0000000000001354.

[23] A. J. Rodriguez-Almeida *et al.*, "Synthetic Patient Data Generation and Evaluation in Disease Prediction Using Small and Imbalanced Datasets," *IEEE J. Biomed. Health Inform.*, vol. 27, no. 6, pp. 2963–2974, 2023, doi: 10.1109/jbhi.2022.3196697.

[24] O. Ali, W. Abdelbaki, A. Shrestha, E. Elbasi, M. A. A. Alryalat, and Y. K. Dwivedi, "A Systematic Literature Review of Artificial Intelligence in the Healthcare Sector: Benefits, Challenges, Methodologies, and Functionalities," *J. Innov. Knowl.*, vol. 8, no. 1, p. 100333, 2023, doi: 10.1016/j.jik.2023.100333.

[25] M. Abdullah and S. Said, "Performance Evaluation of Machine Learning Regression Models for Rainfall Prediction," *Iran. J. Sci. Technol. Trans. Civ. Eng.*, vol. 49, no. 4, pp. 2105–2120, 2025, doi: 10.1007/s40996-024-01691-4.

[26] A. K. Kleine, E. Kokje, P. Hummelsberger, E. Lermer, I. Schaffernak, and S. Gaube, "AI-enabled Clinical Decision Support Tools for Mental Healthcare: A Product Review," *Artif. Intell. Med.*, vol. 159, p. 103052, 2025, doi: 10.1016/j.artmed.2024.103052.

[27] M. S. Thaiya, K. Julia, and S. Mbugua, "On Software Modular Architecture: Concepts, Metrics and Trends," *Int. J. Comput. Organ. Trends*, vol. 12, no. 1, pp. 14–22, 2022, doi: 10.14445/22492593/ijcot-v12i1p302.

[28] J. Fiser and G. Lengyel, "Statistical Learning in Vision," *Annu. Rev. Vis. Sci.*, vol. 8, pp. 25–49, 2022, doi: 10.1146/annurev-vision-100720-103343.

[29] A. Géron, *Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow*, 3rd ed. Sebastopol, CA: O'Reilly Media, Inc., 2022. [Online]. Available: https://www.oreilly.com/library/view/hands-on-machine-learning/9781098125967

[30] T. Hulsen, "Explainable Artificial Intelligence (XAI): Concepts and Challenges in Healthcare," *AI (Switzerland)*, vol. 4, no. 3, pp. 601–615, 2023, doi: 10.3390/ai4030034.

[31] D. Li *et al.*, "From Generation to Judgment: Opportunities and Challenges of LLM-as-a-judge," in *Proc. 2025 Conf. Empirical Methods in Natural Language Processing*, 2025, pp. 2110–2135, doi: 10.18653/v1/2025.emnlp-main.138.

p-ISSN : 2803-1507 e-ISSN : 2803-1531

[32] B. Karthikeyan, K. Nithya, A. Alkhayyat, and Y. K. Yousif, "Artificial Intelligence Enabled Decision Support System on E-Healthcare Environment," *Intell. Autom. Soft Comput.*, vol. 36, no. 2, pp. 1953–1968, 2023, doi: 10.32604/iasc.2023.032585.

[33] H. Ponnambalath Mohanadas, A. Manikandan, A. F. Ismail, N. Tucker, and S. K. Jaganathan, "Artificial Intelligence Applications in Medical Devices for Personalized Health Care Solutions: Systematic Review," *J. Med. Internet Res.*, vol. 28, p. e72410, Mar. 2026, doi: 10.2196/72410.

[34] Y. Liu, D. Iter, Y. Xu, S. Wang, R. Xu, and C. Zhu, "G-EVAL: NLG Evaluation using GPT-4 with Better Human Alignment," in *Proc. EMNLP 2023*, 2023, pp. 2510–2522, doi: 10.18653/v1/2023.emnlp-main.153.

[35] A. S. Thakur, K. Choudhary, V. S. Ramayapally, S. Vaidyanathan, and D. Hupkes, "Judging the Judges: Evaluating Alignment and Vulnerabilities in LLMs-as-Judges," in *Proc. 4th Workshop on Generation, Evaluation and Metrics (GEM)*, Vienna, Austria: Association for Computational Linguistics, Jul. 2025, pp. 404–430. [Online]. Available: https://aclanthology.org/2025.gem-1.33/

[36] Q. Li, S. Dou, K. Shao, C. Chen, and H. Hu, "Evaluating Scoring Bias in LLM-as-a-judge," *arXiv preprint arXiv:2506.22316*, 2025. [Online]. Available: https://arxiv.org/abs/2506.22316

[37] W. Cai, J. Jiang, F. Wang, J. Tang, S. Kim, and J. Huang, "A Survey on Mixture of Experts in Large Language Models," *IEEE Trans. Knowl. Data Eng.*, vol. 37, no. 7, pp. 1540–1565, 2025, doi: 10.1109/tkde.2025.3554028.

[38] G. Bai *et al.*, "Beyond Efficiency: A Systematic Survey of Resource-Efficient Large Language Models," *arXiv preprint arXiv:2401.00625*, 2024. [Online]. Available: https://arxiv.org/abs/2401.00625

[39] Z. Wang, H. Li, D. Huang, H. S. Kim, C. W. Shin, and A. M. Rahmani, "HealthQ: Unveiling Questioning Capabilities of LLM Chains in Healthcare Conversations," *Smart Health*, vol. 34, p. 100570, 2025, doi: 10.1016/j.smhl.2025.100570.

[40] J. Allgaier, L. Mulansky, R. L. Draelos, and R. Pryss, "How Does the Model Make Predictions? A Systematic Literature Review on the Explainability Power of Machine Learning in Healthcare," *Artif. Intell. Med.*, vol. 143, p. 102616, 2023, doi: 10.1016/j.artmed.2023.102616.