



PERANCANGAN SCHEMA DATA WAREHOUSE STUDI KASUS PENYAKIT DIABETES MELITUS

Ardi Riyadi¹, Johan Abisay Tambunan², Andri Wijaya³

Sistem Informasi, Sains and Tecnology, Universitas Katolik Musi Charitas
Palembang ^{1,2,3}

Email: 2314024@student.ukmc.ac.id¹, 2314007@student.ukmc.ac.id²,
andri_wijaya@ukmc.ac.id³

ABSTRACT

Diabetes Mellitus is a chronic disease with an increasing global prevalence, requiring integrated and data-driven health data management. However, healthcare institutions often face challenges in managing patient data that are distributed across multiple formats and fragmented storage systems (data silos), which hinders strategic analysis and effective decision-making. This study aims to design and implement a Data Warehouse to integrate diabetes patient data as a foundation for clinical and managerial decision support. The research methodology applies a multidimensional schema design using the Star Schema approach, consisting of one fact table and three main dimension tables: demographics, lifestyle, and medical conditions. The Extract, Transform, and Load (ETL) process was implemented using SQL Server Integration Services (SSIS) to cleanse and centralize data from operational sources. The dataset used in this study consists of 10,000 anonymized patient records that have undergone data profiling and data cleansing processes. The results indicate that the developed Data Warehouse is capable of integrating data consistently and supporting multidimensional analysis. Data visualization using Tableau Public reveals a correlation between Body Mass Index (BMI) and diabetes status, where patients diagnosed with diabetes exhibit a higher average BMI compared to non-diabetic patients. This implementation improves data access efficiency and facilitates the identification of health risk patterns. Therefore, the proposed Data Warehouse can serve as a foundation for a healthcare analytics system that supports data-driven strategies for diabetes prevention and management.

Keywords : Data Warehouse, Diabetes Mellitus, Star Schema, Business Intelligence, Healthcare Analytics

ABSTRAK

Diabetes Melitus merupakan penyakit kronis dengan prevalensi yang terus meningkat secara global, sehingga menuntut pengelolaan data kesehatan yang terintegrasi dan berbasis analitik. Namun, fasilitas kesehatan masih menghadapi kendala dalam mengelola data pasien yang tersebar dalam berbagai format dan lokasi penyimpanan yang terfragmentasi (data silo), yang menyulitkan proses analisis strategis dan pengambilan keputusan. Penelitian ini bertujuan untuk merancang dan mengimplementasikan Data Warehouse guna mengintegrasikan data pasien diabetes sebagai dasar pendukung keputusan klinis dan manajerial. Metodologi penelitian menggunakan perancangan skema multidimensi dengan pendekatan Star Schema, yang terdiri dari satu tabel fakta dan tiga tabel dimensi utama, yaitu dimensi demografi, gaya hidup, dan kondisi medis. Proses Extract, Transform, and Load (ETL) diimplementasikan menggunakan SQL Server Integration Services (SSIS) untuk membersihkan dan memusatkan data dari sumber operasional. Dataset yang digunakan terdiri dari 10.000 record pasien yang bersifat anonim dan telah melalui proses data profiling serta pembersihan data. Hasil penelitian menunjukkan bahwa Data Warehouse yang dibangun mampu mengintegrasikan data secara konsisten dan mendukung analisis multidimensi. Visualisasi data menggunakan Tableau Public memperlihatkan adanya korelasi antara Body Mass Index (BMI) dan status diabetes, di mana kelompok pasien diabetes memiliki rata-rata BMI yang lebih tinggi dibandingkan kelompok non-diabetes. Implementasi ini terbukti meningkatkan efisiensi akses data serta mempermudah identifikasi pola risiko kesehatan. Dengan demikian, Data Warehouse yang dirancang dapat menjadi fondasi sistem analitik kesehatan yang mendukung perumusan strategi pencegahan dan penanganan diabetes berbasis data.

Kata Kunci : Data Warehouse, Diabetes Melitus, Star Schema, Business Intelligence, Analisis Kesehatan

PENDAHULUAN

Penyakit diabetes mellitus telah menjadi krisis kesehatan global yang sangat serius, di mana jumlah penderitanya terus meningkat secara signifikan setiap tahunnya. Berdasarkan data terbaru, diabetes tidak hanya berdampak pada penurunan kualitas hidup pasien tetapi juga meningkatkan risiko komplikasi berat seperti penyakit jantung, gagal ginjal, hingga kebutaan jika tidak dikelola dengan tepat [5][10]. Di Indonesia, tantangan ini semakin nyata seiring dengan perubahan gaya hidup masyarakat yang meningkatkan prevalensi penyakit kronis secara nasional. Kondisi kesehatan masyarakat yang memprihatinkan ini menuntut adanya pengelolaan data klinis yang lebih terintegrasi.

Pengelolaan data klinis tersebut kini memasuki era baru di mana volume data kesehatan digital meledak dalam jumlah yang sangat besar atau *big data* pada

periode 2024-2025. Meskipun institusi kesehatan telah mulai mendokumentasikan rekam medis secara elektronik, kenyataannya data tersebut seringkali hanya tersimpan dalam repositori statis yang sulit untuk diolah kembali menjadi informasi strategis [6][26]. Fenomena ini menyebabkan tumpukan data rekam medis menjadi aset yang tidak terutilisasi secara maksimal bagi para praktisi kesehatan. Akibatnya, informasi penting mengenai pola risiko kesehatan pasien seringkali terkubur dalam basis data yang tidak terstruktur.

Ketidakteraturan akses terhadap informasi ini berakar dari sifat data kesehatan yang bersifat terfragmentasi dan heterogen, yang seringkali menyebabkan inkonsistensi dalam proses pelaporan dan analisis [2][15]. Masalah utama yang dihadapi adalah sulitnya melakukan *query* lintas variabel, seperti menghubungkan faktor gaya hidup dengan indikator medis secara cepat, karena struktur database operasional yang tidak dirancang untuk tujuan analitik. Inefisiensi dalam pengolahan data ini memperlambat proses pengambilan keputusan klinis yang seharusnya bisa dilakukan secara dini untuk mencegah komplikasi diabetes. Guna mengatasi inefisiensi tersebut, implementasi sistem *Data Warehouse* (DWH) menjadi solusi yang mendesak.

Implementasi *Data Warehouse* dilakukan melalui perancangan skema multidimensi, khususnya *Star Schema*, yang mampu mengintegrasikan data dari berbagai dimensi seperti demografi, gaya hidup, dan kondisi medis ke dalam satu tabel fakta yang terpusat [1][16]. Dengan memanfaatkan alat bantu seperti *SQL Server Integration Services* (SSIS) untuk proses *ETL* (*Extract, Transform, Load*) dan Tableau untuk visualisasi analitik, data mentah yang kompleks dapat diubah menjadi dasbor interaktif yang mudah dipahami [3][24]. Melalui pendekatan ini, pola tersembunyi seperti korelasi antara indeks massa tubuh (*BMI*) dan aktivitas fisik terhadap risiko diabetes dapat diidentifikasi secara akurat. Berdasarkan urgensi tersebut, penelitian ini difokuskan pada perancangan skema *Data Warehouse* dengan studi kasus penyakit diabetes untuk menghasilkan sistem pendukung keputusan yang lebih cerdas dan responsif.

METODE PENELITIAN

Tahapan Penelitian

Penelitian ini mengikuti *Nine-Step Methodology* yang dikemukakan oleh Ralph Kimball sebagai pendekatan sistematis dalam perancangan *Data Warehouse* berbasis dimensional. Metodologi ini dipilih karena kemampuannya dalam menangani data heterogen berskala besar serta kesesuaian untuk kebutuhan analisis kesehatan yang bersifat multidimensi [1][28].

Tahapan penelitian diawali dengan identifikasi proses bisnis utama, yaitu analisis faktor risiko dan status diagnosis diabetes mellitus. Selanjutnya dilakukan penentuan grain data, di mana satu baris data merepresentasikan satu pasien pada

satu periode observasi, sehingga konsistensi data analitik dapat terjaga.

Pada tahap berikutnya dilakukan identifikasi tabel dimensi dan tabel fakta berdasarkan kebutuhan analisis klinis dan manajerial. Tabel dimensi mencakup aspek demografi, gaya hidup, dan kondisi medis pasien, sedangkan tabel fakta merepresentasikan status diagnosis diabetes beserta atribut numerik pendukung. Hasil dari tahapan ini adalah rancangan skema multidimensi dengan pendekatan Star Schema.

Setelah rancangan logis terbentuk, dilakukan perancangan fisik database pada *SQL Server*, diikuti dengan implementasi proses *Extract, Transform, and Load (ETL)* untuk memindahkan data dari sumber operasional ke dalam *Data Warehouse*. Tahap akhir penelitian adalah pengujian hasil melalui visualisasi data menggunakan alat *Business Intelligence* guna memastikan bahwa struktur *Data Warehouse* mampu mendukung analisis data secara efektif [16][25]. Pendekatan bertahap ini memastikan integritas data tetap terjaga sejak tahap perancangan hingga tahap analisis, serta memudahkan pengembangan sistem di masa mendatang.

Sumber dan Profiling Data

Dataset yang digunakan dalam penelitian ini merupakan data sekunder yang bersifat anonim dan tidak mengandung informasi identitas pribadi pasien, sehingga aman digunakan untuk kepentingan penelitian akademik. Data terdiri dari 10.000 record dengan 22 atribut yang mencakup informasi demografis, gaya hidup, serta indikator medis pasien. Dataset ini digunakan sebagai representasi data klinis pasien diabetes untuk keperluan perancangan dan pengujian sistem *Data Warehouse*.

Sebelum dilakukan proses *ETL*, diterapkan tahap data profiling untuk mengevaluasi kualitas data awal. Tahap ini bertujuan untuk mengidentifikasi keberadaan nilai kosong (*missing values*), duplikasi data, serta inkonsistensi format dan rentang nilai pada atribut tertentu. Hasil profiling menunjukkan bahwa beberapa atribut numerik, seperti *Body Mass Index (BMI)* dan kadar glukosa darah, memerlukan proses standarisasi nilai agar konsisten dan layak dianalisis.

Proses pembersihan data dilakukan dengan menghilangkan data duplikat, menyesuaikan format tipe data, serta melakukan penanganan nilai kosong pada atribut tertentu sebelum data dimuat ke dalam *Data Warehouse*. Tahapan ini berperan penting dalam menjamin akurasi dan reliabilitas analisis yang dihasilkan [5][22]. Data yang telah tervalidasi kemudian digunakan sebagai dasar pembentukan tabel dimensi dan tabel fakta pada skema multidimensi, sehingga kualitas data yang baik menjadi fondasi utama dalam perancangan *Data Warehouse* yang akurat dan konsisten.

Perancangan Skema Bintang (*Star Schema*)

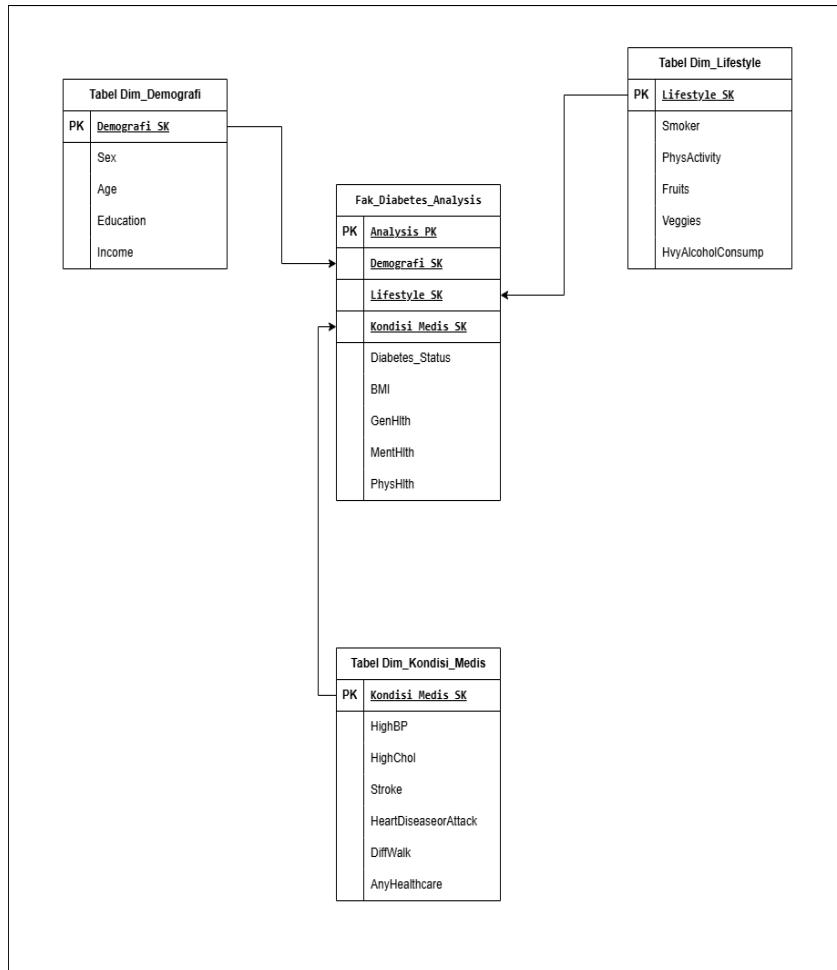
Model data dilakukan menggunakan alat bantu *Draw.io* dengan menerapkan arsitektur *Star Schema*, sebagaimana ditunjukkan pada Gambar 1. Skema ini terdiri dari satu tabel fakta terpusat, yaitu *Fact_Diabetes_Analysis*, yang dikelilingi oleh

tiga tabel dimensi utama, yaitu Dim_Demografi, Dim_Lifestyle, dan Dim_Kondisi_Medis [2][21].

Tabel Fact_Diabetes_Analysis dirancang dengan grain data berupa satu baris merepresentasikan satu pasien pada satu periode observasi, serta menyimpan atribut numerik dan indikator kesehatan seperti *Diabetes_Status*, *BMI*, *General Health*, *Mental Health*, dan *Physical Health*. Tabel ini juga memiliki *foreign key* yang mengacu pada masing-masing tabel dimensi untuk mendukung analisis multidimensi.

Tabel Dim_Demografi menyimpan atribut deskriptif pasien, meliputi jenis kelamin (*Sex*), usia (*Age*), tingkat pendidikan (*Education*), dan tingkat pendapatan (*Income*). Tabel Dim_Lifestyle merepresentasikan faktor gaya hidup pasien seperti status merokok (*Smoker*), tingkat aktivitas fisik (*PhysActivity*), pola konsumsi buah dan sayur, serta konsumsi alkohol. Sementara itu, tabel Dim_Kondisi_Medis menyimpan informasi riwayat kondisi kesehatan pasien, antara lain tekanan darah tinggi (*HighBP*), kolesterol tinggi (*HighChol*), riwayat stroke, penyakit jantung, serta akses layanan kesehatan.

Pemilihan arsitektur *Star Schema* didasarkan pada keunggulannya dalam menyederhanakan struktur data, mempercepat eksekusi query analitik, serta meningkatkan kemudahan pemahaman bagi pengguna akhir seperti dokter dan analis kesehatan [1][27]. Struktur skema ini juga mendukung proses *ETL* yang efisien serta integrasi dengan alat Business Intelligence untuk analisis data kesehatan secara komprehensif.



Gambar 1. Rancangan Star Schema Data Warehouse untuk Analisis Penyakit Diabetes Melitus

Implementasi ETL menggunakan SSIS

Proses *Extract, Transform, and Load* (ETL) pada penelitian ini diimplementasikan menggunakan *SQL Server Integration Services* (SSIS). Alur ETL dirancang secara terstruktur dengan memisahkan proses pemuatan tabel dimensi dan tabel fakta guna menjaga integritas referensial serta mengikuti praktik terbaik dalam perancangan Data Warehouse.

Pada tahap awal, dilakukan proses ETL untuk masing-masing tabel dimensi secara terpisah. Data diekstraksi dari sumber file datar (*flat file* berformat *.csv*), kemudian melalui komponen Data Conversion untuk menyelaraskan tipe data agar sesuai dengan skema database *SQL Server*. Selanjutnya, data dimuat ke dalam tabel Dim_Demografi, Dim_Kondisi_Medis, dan Dim_Lifestyle. Proses ini memastikan bahwa setiap tabel dimensi telah terisi dan memiliki *surrogate key* sebelum digunakan pada tabel fakta.

Setelah seluruh tabel dimensi berhasil dimuat, proses ETL dilanjutkan dengan pengisian tabel fakta Fact_Diabetes_Analysis. Pada tahap ini, data diekstraksi dari database staging menggunakan *OLE DB Source*, kemudian dilakukan Data

Conversion serta transformasi tambahan melalui komponen *Derived Column* untuk membentuk atribut analitik yang dibutuhkan. Selanjutnya, digunakan *Lookup Transformation* secara berurutan ke masing-masing tabel dimensi untuk memperoleh *foreign key* yang merepresentasikan hubungan antar data.

Tahap akhir dari proses *ETL* adalah pemuatan data ke tabel fakta setelah seluruh relasi dimensi berhasil dipetakan. Seluruh proses *ETL* berjalan secara otomatis dan berhasil memuat data tanpa adanya kesalahan atau duplikasi. Dengan alur *ETL* yang terstruktur ini, Data Warehouse yang dihasilkan memiliki data yang konsisten, terintegrasi, dan siap digunakan untuk analisis multidimensi serta visualisasi data pada tahap berikutnya.

Visualisasi Data menggunakan Tableau

Tahap akhir dari metodologi penelitian ini adalah pengembangan dashboard interaktif menggunakan Tableau Public sebagai alat *Business Intelligence*. Data yang telah tersimpan pada *Data Warehouse* di *SQL Server* dihubungkan secara langsung ke Tableau untuk dianalisis berdasarkan struktur tabel fakta dan tabel dimensi yang telah dirancang sebelumnya.

Visualisasi data dikembangkan dengan memanfaatkan tabel fakta *Fact_Diabetes_Analysis* sebagai sumber utama *measure*, serta tabel dimensi sebagai dasar analisis multidimensi. Berbagai jenis visualisasi, digunakan untuk menggambarkan distribusi status diabetes, serta hubungan antara faktor risiko seperti indeks massa tubuh (*BMI*), gaya hidup, dan kondisi medis pasien [3][9].

Dashboard yang dihasilkan dirancang untuk memberikan wawasan secara cepat dan intuitif mengenai pola penyebaran penyakit diabetes dan faktor risiko dominan. Selain berfungsi sebagai sarana pendukung pengambilan keputusan medis berbasis data, visualisasi ini juga digunakan untuk memvalidasi efektivitas struktur *Data Warehouse* dalam mendukung analisis data kesehatan secara multidimensi. Integrasi antara Data Warehouse dan alat *Business Intelligence* ini menghasilkan sistem analitik yang komprehensif dan mudah digunakan oleh pengguna akhir.

HASIL DAN PEMBAHASAN

Implementasi Data Warehouse

Implementasi *Data Warehouse* pada penelitian ini diawali dengan pembentukan struktur fisik basis data menggunakan *SQL Server 2022*. Struktur data dirancang berdasarkan pendekatan *Star Schema* yang telah ditetapkan pada tahap perancangan, sehingga menghasilkan empat tabel utama yang saling terintegrasi, yaitu tiga tabel dimensi (*Dim_Demografi*, *Dim_Lifestyle*, dan *Dim_Kondisi_Medis*) serta satu tabel fakta (*Fact_Diabetes_Analysis*). Tabel dimensi berfungsi untuk menyimpan atribut deskriptif pasien, sedangkan tabel fakta digunakan untuk menyimpan data numerik dan indikator kesehatan yang menjadi dasar analisis.

Seperti ditunjukkan pada Gambar 2, data mentah yang bersumber dari file CSV

berhasil dimuat dan disesuaikan ke dalam struktur Data Warehouse. Setiap tabel dimensi menggunakan surrogate key sebagai primary key untuk memastikan bahwa setiap baris data memiliki identitas unik yang independen dari operational key pada sistem sumber. Pendekatan ini bertujuan untuk menjaga konsistensi historis data serta mendukung fleksibilitas perubahan data di masa mendatang [1].

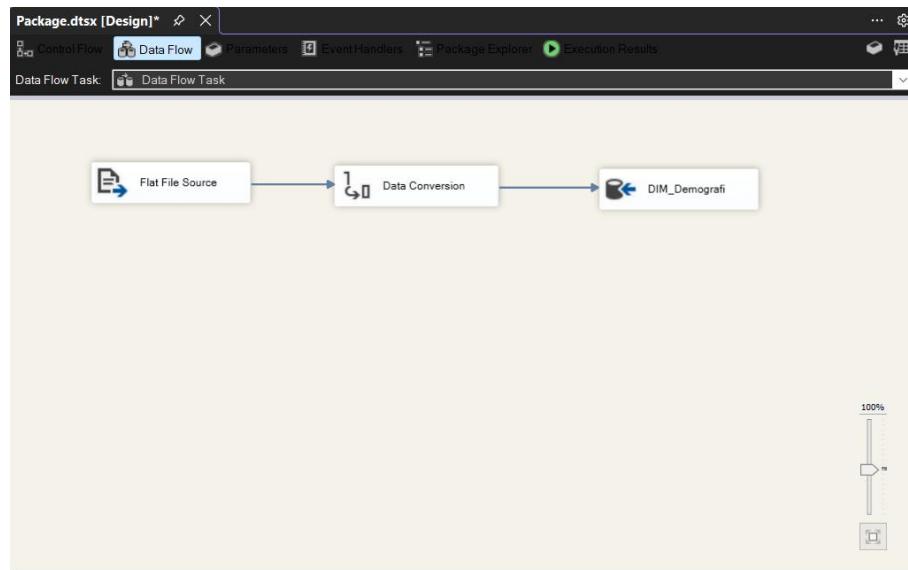
Tabel fakta *Fact_Diabetes_Analysis* dirancang dengan grain satu baris merepresentasikan satu pasien pada satu periode observasi. Tabel ini menyimpan sejumlah atribut numerik utama yang digunakan dalam analisis, antara lain *Diabetes_Status*, *Body Mass Index (BMI)*, *General Health (GenHlth)*, *Mental Health (MentHlth)*, dan *Physical Health (PhysHlth)*, serta *foreign key* yang mengacu pada masing-masing tabel dimensi. Berdasarkan hasil implementasi, nilai *BMI* pasien berada pada rentang 24 hingga 40, sementara status diabetes terbagi ke dalam kategori non-diabetes (0), pre-diabetes (1), dan diabetes (2). Struktur ini memungkinkan analisis multidimensi seperti perhitungan rata-rata *BMI* berdasarkan status diabetes atau kondisi kesehatan pasien.

Struktur fisik *Data Warehouse* yang dibangun terbukti mampu menampung sebanyak 10.000 record data pasien tanpa mengalami kendala performa pada saat dilakukan query agregasi sederhana. Pengujian awal menunjukkan bahwa skema Star Schema yang diterapkan mampu mendukung proses analitik dengan respons yang cepat dan konsisten. Keberhasilan implementasi struktur *Data Warehouse* ini menjadi fondasi utama bagi proses pengisian data melalui mekanisme *ETL* serta analisis lanjutan pada tahap visualisasi data.

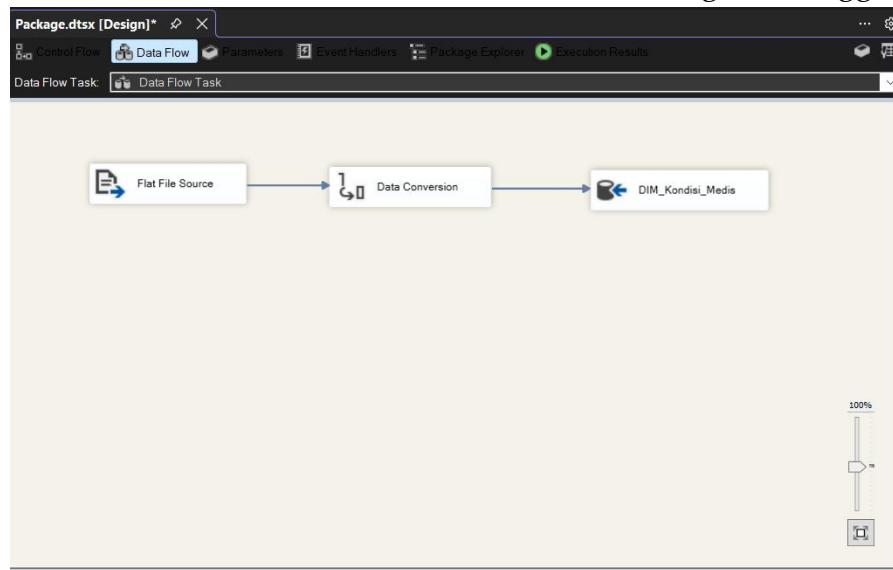
Analysis_PK	Demografi_SK	Lifestyle_SK	Kondisi_Medis...	Diabetes_Status	BMI	GenHlth	MentHlth	PhysHlth
1	1	1	1	0	40	5	18	15
2	1	1	1	0	25	3	0	0
3	1	1	1	0	28	5	30	30
4	1	1	1	0	27	2	0	0
5	1	1	1	0	24	2	3	0
6	1	1	1	0	25	2	0	2
7	1	1	1	0	30	3	0	14
8	1	1	1	0	25	3	0	0
9	1	1	1	2	30	5	30	30
10	1	1	1	0	24	2	0	0

Gambar 2. Implementasi Struktur Fisik Data Warehouse pada SQL Server 2022
Hasil Eksekusi ETL (Extract, Transform, Load)

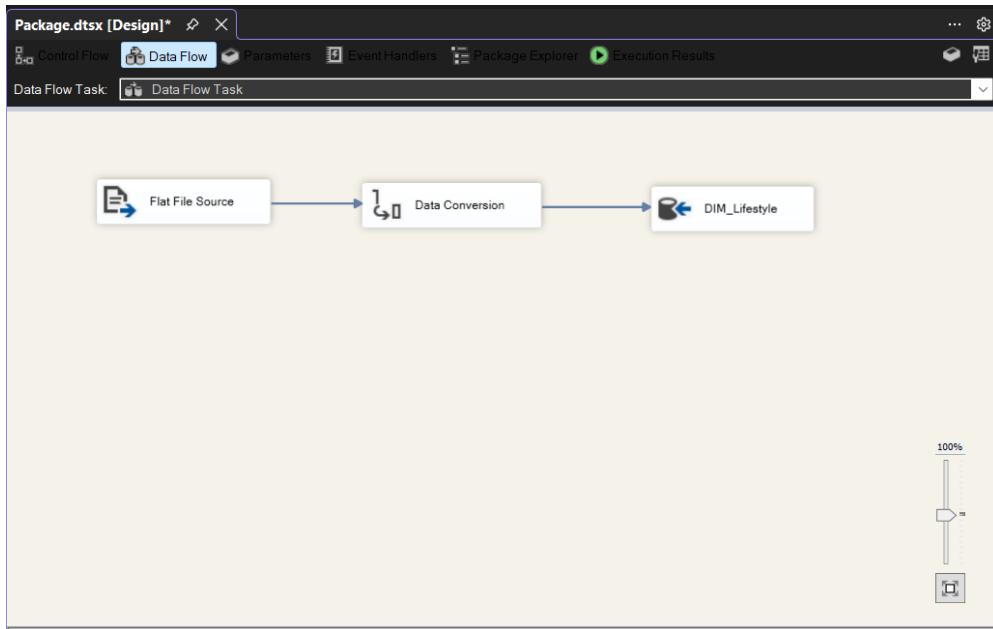
Proses *Extract, Transform, and Load (ETL)* pada penelitian ini diimplementasikan menggunakan *SQL Server Integration Services (SSIS)* dengan skenario yang terstruktur dan berurutan. Pada tahap *Control Flow*, alur eksekusi dirancang secara *sequential*, dimulai dari proses pemuatan tabel dimensi terlebih dahulu, kemudian dilanjutkan dengan pemuatan tabel fakta. Strategi ini diterapkan untuk memastikan bahwa seluruh *surrogate key* pada tabel dimensi telah tersedia sebelum digunakan sebagai *foreign key* pada tabel fakta, sehingga menghindari terjadinya kegagalan constraint referensial pada basis data.



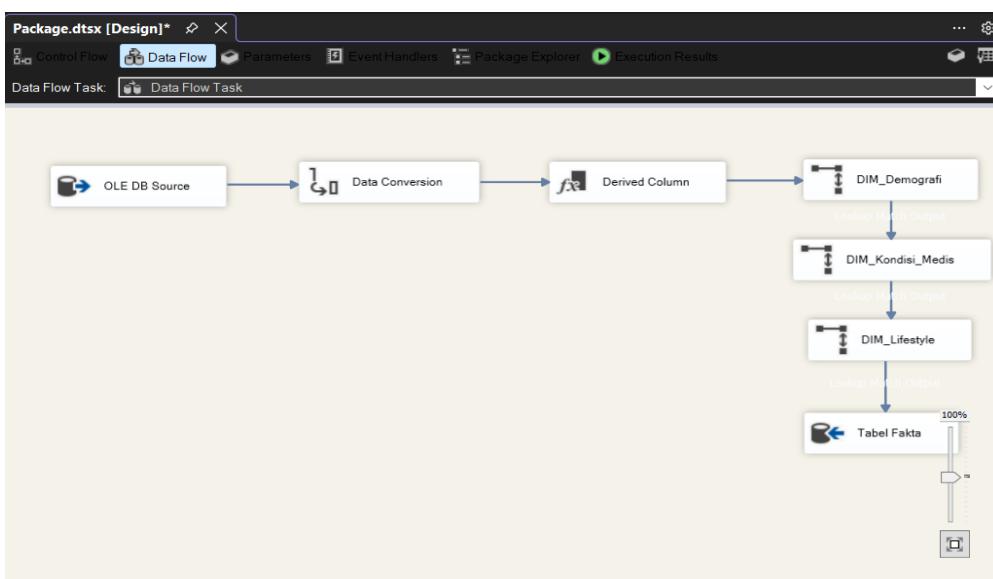
Gambar 3. Data Flow ETL untuk Pemuatan Dimensi Demografi menggunakan SSIS



Gambar 4. Data Flow ETL untuk Pemuatan Dimensi Kondisi Medis menggunakan SSIS



Gambar 5. Data Flow ETL untuk Pemuatan Dimensi Gaya Hidup menggunakan SSIS



Gambar 6. Data Flow ETL untuk Pemuatan Tabel Fakta Fact_Diabetes_Analysis

Pada tahap Data Flow untuk tabel dimensi, data diekstraksi dari sumber file datar (*flat file* berformat CSV) menggunakan komponen *Flat File Source*. Selanjutnya dilakukan proses Data Conversion untuk menyesuaikan tipe data agar sesuai dengan skema database *SQL Server*. Data yang telah dikonversi kemudian dimuat ke masing-masing tabel dimensi, yaitu *Dim_Demografi*, *Dim_Kondisi_Medis*, dan *Dim_Lifestyle*, sebagaimana ditunjukkan pada Gambar 3, Gambar 4, dan Gambar 5. Proses ini memastikan bahwa setiap tabel dimensi terisi dengan data yang konsisten dan siap digunakan dalam proses analitik.

Setelah seluruh tabel dimensi berhasil dimuat, proses *ETL* dilanjutkan dengan pengisian tabel fakta *Fact_Diabetes_Analysis*. Seperti ditunjukkan pada Gambar 6,

data diekstraksi dari database staging menggunakan *OLE DB Source*, kemudian melalui proses *Data Conversion* dan *Derived Column* untuk membentuk atribut analitik yang dibutuhkan. Selanjutnya, dilakukan proses *Lookup* secara berurutan ke tabel *Dim_Demografi*, *Dim_Kondisi_Medis*, dan *Dim_Lifestyle* untuk memperoleh *foreign key* yang sesuai dengan masing-masing dimensi.

Hasil pemantauan eksekusi paket *SSIS* menunjukkan bahwa seluruh komponen *Data Flow* berhasil dijalankan dengan indikator status hijau, yang menandakan tidak adanya error selama proses *ETL*. Sebanyak 10.000 baris data berhasil dimuat sepenuhnya ke dalam tabel fakta tanpa adanya data yang tertolak (*rejected rows*).

Keberhasilan proses ini menunjukkan bahwa arsitektur *ETL* yang dirancang efektif dalam menjaga konsistensi, integritas, dan validitas data sejak tahap ekstraksi hingga pemuatan ke *Data Warehouse*. Data yang telah bersih dan terstruktur ini selanjutnya digunakan sebagai basis utama dalam analisis visualisasi dan pembahasan pada tahap berikutnya.

Analisis Visualisasi dan Temuan

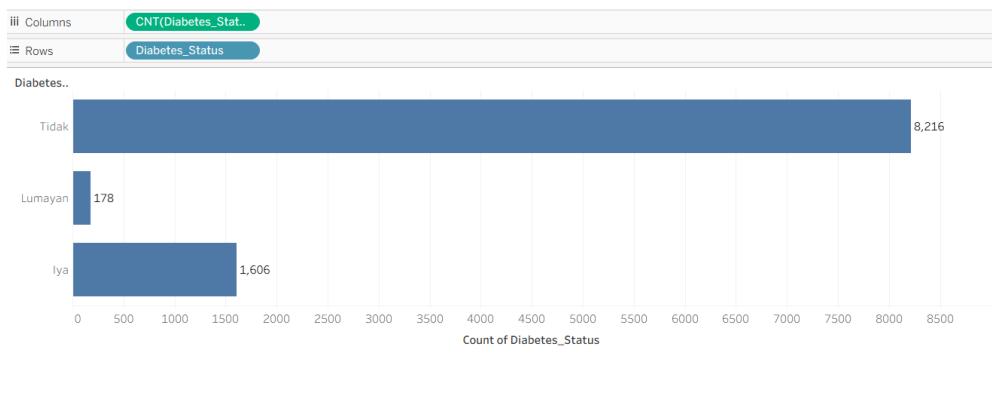
Pemanfaatan Tableau Public sebagai alat *Business Intelligence* menghasilkan *dashboard* interaktif yang memvisualisasikan korelasi antar variabel risiko diabetes. Berikut adalah temuan utama berdasarkan visualisasi data:

Distribusi Prevalensi Diabetes

Visualisasi distribusi status pasien menunjukkan bahwa dari total 10.000 data pasien yang dianalisis, sebanyak 8.216 pasien (82,16%) berada pada kategori non-diabetes, 178 pasien (1,78%) termasuk dalam kategori pre-diabetes, dan 1.606 pasien (16,06%) terdiagnosis diabetes. Distribusi ini memberikan gambaran makro mengenai kondisi kesehatan populasi sampel, di mana proporsi pasien dengan diabetes menunjukkan angka yang cukup signifikan.

Persentase penderita diabetes yang mencapai lebih dari 16% mengindikasikan tingginya prevalensi penyakit ini dalam populasi penelitian, sehingga diperlukan perhatian khusus dalam upaya pencegahan dan pengelolaan penyakit secara berkelanjutan. Sementara itu, keberadaan kelompok pre-diabetes meskipun relatif kecil tetap menjadi indikator penting bagi institusi kesehatan untuk melakukan intervensi dini guna mencegah perkembangan penyakit ke tahap yang lebih berat.

Hasil visualisasi ini menegaskan pentingnya pemanfaatan *Data Warehouse* sebagai sarana pemetaan populasi berisiko tinggi secara cepat dan akurat. Dengan struktur data yang terintegrasi, pihak manajemen dan tenaga medis dapat menggunakan informasi ini sebagai dasar dalam perencanaan strategi preventif, pengalokasian sumber daya, serta pengambilan keputusan berbasis data di bidang kesehatan [2].



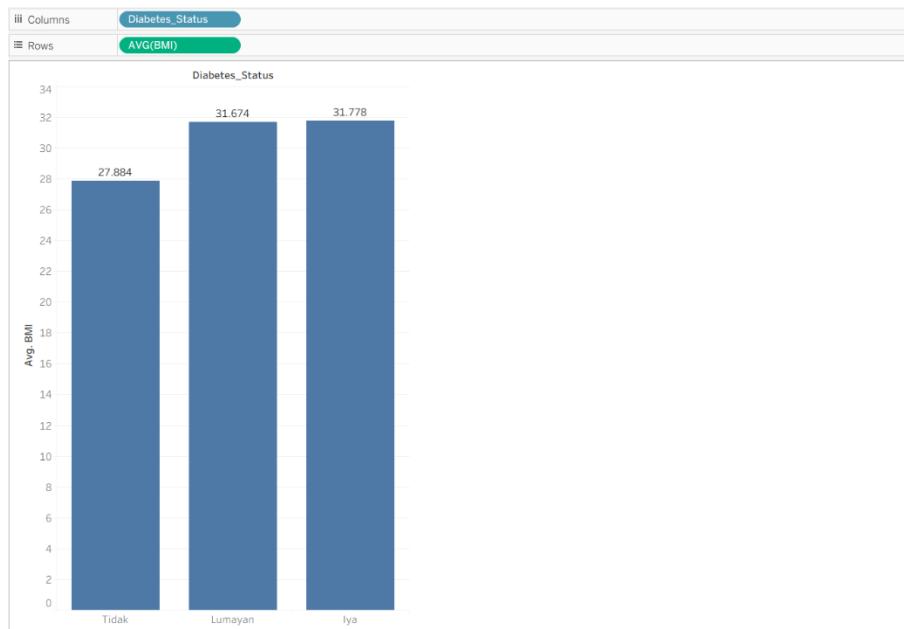
Gambar 7. Dashboard Distribusi Status Diabetes Pasien

Analisis Korelasi BMI dan Risiko Diabetes

Analisis korelasi antara *Body Mass Index (BMI)* dan status diabetes dilakukan untuk mengidentifikasi pola risiko berdasarkan kategori indeks massa tubuh pasien. Berdasarkan visualisasi pada Gambar 8, terlihat adanya tren peningkatan nilai rata-rata *BMI* seiring dengan meningkatnya status diabetes. Pasien dengan status tidak diabetes memiliki rata-rata *BMI* sebesar 27,88, yang masih berada pada kategori *overweight*. Sementara itu, pasien dengan status pre-diabetes dan diabetes masing-masing memiliki rata-rata *BMI* sebesar 31,67 dan 31,78, yang termasuk dalam kategori *obesitas*.

Pola ini menunjukkan bahwa peningkatan risiko diabetes mulai signifikan ketika *BMI* memasuki kategori *obesitas*, bahkan sebelum pasien secara klinis terdiagnosa diabetes. Temuan ini mengindikasikan bahwa fase pre-diabetes merupakan periode kritis untuk intervensi medis dan perubahan gaya hidup. Hasil ini sejalan dengan penelitian sebelumnya yang menyatakan bahwa obesitas merupakan faktor risiko utama diabetes mellitus tipe 2 karena berpengaruh terhadap resistensi insulin dan gangguan metabolisme glukosa [5].

Dari perspektif manajerial dan klinis, informasi ini sangat penting bagi fasilitas kesehatan dalam merancang program pencegahan berbasis data, seperti intervensi penurunan berat badan dan peningkatan aktivitas fisik yang ditargetkan pada kelompok pasien dengan *BMI* tinggi. Dengan demikian, *Data Warehouse* yang dibangun terbukti mampu mendukung analisis risiko kesehatan secara multidimensi dan menghasilkan wawasan yang bernilai bagi pengambilan keputusan berbasis data.



Gambar 8. Rata-Rata Body Mass Index (BMI) Berdasarkan Status Diabetes

KESIMPULAN

Penelitian ini berhasil merancang dan mengimplementasikan sebuah *Data Warehouse* berbasis skema multidimensi dengan pendekatan *Star Schema* untuk studi kasus penyakit Diabetes Melitus. Perancangan *Data Warehouse* dilakukan dengan mengintegrasikan data pasien yang sebelumnya tersebar dan terfragmentasi ke dalam satu repositori terpusat, sehingga memungkinkan proses analisis data kesehatan dilakukan secara lebih terstruktur, konsisten, dan efisien. Struktur yang dibangun, terdiri dari satu tabel fakta dan tiga tabel dimensi utama, mampu merepresentasikan kebutuhan analitik baik dari sisi klinis maupun manajerial.

Implementasi proses *Extract, Transform, and Load (ETL)* menggunakan *SQL Server Integration Services (SSIS)* berhasil memuat seluruh data pasien sebanyak 10.000 record ke dalam *Data Warehouse* tanpa kehilangan maupun duplikasi data. Penggunaan *surrogate key* pada tabel dimensi terbukti mampu menjaga konsistensi historis data serta mendukung performa query analitik. Hasil ini menunjukkan bahwa arsitektur *Data Warehouse* yang dirancang telah memenuhi prinsip-prinsip dasar perancangan sistem analitik dan siap digunakan untuk analisis multidimensi.

Hasil analisis dan visualisasi data menggunakan Tableau Public menunjukkan bahwa *Data Warehouse* yang dibangun mampu menghasilkan wawasan yang bernilai. Analisis korelasi antara *Body Mass Index (BMI)* dan status diabetes memperlihatkan bahwa kelompok pasien dengan status diabetes memiliki rata-rata *BMI* yang berada pada kategori *obesitas*, lebih tinggi dibandingkan kelompok non-diabetes. Temuan ini mengonfirmasi bahwa struktur *Data Warehouse* yang dirancang efektif dalam mendukung identifikasi pola risiko kesehatan dan eksplorasi data secara komprehensif.

Secara keseluruhan, penelitian ini membuktikan bahwa penerapan *Data Warehouse* pada domain kesehatan dapat meningkatkan efisiensi pengolahan data, mempercepat proses analisis, serta mendukung pengambilan keputusan berbasis data bagi tenaga medis dan manajemen fasilitas kesehatan. Meskipun demikian, penelitian ini masih memiliki keterbatasan, antara lain penggunaan dataset sekunder dan belum diterapkannya pembaruan data secara *real-time*. Oleh karena itu, penelitian selanjutnya disarankan untuk mengintegrasikan data *real-time*, memperluas variabel klinis, serta mengembangkan analisis prediktif guna menghasilkan sistem analitik kesehatan yang lebih cerdas dan berkelanjutan.

DAFTAR PUSTAKA

- [1] W. Suharso *et al.*, "IMPLEMENTASI STAR SCHEMA PADA STUDI KASUS PERPUSTAKAAN," vol. 4, no. 1, pp. 1-11, 2021.
- [2] A. Ado, A. Aliyu, S. A. Bello, and A. S. Gezawa, "Building a Diabetes Data Warehouse to Support Decision making in healthcare industry," vol. 16, no. 2, pp. 138-143, 2014.
- [3] D. Analytics, "Business Intelligence Data Visualization for Diabetes Health Prediction Data Analytics and Insights for Diabetes Prediction," vol. 14, no. 1, pp. 818-831, 2023.
- [4] E. T. Inau *et al.*, "The Journey to a FAIR CORE DATA SET for Diabetes Research in Germany," pp. 1-13, 2024, doi: 10.1038/s41597-024-03882-0.
- [5] I. Z. Sadiq *et al.*, "Data-driven diabetes mellitus prediction and management : a comparative evaluation of decision tree classifier and artificial neural network models along with statistical analysis," pp. 1-16, 2025.
- [6] K. S. Adewole, E. Alozie, H. Olagunju, N. Faruk, and R. Yusuf, *A systematic review and meta - data analysis of clinical data repositories in Africa and beyond : recent development , challenges , and future directions*. Springer International Publishing, 2024. doi: 10.1007/s44248-024-00012-4.
- [7] A. D. Detection, "Analyzing Diabetes Detection and Classification: A Bibliometric Review (2000-2023)," 2024.
- [8] R. D. Thantilage, N. Le-khac, and M. Kechadi, "Informatics in Medicine Unlocked Healthcare data security and privacy in Data Warehouse architectures ☆," *Informatics Med. Unlocked*, vol. 39, no. May, p. 101270, 2023, doi: 10.1016/j.imu.2023.101270.
- [9] B. Lockee *et al.*, "Establishment of a Diabetes-Tailored Data Intelligence Platform Enhances Clinical Care , Enables Risk-Based Monitoring , and Facilitates Population-Health-Based Approaches at a Pediatric Diabetes Network," no. X, 2025, doi: 10.1177/19322968251367776.
- [10] K. Kim, M. Hong, S. G. Lee, H. Chang, and T. H. Kim, "Trends in Health Expenditures for Chronic Disease Management: Hypertension , Diabetes

Mellitus , and Dyslipidemia from 2013 to 2020,” vol. 66, no. 12, pp. 874–882, 2025.

[11] O. Article, F. Amin, M. Imran, S. A. Hafeez, and B. Zehra, “Diabetes and its associated factors : A Retrospective cohort analysis of a large database at Indus Hospital Health Network,” vol. 40, no. 2, pp. 10–14, 2024.

[12] A. Obbo, W. Wasswa, and A. Basaza-ejiri, “Leveraging Data Warehousing for Healthcare Delivery – A case of Outpatient Department Leveraging Data Warehousing for Healthcare Delivery – A Case of Outpatient Department,” 2024, doi: 10.20944/preprints202406.1899.v1.

[13] S. Malunjkar, S. Weber, and S. Datta, “A highly scalable repository of waveform and vital signs data from bedside monitoring devices,” pp. 1–12.

[14] Z. Li, X. Liu, Z. Tang, P. Zhang, and N. Jin, “TrajVis : a visual clinical decision support system to translate artificial intelligence trajectory models in the precision management of chronic kidney disease”.

[15] R. H. Systems *et al.*, “Design and Implementation of a Scalable Clinical Data Warehouse for Resource-Constrained Healthcare Systems”.

[16] G. Turcan and S. Peker, “A multidimensional data warehouse design to combat the health pandemics,” *J. Data, Inf. Manag.*, pp. 371–386, 2022, doi: 10.1007/s42488-022-00082-6.

[17] Y. Fu *et al.*, “Stacking model framework reveals clinical biochemical data and dietary behavior features associated with type 2 diabetes : A retrospective cohort study,” vol. 046111, pp. 1–15, 2024, doi: 10.1063/5.0207658.

[18] V. Burda, J. Schneider, and D. Nov, “Managing Diabetes Using Mobiab : Long-Term Case Study of the Impact of a Mobile App on Self-management Corresponding Author :,” vol. 7, pp. 1–12, 2022, doi: 10.2196/36675.

[19] K. Kim, B. Kim, K. Han, K. Kim, B. Kim, and K. Han, “Big Data Research for Diabetes-Related Diseases Using the Korean National Health Information Database Big Data Research for Diabetes-Related Diseases Using the Korean National Health Information Database,” pp. 13–21, 2025.

[20] A. A. Abu-shareha, H. Qutaishat, and A. Al-khayat, “A Framework for Diabetes Detection Using Machine Learning and Data Preprocessing,” vol. 5, no. 4, pp. 1654–1667, 2024.

[21] Z. Wang *et al.*, “Clinical Data Warehousing : A Scoping Review,” vol. 4, no. 1, pp. 1–19, 2024.

[22] I. Abousaber, H. F. Abdallah, and H. El-ghaish, “Robust predictive framework for diabetes classification using optimized machine learning on imbalanced datasets,” 2022.

[23] I. Madiha, B. Nikitha, S. Trupthi, M. Aamir, and V. Ambika, “Designing a data warehouse for healthcare analytics using snowflake and big query-A Review,” vol. 10, no. 9, pp. 5–10, 2025.

- [24] R. Panthong, "Data Warehouse System to Support Condition Health Care of Elderly," 2024, doi: 10.3844/jcssp.2024.1602.1609.
- [25] A. Lamer, C. Saint-dizier, N. Paris, and E. Chazard, "Data Lake , Data Warehouse , Datamart , and Feature Store : Their Contributions to the Complete Data Reuse Pipeline," vol. 12, pp. 1-7, 2024, doi: 10.2196/54590.
- [26] W. Chen *et al.*, "Research data warehouse: using electronic health records to conduct population-based observational studies," vol. 6, no. May, 2023.
- [27] F. Pecoraro, D. Luzi, and F. L. Ricci, "A clinical data warehouse architecture based on the electronic healthcare record infrastructure," *Heal. 2014 - 7th Int. Conf. Heal. Informatics, Proceedings; Part 7th Int. Jt. Conf. Biomed. Eng. Syst. Technol. BIOSTEC 2014*, pp. 287-294, 2014, doi: 10.5220/0004764502870294.
- [28] A. Labellapansa and A. Yulianti, "Gross Death Rate (GDR), Anasthesia Death Rate (ADR), Post Operasi Death Rate (PODR), Post Interactive Death Rate (PIDR), Normal Tissue Removal Rate (NTRR), Maternal Death Rate (MDR)," no. September, 2012.
- [29] D. Hafiz, A. Rahman, and K. Kunci, "Perancangan Data Warehouse Untuk Pasien Rawat Inap Pada Rumah Sakit Siti Fatimah Az-Zahra," *Mdp Student Conf.*, pp. 241-247, 2024.
- [30] M. Doutreligne, A. Degremont, P. A. Jachiet, A. Lamer, and X. Tannier, "Good practices for clinical data warehouse implementation: A case study in France," *PLOS Digit. Heal.*, vol. 2, no. 7, 2023, doi: 10.1371/journal.pdig.0000298.