

**PENERAPAN RAG DIGITAL TWIN
FRAMEWORK UNTUK PREDIKSI KADAR
GLUKOSA DARAH PADA PASIEN DIABETES
TIPE 2 BERDASARKAN DATA SIMULATIF**

Proposal Tugas Akhir

Oleh

**Daffari Adiyatma
18222003**



**PROGRAM STUDI SISTEM DAN TEKNOLOGI INFORMASI
SEKOLAH TEKNIK ELEKTRO DAN INFORMATIKA
INSTITUT TEKNOLOGI BANDUNG
Nopember 2025**

LEMBAR PENGESAHAN

PENERAPAN RAG DIGITAL TWIN FRAMEWORK UNTUK PREDIKSI KADAR GLUKOSA DARAH PADA PASIEN DIABETES TIPE 2 BERDASARKAN DATA SIMULATIF

Proposal Tugas Akhir

Oleh

**Daffari Adiyatma
18222003**

Program Studi Sistem dan Teknologi Informasi
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung

Proposal Tugas Akhir ini telah disetujui dan disahkan
di Bandung, pada tanggal 20 Nopember 2025

Pembimbing 1

Pembimbing 2

Prof. Dr. Ir. Suhono Harso Supangkat, Ir. Devi Willieam Anggara S.T., M.Phil.,
M. Eng. Ph.D
NIP. 196212031988111001 No Peg. 124110055

DAFTAR ISI

DAFTAR GAMBAR

DAFTAR TABEL

DAFTAR KODE

BAB I

PENDAHULUAN

I.1 Latar Belakang

Penyakit diabetes melitus (DM) merupakan salah satu masalah kesehatan global yang paling serius pada abad ke-21. Menurut laporan International Diabetes Federation (IDF) tahun 2024, jumlah penderita diabetes di Indonesia mencapai lebih dari 19 juta orang dan diperkirakan terus meningkat seiring dengan perubahan gaya hidup dan urbanisasi cepat di kawasan Asia Tenggara (**IDF2024**). Dari seluruh kasus tersebut, sekitar 90–95% tergolong Diabetes Melitus Tipe 2 (T2DM), yang ditandai dengan resistensi insulin dan penurunan sensitivitas sel terhadap glukosa (**WHO2023**). Kondisi ini meningkatkan risiko komplikasi kronis seperti penyakit jantung, gagal ginjal, neuropati, dan kebutaan apabila tidak ditangani dengan manajemen glukosa darah yang baik.

Sebagian besar sistem pengelolaan diabetes di Indonesia masih bersifat reaktif. Data dari **Riskesdas2018; KSI2020<empty citation>** menunjukkan bahwa sebagian besar kasus diabetes di Indonesia tidak terdiagnos, dengan hanya sekitar 26% penderita yang mengetahui status penyakitnya. Hal ini sejalan dengan tinjauan sistematis oleh **Alkaff2021<empty citation>** yang menyimpulkan bahwa sistem kesehatan di Indonesia masih berfokus pada pendekatan kuratif dibandingkan pencegahan. Penggunaan continuous glucose monitoring (CGM) dan insulin pump telah terbukti membantu pasien dalam memantau kadar glukosa secara real-time dan menyesuaikan dosis insulin secara lebih presisi (**Battelino2019**). Selain itu, pendekatan berbasis machine learning mulai digunakan untuk memprediksi fluktuasi glukosa darah berdasarkan data historis pasien **Woldaregay2019<empty citation>**. Teknologi digital twin, yang merupakan replika virtual dari kondisi fisiologis pasien, mulai diterapkan dalam manajemen diabetes untuk mensimulasikan respons metabolismik pasien terhadap berbagai skenario pengobatan (**Bruynseels2018**).

Penelitian terkini oleh **Rad2024<empty citation>** mengusulkan framework digital twin komprehensif berbasis Personal Health Knowledge Graph (PHKG) yang mampu mengintegrasikan data dari Electronic Health Records (EHR), wearable devices, dan mobile health applications dengan standar HL7 FHIR. Framework ini telah terbukti efektif dalam prediksi glukosa dengan Root Mean Square Error (RMSE) 19,83 mg/dL dan mampu memberikan rekomendasi insulin personal serta saran diet yang disesuaikan. Penelitian serupa oleh **Zhang2024<empty citation>** mengintegrasikan machine learning dengan data multiomic untuk memprediksi progresi diabetes tipe 2, menunjukkan potensi digital twin dalam personalized medicine. **Cappon2024<empty citation>** dalam systematic review mereka menemukan bahwa meskipun pendekatan digital twin menjanjikan, sebagian besar implementasinya masih mengandalkan infrastruktur teknologi yang kompleks dan perangkat medis yang mahal.

Meskipun framework-framework tersebut menunjukkan hasil yang menjanjikan, implementasinya menghadapi hambatan signifikan di konteks Indonesia. Pertama, dari sisi infrastruktur digital kesehatan, meskipun pemerintah Indonesia mewajibkan adopsi rekam medis elektronik (EMR) pada akhir 2023, transisi ini masih menghadapi berbagai tantangan teknologi, budaya, dan infrastruktur **Harahap2024<empty citation>**. Studi yang melibatkan 9 provinsi di Indonesia menunjukkan variasi signifikan dalam kesiapan teknologi informasi dan komunikasi (TIK) antarfasilitas kesehatan, dengan perlunya peningkatan sumber daya manusia (SDM), infrastruktur, perangkat keras, dan optimalisasi sistem informasi untuk mencapai kemajuan TIK (**Aisyah2024**). Sebagian besar fasilitas kesehatan belum menyediakan akses terintegrasi ke rekam kesehatan pasien dengan pertukaran informasi yang masih bersifat satu arah, dari fasilitas kesehatan ke pasien (**Harahap2023**).

Kedua, dari sisi keterjangkauan perangkat monitoring, framework **Rad2024<empty citation>** mengasumsikan ketersediaan Continuous Glucose Monitoring (CGM) untuk data real-time. Namun, studi **Ramadaniati2024<empty citation>** menunjukkan bahwa CGM memerlukan biaya setara satu bulan gaji untuk membeli reader dan dua bulan gaji untuk pasokan sensor bulanan, dengan upah minimum harian di Indonesia sekitar US\$3,50. Hal ini menyebabkan CGM tidak terjangkau bagi mayoritas pasien diabetes di Indonesia, terutama mengingat bahwa untuk membeli pasokan pengobatan selama 30 hari (insulin pen, jarum pen, dan monitoring mandiri berdasarkan 5 kali tes per hari), pasien perlu menghabiskan hampir seluruh gaji bulanan mereka.

Ketiga, kompleksitas teknis framework **Rad2024<empty citation>** yang memer-

lukan pengembangan ontology berbasis HL7 FHIR, implementasi GLAV (Global-Local as View) framework untuk integrasi data, dan penggunaan Conditional Random Fields untuk mapping data, membutuhkan expertise spesialis yang belum banyak tersedia di Indonesia. Penelitian menunjukkan bahwa di negara berkembang, adopsi EMR berbeda karena beberapa faktor termasuk infrastruktur sistem kesehatan, tingkat pendidikan dan pelatihan tenaga kesehatan, pendanaan, dan penerimaan budaya terhadap EMR, sehingga di banyak negara berkembang, penggunaan EMR belum sepenuhnya diterapkan (**Abodunrin2020**).

Dari uraian tersebut, tampak bahwa pengelolaan diabetes di Indonesia masih menghadapi hambatan dalam pemanfaatan teknologi prediktif yang efisien dan terjangkau. Kondisi ini menunjukkan adanya kesenjangan antara potensi teknologi digital twin yang telah terbukti efektif di negara maju dengan kemampuan implementasinya di Indonesia. Diperlukan studi lebih lanjut untuk memahami bagaimana konsep digital twin dapat disesuaikan dengan keterbatasan infrastruktur, biaya, serta sumber daya lokal sehingga mampu memberikan manfaat nyata dalam konteks sistem kesehatan nasional.

I.2 Rumusan Masalah

Berdasarkan latar belakang tersebut, permasalahan utama dalam tugas akhir ini adalah kesenjangan antara framework digital twin yang canggih (seperti yang dikembangkan oleh **Rad2024<empty citation>**) dengan kemampuan implementasi di Indonesia yang terkendala oleh keterbatasan infrastruktur EHR, tingginya biaya perangkat monitoring real-time, dan kompleksitas teknis yang memerlukan expertise spesialis. Kesenjangan ini penting untuk diatasi karena mayoritas penderita diabetes di Indonesia (74%) belum terdiagnosis dan memerlukan sistem prediktif yang terjangkau untuk deteksi dini dan manajemen yang lebih baik.

Berdasarkan permasalahan tersebut, penelitian ini difokuskan untuk menelusuri bagaimana konsep digital twin dapat disederhanakan dan diterapkan secara kontekstual di Indonesia, khususnya dalam mendukung prediksi kadar glukosa darah pasien diabetes tipe 2 berbasis data simulatif.

Secara khusus, rumusan masalah dalam tugas akhir ini adalah sebagai berikut:

1. Bagaimana menganalisis dan memformulasikan kebutuhan arsitektur sistem digital twin yang disederhanakan untuk prediksi kadar glukosa darah tanpa ketergantungan pada Personal Health Knowledge Graph dan infrastruktur HL7 FHIR?

2. Bagaimana mengembangkan model prediktif kadar glukosa darah berbasis data simulatif dari dataset publik menggunakan pendekatan pembelajaran mesin?
3. Bagaimana metode validasi yang tepat untuk memastikan hasil prediksi sistem digital twin yang disederhanakan memiliki akurasi yang sebanding dengan pendekatan state-of-the-art?
4. Bagaimana sistem digital twin yang disederhanakan ini dapat digunakan sebagai alat bantu prediksi yang feasible untuk implementasi di fasilitas kesehatan dengan keterbatasan infrastruktur?

I.3 Tujuan

Tujuan umum penelitian ini adalah mengeksplorasi dan mengembangkan pendekatan digital twin berbasis data simulatif untuk prediksi kadar glukosa darah pada pasien diabetes, yang dapat diimplementasikan di Indonesia tanpa memerlukan infrastruktur kompleks dan perangkat monitoring real-time yang mahal.

Secara khusus, tujuan penelitian ini meliputi:

1. Merumuskan kebutuhan dan desain konseptual sistem digital twin yang disederhanakan dengan pendekatan pembelajaran mesin langsung, tanpa memerlukan pengembangan knowledge graph dan integrasi HL7 FHIR.
2. Mengembangkan modul prediktif kadar glukosa berbasis pembelajaran mesin (LSTM atau Random Forest) yang dapat beroperasi menggunakan data minimal dari dataset simulatif publik.
3. Melakukan validasi terhadap hasil prediksi sistem menggunakan dataset terbuka (OhioT1DM Dataset atau UVA/Padova Simulator) dengan metrik evaluasi standar seperti RMSE, MAE, dan Clarke Error Grid Analysis.
4. Mengevaluasi akurasi dan reliabilitas sistem dalam memprediksi perubahan kadar glukosa darah dan membandingkannya dengan baseline pendekatan yang ada.

Kriteria keberhasilan tugas akhir ini adalah terciptanya prototipe sistem yang mampu menghasilkan prediksi kadar glukosa dengan akurasi yang sebanding dengan state-of-the-art (target $\text{RMSE} \leq 25 \text{ mg/dL}$) namun dengan kompleksitas implementasi yang lebih rendah, sehingga dapat menjadi *proof-of-concept* untuk implementasi di fasilitas kesehatan dengan infrastruktur terbatas.

I.4 Batasan Masalah

Penelitian ini dibatasi oleh hal-hal berikut:

1. Data yang digunakan berasal dari dataset terbuka (OhioT1DM Dataset atau UVA/Padova Simulator) tanpa pengumpulan data pasien nyata.
2. Sistem tidak melibatkan perangkat keras, sensor IoT, atau integrasi dengan sistem EHR yang ada.
3. Fokus penelitian adalah pada pengembangan dan pengujian sistem digital twin berbasis perangkat lunak dengan pendekatan pembelajaran mesin langsung, tanpa implementasi Personal Health Knowledge Graph.
4. Model hanya mencakup prediksi kadar glukosa berdasarkan variabel pendukung yang tersedia dalam dataset (asupan karbohidrat, dosis insulin, aktivitas fisik), tanpa melibatkan faktor genetik, psikologis, atau data multiomic.
5. Evaluasi dilakukan terhadap performa sistem dalam skenario simulatif sebagai *proof-of-concept*, bukan pada uji klinis langsung dengan pasien nyata.
6. Sistem yang dikembangkan berfokus pada satu *use case* utama yaitu prediksi glukosa darah, tidak mencakup optimasi insulin atau rekomendasi *meal planning*.

I.5 Metodologi

Tahapan metodologi dirancang untuk memastikan proses penelitian berjalan secara sistematis, mulai dari pengumpulan informasi awal hingga validasi hasil prediksi. Metodologi penelitian ini terdiri dari lima tahap utama:

I.5.1 Studi Literatur

Melakukan kajian pustaka terhadap konsep digital twin, pengelolaan penyakit diabetes, serta penelitian terdahulu terkait model simulatif dan prediktif. Sumber literatur berasal dari jurnal ilmiah bereputasi seperti IEEE Xplore, ScienceDirect, Nature Digital Medicine, dan Journal of Personalized Medicine. Pencarian literatur dilakukan dengan kata kunci “digital twin diabetes”, “glucose prediction machine learning”, “diabetes simulation model”, “simplified digital twin framework”, dan kombinasi kata kunci terkait.

Literatur yang dikumpulkan kemudian dikelompokkan berdasarkan tema: (a) konsep dan arsitektur digital twin dalam kesehatan, dengan fokus pada framework state-of-the-art seperti **Rad2024<empty citation>** dan **Zhang2024<empty citation>**; (b) metode prediksi glukosa darah menggunakan machine learning dan deep learning;

(c) dataset simulatif diabetes yang tersedia secara publik; (d) metrik evaluasi sistem prediktif kesehatan; dan (e) tantangan implementasi teknologi kesehatan digital di negara berkembang.

I.5.2 Analisis Kebutuhan dan Perancangan Sistem

Menentukan kebutuhan fungsional sistem digital twin yang disederhanakan, dengan fokus pada:

- **Pengelolaan data simulatif:** Kemampuan untuk membaca, memproses, dan menyimpan data dari dataset publik dalam format yang konsisten.
- **Modul preprocessing:** Pembersihan data, normalisasi, dan feature engineering untuk persiapan training model.
- **Modul prediksi:** Implementasi model pembelajaran mesin untuk prediksi glukosa darah.
- **Modul evaluasi:** Perhitungan metrik akurasi dan visualisasi hasil prediksi.

Mendesain arsitektur sistem yang terdiri dari tiga komponen utama yang disederhanakan:

1. **Data Management Module:** Modul untuk loading dan preprocessing data dari dataset publik (OhioT1DM atau UVA/Padova), tanpa perlu integrasi dengan sistem EHR atau sensor real-time.
2. **Simplified Patient Digital Model:** Representasi pasien berbasis feature vector yang berisi variabel-variabel penting (glucose history, carbohydrate intake, insulin dosage, physical activity) tanpa menggunakan knowledge graph.
3. **Prediction Engine:** Model pembelajaran mesin (LSTM atau Random Forest) yang dilatih untuk memprediksi kadar glukosa darah berdasarkan data historis, dengan fokus pada efisiensi komputasi dan kemudahan deployment.

I.5.3 Implementasi Sistem

Mengembangkan sistem berbasis Python dengan menggunakan framework dan library berikut:

- **Data processing:** Pandas, NumPy untuk manipulasi data.
- **Machine learning:** Scikit-learn untuk model tradisional (Random Forest, SVM).
- **Deep learning:** TensorFlow atau PyTorch untuk model LSTM/GRU.
- **Visualisasi:** Matplotlib, Seaborn untuk visualisasi hasil prediksi.
- **Evaluasi:** Implementasi metrik RMSE, MAE, dan Clarke Error Grid Analysis.

Tahapan pengembangan model meliputi:

1. Eksplorasi dan analisis dataset untuk memahami distribusi dan karakteristik data.
2. Feature engineering untuk mengekstrak fitur-fitur yang relevan.
3. Pembagian data menjadi training, validation, dan testing set.
4. Pelatihan model dengan hyperparameter tuning.
5. Evaluasi performa model pada test set.

I.5.4 Validasi dan Evaluasi

Melakukan pengujian model dengan metrik evaluasi berikut:

- **Root Mean Square Error (RMSE)**: Mengukur rata-rata deviasi prediksi dari nilai aktual.
- **Mean Absolute Error (MAE)**: Mengukur rata-rata absolut error.
- **Clarke Error Grid Analysis (EG)**: Mengukur clinical accuracy dengan meng-kategorikan error berdasarkan risk.

BAB II

STUDI LITERATUR

II.1 Konsep Digital Twin

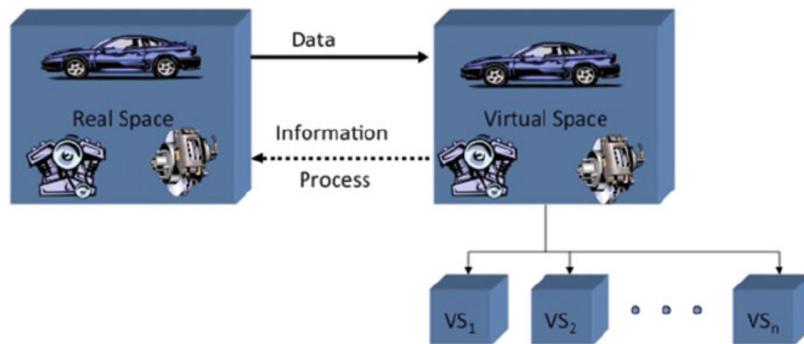
Konsep *Digital Twin* (DT) pertama kali diperkenalkan oleh grieves^{2017digital<empty citation>} sebagai bagian dari paradigma *Product Lifecycle Management* (PLM). DT didefinisikan sebagai representasi digital dari suatu entitas fisik yang secara dinamis diperbarui melalui aliran data dua arah antara dunia nyata (*real space*) dan dunia virtual (*virtual space*). Melalui keterhubungan tersebut, sistem DT dapat melakukan pemantauan, simulasi, serta prediksi terhadap perilaku dan kondisi objek fisik secara berkelanjutan.

Secara umum, arsitektur DT terdiri dari tiga komponen utama:

1. **Physical Entity**, yaitu objek nyata yang menjadi sumber data (contohnya mesin industri, kendaraan, atau pasien dalam konteks medis);
2. **Digital Representation**, yaitu model virtual yang merepresentasikan perilaku dan karakteristik objek fisik;
3. **Data Connection Layer**, yaitu lapisan penghubung yang mengintegrasikan aliran data dua arah antara model digital dan entitas fisik, baik untuk pembauran kondisi maupun umpan balik hasil simulasi.

Contoh arsitektur dasar DT dapat dilihat pada Gambar II.1. Diagram tersebut merupakan model konseptual awal yang dikembangkan oleh Grieves untuk menjelaskan hubungan dua ruang antara sistem fisik dan sistem virtual pada konteks PLM.

Gambar II.1 menampilkan dua ruang utama, yaitu *Real Space* dan *Virtual Space*. Pada bagian *Real Space*, terdapat entitas fisik yang menghasilkan data melalui sensor dan sistem pengukuran. Data tersebut dikirim ke *Virtual Space* untuk memperbaiki model digital secara berkelanjutan. Sebaliknya, hasil simulasi dan analisis dari model virtual dikirim kembali ke sistem fisik sebagai informasi atau proses kontrol.



Gambar II.1 Model konseptual *Digital Twin* yang menunjukkan keterhubungan antara ruang fisik dan ruang virtual. Diadaptasi dari [grieves2017digital<empty citation>](#).

Siklus pertukaran data ini menciptakan hubungan dua arah yang memungkinkan DT berfungsi sebagai sistem adaptif dan prediktif ([grieves2017digital](#)).

Menurut [Bruynseels2018<empty citation>](#), keunggulan utama DT terletak pada kemampuannya untuk menghasilkan *living model*, yakni model yang tidak hanya merepresentasikan kondisi statis suatu sistem, tetapi juga mampu berevolusi mengikuti dinamika data aktual. Hal ini menjadikan DT sangat potensial dalam pengembangan sistem prediktif di berbagai bidang, termasuk manufaktur, pertahanan, energi, dan kesehatan.

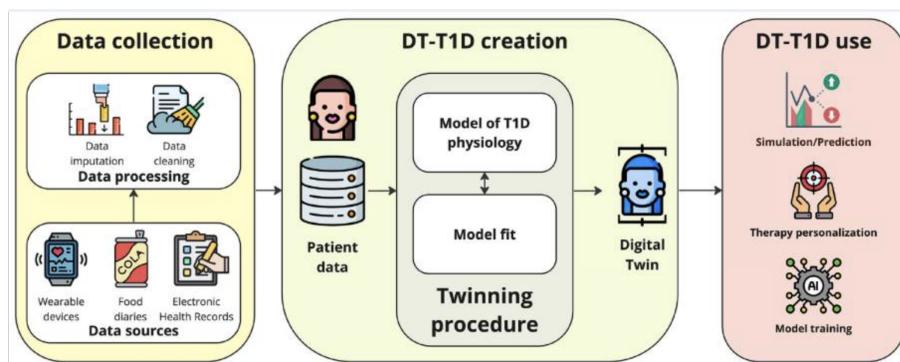
Lebih lanjut, [Cappon2024<empty citation>](#) menjelaskan bahwa arsitektur DT modern telah berevolusi dari model PLM klasik menjadi sistem terintegrasi multi-sumber data. Model ini tidak hanya menerima input dari sensor fisik, tetapi juga dari basis pengetahuan eksternal seperti rekam medis elektronik, data *wearable devices*, dan sistem informasi klinis. Dengan kemampuan integrasi ini, DT berfungsi sebagai platform analitik cerdas yang mampu melakukan prediksi, optimasi, dan personalisasi secara simultan.

Dalam konteks penelitian ini, pemahaman terhadap konsep dan arsitektur dasar DT menjadi landasan untuk mengembangkan model *Digital Twin* yang dapat merepresentasikan kondisi fisiologis pasien diabetes tipe 2 secara virtual. Model digital tersebut akan digunakan untuk proses simulasi dan prediksi kadar glukosa darah berdasarkan data simulatif, sebelum nantinya dikembangkan lebih lanjut melalui pendekatan *Retrieval-Augmented Generation (RAG)* di bab berikutnya.

II.2 Digital Twin dalam Bidang Kesehatan

Konsep *Digital Twin* telah berkembang pesat dalam bidang kesehatan dan kedokteran presisi selama satu dekade terakhir. Menurut **Bruynseels2018<empty citation>**, penerapan DT di bidang medis membuka peluang untuk mengembangkan model digital pasien yang mampu merepresentasikan kondisi fisiologis individu secara dinamis. Model tersebut dapat digunakan untuk melakukan prediksi hasil klinis, simulasi pengobatan, hingga personalisasi terapi berdasarkan karakteristik spesifik pasien.

Salah satu penerapan paling menonjol adalah pada bidang manajemen penyakit kronis seperti diabetes. **Cappon2024<empty citation>** dalam tinjauan sistematisnya menunjukkan bahwa DT digunakan untuk mensimulasikan perilaku metabolismik tubuh dan memprediksi kadar glukosa darah berdasarkan data historis pasien. Sistem ini berfungsi sebagai *virtual replica* dari tubuh pasien yang beroperasi secara parallel dengan sistem nyata untuk memberikan wawasan tentang status kesehatan serta potensi respon terhadap terapi.



Gambar II.2 Arsitektur umum *Digital Twin* dalam bidang kesehatan yang mengintegrasikan data dari perangkat medis, sensor, dan catatan kesehatan elektronik. Diadaptasi dari **Cappon2024<empty citation>**.

Gambar II.2 menampilkan arsitektur umum *Digital Twin* dalam bidang kesehatan sebagaimana dijelaskan oleh **Cappon2024<empty citation>**. Diagram tersebut memperlihatkan bahwa DT medis terdiri atas beberapa lapisan integrasi data, meliputi:

1. **Sumber Data Fisik**, yang mencakup perangkat medis seperti *Continuous Glucose Monitoring* (CGM), sensor aktivitas, dan perangkat *wearable*;
2. **Lapisan Data dan Analitik**, tempat data klinis, fisiologis, dan perilaku pasien diproses serta diintegrasikan ke dalam model digital;
3. **Model Digital Pasien (Patient Twin)**, yaitu representasi virtual pasien yang digunakan untuk prediksi, simulasi, dan evaluasi terapi; serta
4. **Interface Klinik**, yang memungkinkan dokter atau sistem pendukung kepu-

tusan memanfaatkan hasil simulasi untuk tindakan medis.

Melalui arsitektur tersebut, sistem DT memungkinkan analisis prediktif yang lebih akurat karena dapat memperhitungkan variabilitas fisiologis antar individu. **Rad2024**<empty citation> bahkan mengusulkan framework *Patient-Centric Digital Twin* berbasis *Personal Health Knowledge Graph (PHKG)* yang mampu mengintegrasikan berbagai sumber data klinis dengan standar interoperabilitas HL7 FHIR. Framework tersebut terbukti mampu meningkatkan akurasi prediksi kadar glukosa dengan nilai *Root Mean Square Error (RMSE)* sebesar 19,83 mg/dL.

Zhang2024<empty citation> memperluas konsep ini dengan mengintegrasikan data multi-omics ke dalam model DT untuk memprediksi progresi penyakit *Type 2 Diabetes Mellitus* (T2DM). Pendekatan tersebut menunjukkan potensi besar DT dalam mendukung *personalized medicine* dengan memanfaatkan data fisiologis, genetik, dan perilaku secara holistik.

Meskipun hasilnya menjanjikan, sebagian besar penerapan DT di bidang kesehatan masih menghadapi kendala berupa kebutuhan akan infrastruktur teknologi tinggi, biaya perangkat medis, dan ketersediaan data real-time yang memadai (**Cappon2024**; **Rad2024**). Tantangan ini semakin besar di negara berkembang, di mana infrastruktur digital dan rekam medis elektronik belum merata. Oleh karena itu, adaptasi konsep DT dengan pendekatan berbasis data simulatif dan kecerdasan buatan menjadi alternatif yang relevan untuk konteks sistem kesehatan seperti di Indonesia.

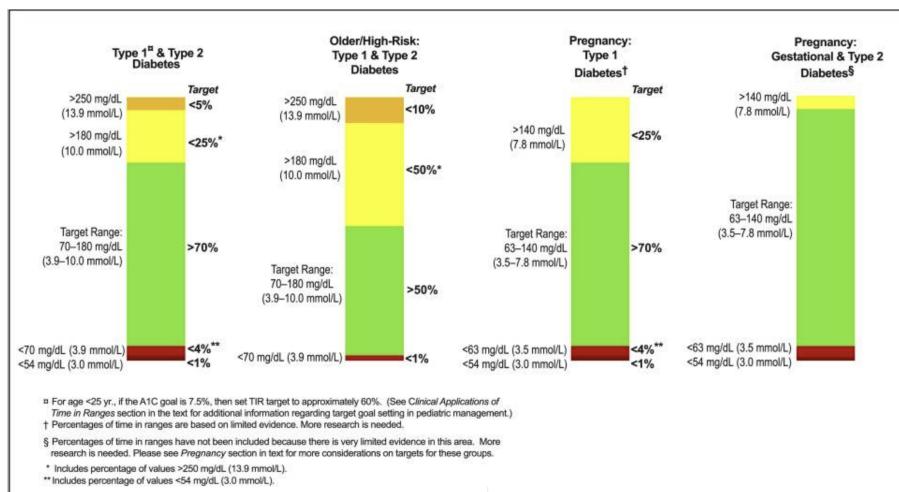
II.3 Manajemen Diabetes dan Pendekatan Digital

Diabetes Mellitus (DM) merupakan penyakit metabolismik kronis yang ditandai oleh meningkatnya kadar glukosa darah akibat gangguan pada sekresi atau kerja insulin. Bentuk yang paling umum adalah *Diabetes Mellitus Tipe 2* (T2DM) yang menyumbang lebih dari 90% kasus diabetes di dunia dan di Indonesia (**WHO2023**). Penyakit ini berhubungan erat dengan resistensi insulin serta penurunan sensitivitas sel terhadap glukosa, sehingga pengelolaan kadar glukosa menjadi aspek kunci dalam pencegahan komplikasi kronis seperti penyakit jantung, gagal ginjal, neuropati, dan kebutaan.

Pendekatan konvensional dalam manajemen T2DM melibatkan pemantauan kadar glukosa darah secara berkala, pengaturan pola makan, aktivitas fisik, serta pemberian terapi insulin atau obat oral. Namun, strategi ini seringkali bersifat reaktif, yaitu penyesuaian dilakukan setelah kadar glukosa menyimpang dari rentang normal. Pendekatan seperti ini belum sepenuhnya mampu mencegah fluktuasi glukosa

ekstrem yang berisiko menimbulkan komplikasi jangka panjang (**Alkaff2021**).

Untuk meningkatkan efektivitas pengelolaan diabetes, berbagai teknologi digital telah dikembangkan, antara lain *Continuous Glucose Monitoring* (CGM) dan *insulin pump* yang memungkinkan pemantauan dan penyesuaian terapi secara real-time (**Battelino2019**). Sistem ini mengubah paradigma pengelolaan diabetes dari berbasis manual menjadi berbasis data (*data-driven management*). Selain itu, penggunaan aplikasi mobile dan perangkat *wearable* juga memungkinkan pasien untuk melacak pola makan, aktivitas fisik, serta kadar glukosa harian secara otomatis.



Gambar II.3 Contoh sistem pemantauan kadar glukosa darah berkelanjutan (*Continuous Glucose Monitoring*) yang digunakan dalam manajemen diabetes.
Diadaptasi dari **Battelino2019**<empty citation>.

Gambar II.3 memperlihatkan contoh sistem *Continuous Glucose Monitoring* (CGM) yang memungkinkan akuisisi data glukosa secara kontinu dari sensor subkutan. Data dari CGM dikirim ke perangkat penerima atau aplikasi seluler untuk analisis dan visualisasi pola glukosa pasien. Informasi ini dapat digunakan untuk menyesuaikan dosis insulin secara otomatis atau semi-otomatis, sehingga membantu menjaga kadar glukosa dalam rentang aman (**Battelino2019**).

Selanjutnya, kemajuan dalam bidang *machine learning* memungkinkan analisis pola data glukosa dalam jangka panjang untuk tujuan prediktif. **Woldaregay2019**<empty citation> mengembangkan model pembelajaran mesin yang mampu memprediksi fluktuasi glukosa darah jangka pendek berdasarkan data historis pasien. Pendekatan ini menunjukkan bahwa model berbasis data dapat membantu pasien mengantisipasi episode hipoglikemia atau hiperglikemia sebelum terjadi.

Meskipun teknologi tersebut memberikan kemajuan signifikan, penerapannya di

Indonesia masih terbatas. Studi oleh **Ramadaniati2024<empty citation>** menunjukkan bahwa harga perangkat CGM di Indonesia relatif tinggi, dengan biaya yang dapat mencapai satu hingga dua bulan gaji rata-rata pasien. Kondisi ini menimbulkan tantangan dalam pemerataan akses terhadap teknologi pemantauan digital, terutama di fasilitas kesehatan primer.

Oleh karena itu, diperlukan pendekatan alternatif yang lebih terjangkau namun tetap memiliki kemampuan prediktif tinggi. Salah satu solusi yang potensial adalah penggunaan model *Digital Twin* berbasis data simulatif, yang dapat menggantikan kebutuhan perangkat real-time melalui pemanfaatan data historis atau sintetis untuk pelatihan model prediksi glukosa. Pendekatan ini membuka peluang bagi pengembangan sistem prediktif adaptif yang relevan untuk konteks sumber daya terbatas seperti di Indonesia.

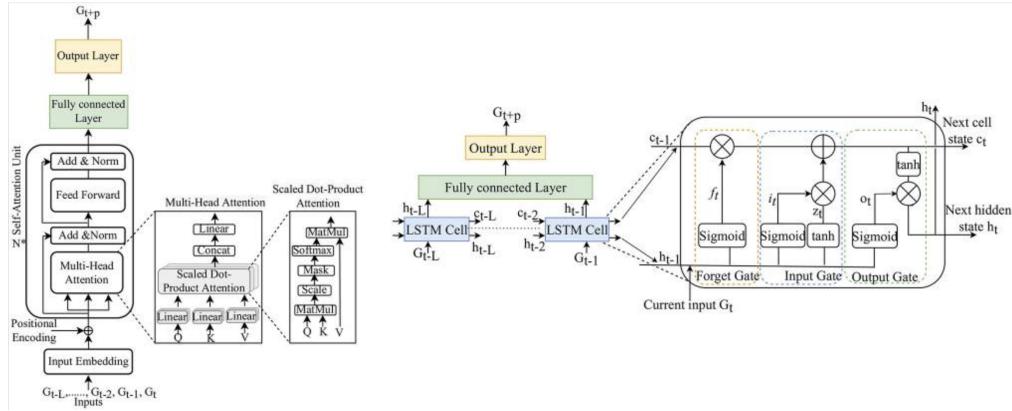
II.4 Metode Machine Learning untuk Prediksi Kadar Glukosa

Prediksi kadar glukosa darah merupakan salah satu tantangan utama dalam manajemen *Diabetes Mellitus*. Kompleksitas hubungan antara berbagai variabel fisiologis seperti asupan karbohidrat, dosis insulin, aktivitas fisik, dan stres membuat pendekatan konvensional berbasis model matematis sulit mencapai akurasi tinggi. Dalam konteks ini, pendekatan *machine learning* (ML) muncul sebagai solusi potensial untuk menangkap pola non-linear yang tidak dapat dimodelkan secara eksplisit.

Menurut **Woldaregay2019<empty citation>**, model pembelajaran mesin dapat mempelajari hubungan kompleks antara variabel fisiologis dari data historis pasien. Pendekatan ini melibatkan pelatihan algoritma menggunakan dataset yang berisi pasangan masukan (misalnya kadar glukosa sebelumnya, dosis insulin, asupan makanan) dan keluaran (kadar glukosa berikutnya). Setelah dilatih, model dapat digunakan untuk memprediksi kadar glukosa masa depan berdasarkan pola yang telah dipelajari.

Beberapa algoritma yang umum digunakan dalam prediksi glukosa antara lain *Random Forest (RF)*, *Support Vector Regression (SVR)*, dan model berbasis jaringan saraf dalam seperti *Long Short-Term Memory (LSTM)* (**li2022time**). Model LSTM populer karena kemampuannya memproses data deret waktu dan mempertahankan informasi jangka panjang melalui mekanisme *memory cell* dan *gating*.

Gambar II.4 menunjukkan arsitektur gabungan *LSTM–Attention* sebagaimana dijelaskan oleh **Ghimire2024<empty citation>**. Model ini terdiri dari tiga bagian uta-



Gambar II.4 Arsitektur model *Long Short-Term Memory (LSTM)* dengan mekanisme *attention* untuk prediksi kadar glukosa darah jangka pendek. Model ini mengombinasikan lapisan *self-attention*, unit *LSTM*, dan lapisan keluaran terhubung penuh untuk memproses deret waktu kadar glukosa. Diadaptasi dari Ghimire2024<empty citation>.

ma. Pertama, lapisan *input embedding* dan *self-attention* bertugas mengekstraksi fitur temporal penting dari data kadar glukosa historis. Kedua, lapisan *LSTM cell* menangkap dependensi jangka panjang dan hubungan non-linear antar nilai glukosa, sementara mekanisme gerbang—*forget gate*, *input gate*, dan *output gate*—mengatur aliran informasi antar langkah waktu. Ketiga, lapisan keluaran (*fully connected layer* dan *output layer*) menghasilkan prediksi kadar glukosa masa depan G_{t+p} berdasarkan masukan deret waktu $G_{t-L}, G_{t-L+1}, \dots, G_t$.

Kombinasi antara *attention mechanism* dan arsitektur LSTM terbukti meningkatkan kemampuan model dalam menangkap hubungan temporal yang kompleks dibandingkan LSTM murni (Ghimire2024). Pendekatan ini memungkinkan model untuk fokus pada segmen data yang paling relevan untuk prediksi, sehingga meningkatkan akurasi dan stabilitas hasil pada berbagai dataset pasien.

Untuk menilai kinerja model prediktif, digunakan berbagai metrik evaluasi kuantitatif. Dua metrik yang paling umum adalah *Root Mean Square Error (RMSE)* dan *Mean Absolute Error (MAE)*. RMSE mengukur akar rata-rata kuadrat selisih antara nilai aktual dan nilai prediksi, sedangkan MAE mengukur rata-rata selisih absolut keduanya. Rumus matematis dari kedua metrik tersebut diberikan pada Persamaan II.1.

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}, \quad \text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (\text{II.1})$$

Nilai RMSE dan MAE yang lebih kecil menunjukkan performa model yang lebih baik. Selain metrik statistik, beberapa penelitian juga menggunakan metrik berbasis klinis seperti *Clarke Error Grid (CEG)* untuk mengevaluasi implikasi klinis dari kesalahan prediksi (**Battelino2019**). Dengan kombinasi evaluasi matematis dan klinis ini, model pembelajaran mesin dapat dinilai tidak hanya berdasarkan akurasi prediksi, tetapi juga keamanan penerapannya dalam konteks medis.

Pendekatan berbasis ML seperti LSTM memiliki potensi besar untuk diintegrasikan dengan konsep *Digital Twin*. Model DT dapat menggunakan hasil prediksi ML untuk memperbarui kondisi virtual pasien secara real-time, sementara data simulatif dari DT dapat digunakan kembali untuk melatih model ML, menciptakan sistem pembelajaran adaptif dua arah antara dunia nyata dan digital.

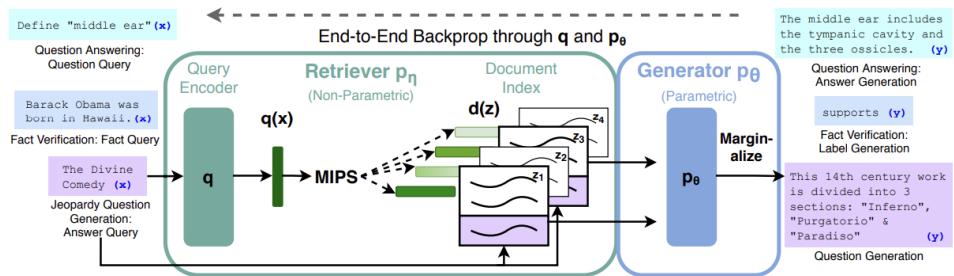
II.5 Retrieval-Augmented Generation (RAG) Framework

Retrieval-Augmented Generation (RAG) merupakan pendekatan dalam bidang *Natural Language Processing (NLP)* yang mengombinasikan dua proses utama, yaitu *retrieval* dan *generation*. Pendekatan ini pertama kali diperkenalkan oleh **Lewis2020**<empty citation> untuk meningkatkan kemampuan model generatif dalam menghasilkan keluaran yang faktual dan berbasis pengetahuan. Tidak seperti model generatif murni yang mengandalkan parameter internal untuk seluruh proses inferensi, RAG memungkinkan model mengakses basis pengetahuan eksternal secara dinamis.

Arsitektur RAG terdiri atas dua komponen utama:

1. **Retriever**, yaitu modul non-parametrik yang bertugas mencari potongan informasi atau dokumen yang relevan dari basis pengetahuan eksternal menggunakan metode pencarian berbasis vektor seperti *Maximum Inner Product Search (MIPS)*;
2. **Generator**, yaitu model bahasa besar berbasis *sequence-to-sequence transformer* yang menghasilkan keluaran akhir berdasarkan masukan dan konteks hasil *retrieval*.

Gambar II.5 memperlihatkan alur kerja RAG sebagaimana dijelaskan oleh **Lewis2020**<empty citation>. Proses dimulai dengan kueri masukan (*query encoder*) yang diubah menjadi representasi vektor. Representasi ini digunakan oleh modul *retriever* untuk mencari dokumen relevan dari indeks pengetahuan eksternal. Dokumen yang ditemukan kemudian diteruskan ke modul *generator*, yang memanfaatkan informasi tersebut untuk membentuk keluaran berbasis konteks. Proses pelatihan dilakukan secara end-to-end dengan pembaruan parameter bersama melalui mekanisme *backpropagation*,



Gambar II.5 Arsitektur *Retrieval-Augmented Generation (RAG)* yang terdiri atas dua komponen utama: *Retriever* untuk mengambil dokumen relevan dari basis pengetahuan eksternal dan *Generator* untuk menghasilkan keluaran berbasis konteks. Model dilatih secara end-to-end sehingga kedua komponen dapat dioptimalkan bersamaan. Diadaptasi dari Lewis2020<empty citation>.

sebagaimana ditunjukkan oleh panah berlabel “End-to-End Backprop through q and p_θ ”.

Pendekatan ini memberikan peningkatan signifikan dibandingkan model generatif murni karena mampu:

- Mengakses pengetahuan terkini tanpa perlu pelatihan ulang model besar;
- Menghasilkan keluaran yang lebih faktual, relevan, dan dapat ditelusuri sumbernya;
- Mengurangi fenomena *hallucination* dalam keluaran model bahasa besar.

Borgeaud2022<empty citation> memperluas konsep ini melalui integrasi retrieval skala besar, memungkinkan model mengakses triliunan token teks eksternal secara efisien untuk meningkatkan penalaran berbasis pengetahuan. Hasil penelitian tersebut menunjukkan bahwa RAG dapat berfungsi sebagai mekanisme penghubung antara model generatif dan sistem pencarian, menjadikannya salah satu pendekatan utama dalam pengembangan kecerdasan buatan berbasis pengetahuan.

Dalam konteks kesehatan digital, framework RAG berpotensi diintegrasikan dengan *Digital Twin*. Model DT yang biasanya bergantung pada data fisiologis pasien dapat diperluas dengan mengakses basis pengetahuan medis eksternal, seperti literatur ilmiah atau panduan klinis. Melalui integrasi ini, *RAG Digital Twin* dapat memanfaatkan data pasien internal bersama dengan informasi medis eksternal untuk menghasilkan rekomendasi yang lebih adaptif, berbasis bukti, dan kontekstual.