

Sistem Personalisasi Pasien Diabetes Berbasis Domain Fitur Menggunakan Algoritma Gradient Boosting Machine

*Feature Domain-Based Diabetes Patient Personalization System Using
Gradient Boosting Machine Algorithm*

Tutus Praningki¹, David Thanlian Kurniawan²

¹Informatika, Universitas Pignatelli Triputra

²Rekayasa Perangkat Lunak, Universitas Pignatelli Triputra

E-mail: ¹praningki86@gmail.ac.id, ² my.upitra3344@gmail.ac.id

Abstrak

Jumlah penderita penyakit diabetes diproyeksikan akan terus meningkat dalam beberapa tahun mendatang. Pengukuran glukosa darah, olahraga, diet, dan pengobatan farmakologis adalah beberapa metode yang dapat digunakan untuk penanganan pasien penderita penyakit diabetes. Penggunaan metode yang sama tidak selalu efektif untuk setiap pasien diabetes, karena respons setiap individu terhadap penggunaan metode terapi dapat berbeda-beda. Pendekatan berbasis Machine Learning telah banyak digunakan untuk penanganan penyakit diabetes, baik untuk deteksi dini maupun proses perawatan pasien diabetes. Pada konteks penanganan pasien diabetes mellitus, pemilihan terapi yang tepat bagi setiap pasien sangat penting untuk dapat mencapai kontrol glikemik yang baik dan mencegah komplikasi jangka panjang. Permasalahan dalam penanganan pasien diabetes adalah menentukan model yang tepat untuk setiap pasien yang berbeda, sehingga pendekatan model berbasis domain fitur menjadi sangat penting untuk diterapkan. Domain fitur yang digunakan dalam penelitian ini adalah demografi, riwayat medis, dan gaya hidup. Hasil pengujian didapatkan bahwa domain fitur riwayat medis menjadi faktor penting untuk deteksi kenaikan kadar gula dalam darah pasien diabetes. Tingkat akurasi yang didapatkan algoritma GBM dengan menggunakan domain fitur demografi, riwayat medis, dan gaya hidup adalah 96%. Dengan hasil pengujian aplikasi rekomendasi personalisasi pasien diabetes dapat digunakan oleh tenaga medis pada program Prolanis.

Kata kunci: Gradient Boosting Machine, Machine Learning, Diabetes, Domain Fitur

Abstract

The number of people with diabetes is projected to continue to increase in the coming years. Blood glucose measurement, exercise, diet, and pharmacological treatment are some of the methods that can be used to manage patients with diabetes. The use of the same method is not always effective for every diabetic patient, as each individual's response to the use of therapy methods may vary. Machine Learning-based approaches have been widely used for diabetes management, both for early detection and the treatment process of diabetic patients. In the context of managing patients with diabetes mellitus, choosing the right therapy for each patient is very important to achieve good glycemic control and prevent long-term complications. The problem in managing diabetic patients is determining the right model for each different patient, so a feature domain-based model approach is very important to apply. The feature domains used in this study are demographics, medical history, and lifestyle. The test results showed that the medical history feature domain is an important factor for the detection of elevated blood sugar levels in diabetic patients. The accuracy rate obtained by the GBM algorithm using the demographic, medical history, and lifestyle feature domains is 96%. With the test results, the application of personalized recommendations for diabetic patients can be used by medical personnel in the Prolanis program.

Keywords: Gradient Boosting Machine, Machine Learning, Diabetes, Feature Domains

1. PENDAHULUAN

Penyakit diabetes mellitus telah menjadi salah satu masalah kesehatan utama di seluruh dunia. Data dari *International Diabetes Federation* menunjukkan bahwa sekitar 537 juta orang di seluruh dunia menderita diabetes pada tahun 2021. Jumlah penderita penyakit diabetes diproyeksikan akan terus meningkat dalam beberapa tahun mendatang. Pengukuran glukosa darah, olahraga, diet, dan pengobatan farmakologis adalah beberapa metode yang dapat digunakan untuk penanganan pasien penderita penyakit diabetes. Penggunaan metode yang sama tidak selalu efektif untuk setiap pasien diabetes, karena respons setiap individu terhadap penggunaan metode terapi dapat berbeda-beda.

Pendekatan berbasis *Machine Learning* (ML) telah banyak digunakan untuk penanganan penyakit diabetes, baik untuk deteksi dini [1] maupun proses perawatan pasien diabetes [2]. Pada konteks penanganan pasien diabetes mellitus, pemilihan terapi yang tepat untuk setiap pasien sangat penting untuk dapat mencapai kontrol glikemik yang baik dan mencegah komplikasi jangka panjang. Pada penelitian oleh [3] teknologi ML dapat dimanfaatkan untuk mendukung keputusan klinis terkait pemilihan obat pada pasien diabetes, sedangkan pada penelitian [4] berfokus pada kontrol gula darah pasien diabetes. Algoritma ML [3] seperti *Naive Bayes* (NB), *Artificial Neural Networks* (ANN), *Random Forest* (RF), *Support Vector Machine* (SVM), dan *Gradient Boosting Machine* (GBM) direkomendasikan untuk digunakan pada proses terapi personalisasi pasien diabetes. Algoritma GBM [5] memiliki tingkat akurasi 96,92%, sehingga lebih baik dari algoritma lain seperti NB 87,43%, SVM 96,15%, RF 95,38%, dan ANN 95,12%.

Permasalahan dalam penanganan pasien diabetes adalah menentukan model yang tepat untuk setiap pasien yang berbeda, beberapa model telah dikembangkan seperti pada [6], namun model yang dikembangkan hanya berdasarkan pada data klinis dan memiliki kekurangan terkait tidak digunakannya faktor gaya hidup pasien. Pada penelitian [7] domain fitur bersumber pada dua faktor, yaitu data klinis dan gaya hidup. Proses rekomendasi penanganan terapi pada pasien diabetes juga dapat dipengaruhi oleh faktor genetik dan demografis pasien diabetes [8], sehingga domain fitur yang dapat dikombinasikan untuk penanganan pasien diabetes adalah riwayat medis, gaya hidup, dan demografis.

Algoritma-algoritma ML telah digunakan untuk personalisasi pasien diabetes [8], namun sangat penting untuk menyesuaikan terapi untuk setiap pasien diabetes karena setiap pasien memiliki karakteristik yang berbeda, termasuk: faktor genetik, respons terhadap pengobatan, pola gaya hidup, dan kondisi kesehatan. Ketidaksesuaian pola penanganan atau terapi dapat menyebabkan kondisi kadar gula yang semakin buruk [9], munculnya komplikasi kesehatan, dan risiko kematian yang lebih tinggi. Oleh sebab itu pemanfaatan data riwayat medis, gaya hidup, demografis dan algoritma GBM untuk sistem rekomendasi terapi personalisasi pasien diabetes diusulkan agar dapat menentukan pola penanganan pasien diabetes secara tepat. Fokus pada penelitian ini membangun sebuah sistem rekomendasi untuk pasien diabetes menggunakan algoritma GBM.

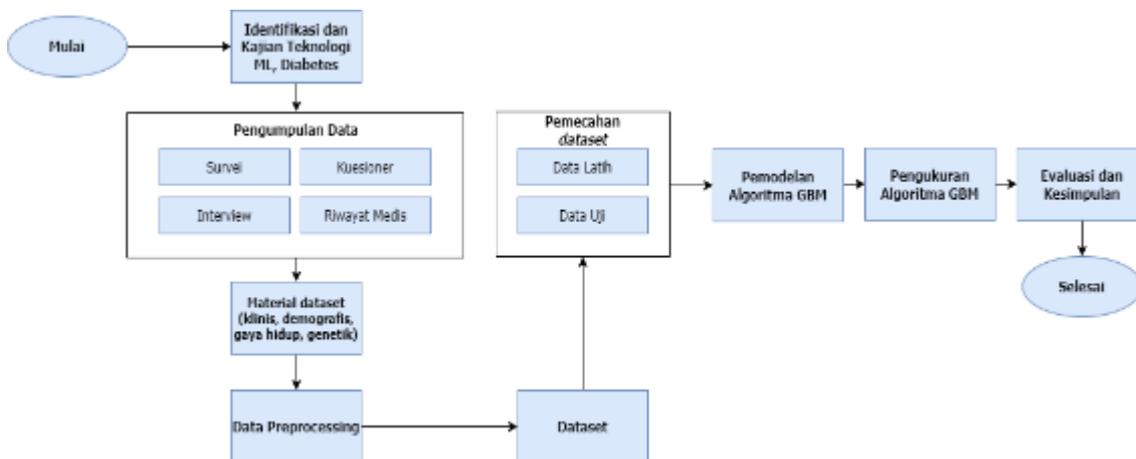
Permasalahan penggunaan metode berbasis ML adalah pada pendekatan yang digunakan untuk meningkatkan kinerja algoritma. Salah satu metode yang populer untuk meningkatkan kinerja algoritma berbasis ML adalah seleksi fitur. Seleksi fitur merupakan metode yang dapat digunakan untuk reduksi dimensi data [10]. Hasil penelitian [10] [11] [12] menunjukkan bahwa penggunaan seleksi fitur pada dataset dapat meningkatkan kinerja algoritma ML, pada penelitian tersebut model ML dibuat berdasarkan fitur-fitur yang telah diseleksi dan dilakukan eliminasi. Pada penelitian ini proses pengujian untuk menentukan fitur yang memiliki pengaruh terhadap hasil klasifikasi adalah menggunakan pendekatan berbasis domain fitur. Pendekatan berbasis domain fitur berfokus pada pembuatan model ML berdasarkan fitur-fitur yang telah dikelompokkan berdasarkan ciri-ciri pada fitur tersebut.

Berdasarkan latar belakang tersebut, rumusan permasalahan pada penelitian ini adalah menggunakan pendekatan berbasis domain fitur, yaitu riwayat medis, gaya hidup, dan demografis menjadi dataset, serta menggunakan dataset tersebut untuk mengkonstruksi pola terapi pasien diabetes menggunakan algoritma GBM. Hasil evaluasi algoritma GBM menjadi acuan untuk

membangun sebuah sistem yang dapat memberikan rekomendasi personalisasi pada pasien diabetes yang mengikuti program Prolanis.

2. METODE PENELITIAN

Pada penelitian ini menggunakan beberapa tahapan dengan kegiatan yang diantaranya sebagai berikut (Gambar 1):



Gambar 1 Diagram alur penelitian

2.1 Identifikasi dan Kajian

Identifikasi dilakukan dengan melihat bagaimana pasien diabetes dirawat di program Prolanis di kota Kediri. Hasil observasi dan metaanalisis [13], metode observasi dalam penelitian ini bertujuan untuk memahami karakteristik pasien diabetes, kondisi klinis pasien, serta variabel-variabel lain yang relevan untuk personalisasi perawatan. Observasi memberikan gambaran awal yang penting dalam membangun model berbasis dataset yang akan digunakan dalam algoritma *Gradient Boosting Machine* (GBM). Metaanalisis digunakan untuk menganalisis secara sistematis dan kuantitatif berbagai penelitian terdahulu yang relevan dengan personalisasi pasien diabetes. Metode Metaanalisis digunakan untuk mendapatkan pemahaman mendalam dan berbasis bukti untuk mendukung algoritma GBM dalam pengolahan domain fitur dataset yang lebih akurat dan relevan.

2.2 Pengumpulan Data

Survei, kuisisioner, interview dan data riwayat klinis [14] digunakan untuk mengumpulkan data primer terkait pasien diabetes, penggunaan 4 metode bertujuan mendapatkan aspek domain fitur dataset yang lebih beragam. Survei, kuisisioner, *interview* dan data riwayat klinis dilakukan di beberapa lokasi penyelenggaraan Prolanis [15] di kota Kediri. Sampel terdiri dari semua pasien diabetes yang mengikuti program Prolanis di beberapa Puskesmas.

2.3 Data Preprocessing

Sebelum dilakukan proses pelatihan model *machine learning*, perlu dilakukan data preprocessing [16] untuk memastikan kesesuaian dan kualitas pada dataset. Langkah-langkah yang dilakukan pada saat *data preprocessing* adalah sebagai berikut:

- Pembersihan data: pada tahap ini dilakukan pembersihan data dari data yang tidak relevan, *missing values*, duplikat, dan *outlier*.
- Normalisasi: Fitur-fitur dengan tipe data bilangan real dilakukan proses normalisasi ke skala standar untuk memastikan memiliki pengaruh yang sama.
- Dataset Split*: Dilakukan pemecahan terhadap dataset menjadi dua bagian, data latih dan data uji. Metode yang digunakan untuk pemecahan dataset adalah *Holdout*, proses pemecahan

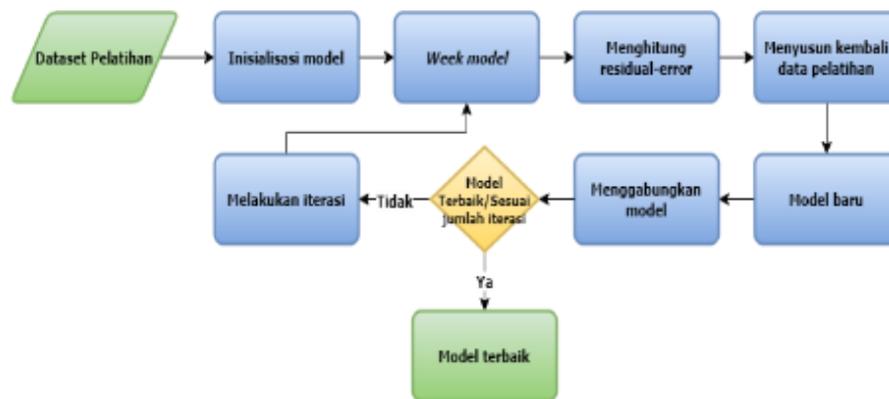
dataset dilakukan secara acak atau *shuffled sampling*. Data latih digunakan untuk proses pembelajaran algoritma GBM, dan data uji digunakan untuk proses pengujian algoritma GBM

d. *Class Balancing* [17] : Teknik seperti *oversampling* atau *undersampling* digunakan jika *dataset* tidak seimbang untuk memastikan setiap kelas terwakili secara memadai dalam proses pelatihan model.

2.4 Pemodelan Algoritma GBM

Mengembangkan model menggunakan algoritma GBM [18] dengan memanfaatkan data yang diproses untuk memberikan rekomendasi terapi diabetes yang disesuaikan dengan kondisi personal pasien diabetes. Pada kegiatan ini mencakup pemilihan fitur pada dataset, pemilihan model, tuning parameter, dan validasi model. Proses GBM [19] dalam membangun model prediktif pada data pelatihan adalah sebagai berikut (gambar 2):

- a. Membangun inisialisasi model
Model awal dibuat sebagai model konstan yang merupakan rata-rata atau median dari variabel terikat.
- b. Membangun model lemah
Model lemah dibuat sebagai model yang mampu memprediksi kesalahan dari model sebelumnya.
- c. Menghitung residual *error*
Residual *error* dihitung sebagai selisih antara nilai prediksi dari model sebelumnya dan nilai asli dari variabel terikat.
- d. Menyusun kembali data latih
Data latih diubah dengan menggunakan residual *error* sebagai variabel terikat.
- e. Membuat model baru
Model baru dibuat dengan memprediksi residual *error* yang dihasilkan dari model sebelumnya.
- f. Menggabungkan model
Model baru yang dibuat digabungkan dengan model sebelumnya untuk membentuk model yang lebih baik.
- g. Melakukan iterasi secara berulang
Tahap b-f diulang hingga mencapai kondisi berhenti yang telah ditentukan, seperti jumlah iterasi yang telah ditentukan atau ketika model tidak mengalami peningkatan yang signifikan.



Gambar 2 Proses pembuatan model algoritma GBM

2.5 Evaluasi Algoritma GBM

Melakukan validasi klinis [20] terhadap model yang dikembangkan oleh algoritma GBM dengan menggunakan data uji. Proses pengujian algoritma GBM dilakukan dengan melakukan simulasi kombinasi dari domain fitur pada dataset, tujuannya agar diketahui domain fitur yang

paling mempengaruhi kenaikan gula dalam darah pasien diabetes. Pada tahap ini metode yang digunakan untuk melakukan evaluasi pada algoritma GBM adalah sebagai berikut [21]:

- a) Akurasi
Akurasi adalah metrik utama untuk menilai kinerja model klasifikasi secara keseluruhan. Akurasi mengukur proporsi contoh yang diprediksi dengan benar di antara semua contoh dalam kumpulan data.
- b) Presisi
Berfungsi mengukur proporsi contoh positif yang diprediksi dengan benar (*true positif*) di antara semua contoh yang diprediksi sebagai positif (*true positif + false positif*)
- c) Recall
Recall juga dikenal sebagai sensitivitas atau tingkat *true positif*, menghitung proporsi contoh positif yang diprediksi dengan benar (*true positif*) di antara semua contoh positif yang sebenarnya (*true positif + false negatif*).
- d) F1 Score
Merupakan rata-rata harmonisasi dari presisi dan *recall*, yang menawarkan keseimbangan antara kedua metrik tersebut. Hal ini sangat berguna untuk set data yang tidak seimbang, menyediakan satu metrik untuk menilai kinerja model. Skor F1, berkisar mulai dari 0 hingga 1 untuk menunjukkan kinerja model, dengan skor yang lebih tinggi menunjukkan tingkat akurasi yang lebih baik.

2.6 Pengujian Aplikasi

Pengujian aplikasi menggunakan metode perbandingan antara hasil deteksi aplikasi yang dikembangkan dengan *ML tool Python Library*. Perbandingan hasil deteksi berfungsi untuk memastikan bahwa implementasi algoritma GBM ke dalam bahasa pemrograman yang digunakan sudah sesuai. *Machine learning tool Python Library* digunakan sebagai pembanding karena sudah terbukti pada area penelitian berbasis ML [1] [22] [23].

3. HASIL DAN PEMBAHASAN

3.1 Dataset

Hasil dari pengumpulan data didapatkan data penderita diabetes dari 7 Puskesmas di kota Kediri Jawa Timur, total data yang telah didapatkan disajikan pada tabel 1.

Tabel 1 Jumlah Pasien

No.	Puskesmas	Jumlah Pasien Diabetes Prolanis
1	Balowerti	112
2	Campurejo	97
3	Pesantren 1	120
4	Pesantren 2	117
5	Ngletih	89
6	Mrican	115
7	Sukorame	121
	Total	771

Dari Tabel 1 jumlah data yang didapatkan sebanyak 771 record, dengan jumlah fitur sebanyak 18, seperti yang ditunjukkan pada tabel 2.

Tabel 2 Fitur-Fitur Dataset

No	Fitur	Deskripsi	Tipe data	Domain Fitur
1	Usia	Usia pasien diabetes (45-76)	Numerik	Demografi
2	Jenis_Kelamin	Jenis kelamin pasien (0=Perempuan; 1= Laki-laki)	Kategorikal yang dinumerikkan	Demografi
3	BMI	Indeks masa tubuh pasien (12-37)	Numerik	Riwayat Medis
4	Merokok	Pasien merokok (0= Tidak; 1= Ya)	Kategorikal yang dinumerikkan	Gaya Hidup
5	Trestbps	Tekanan darah pasien (100 - 185)	Numerik	Riwayat Medis

No	Fitur	Deskripsi	Tipe data	Domain Fitur
6	Frekuensi_Kehadiran	Frekuensi mengikuti kegiatan prolanis dalam 1 bulan (1-4)	Numerik	Gaya Hidup
7	Konsumsi_Sayur_Buah	Mengonsumsi buah atau sayuran tiap hari (0= Tidak; 1= Ya)	Kategorikal yang dinumerikkan	Gaya Hidup
8	Frekuensi_Minuman_Berpemanis	Frekuensi mengonsumsi minuman berpemanis dalam 1 hari (1-5)	Numerik	Gaya Hidup
9	Konsumsi Alkohol	Mengonsumsi minuman berakohol (0= Tidak; 1= Ya)	Kategorikal yang dinumerikkan	Gaya Hidup
10	Kandungan_Obat_Diabetes	Kandungan obat diabetes yang dikonsumsi (0= Gliclazide, 1= Glibenclamide; 2= Metformin; 3= Glimepiride; 4= Insulin Aspart)	Kategorikal yang dinumerikkan	Riwayat Medis
11	Tipe_Diabetes	Tipe diabetes yang diderita pasien	Kategorikal yang dinumerikkan	Riwayat Medis
12	Lipid Profile	Jumlah total kandungan kolesterol dalam darah (176-320)	Numerik	Riwayat Medis
13	HbA1c	Kadar glukosa darah (3,5-9)	Numerik	Riwayat Medis
14	Frekuensi_Konsumsi_Karbohidrat	Frekuensi mengonsumsi karbohidrat dalam sehari (2-4)	Numerik	Gaya Hidup
15	Durasi Tidur	Durasi tidur pasien dalam sehari (6-11)	Numerik	Gaya Hidup
16	Durasi_Diabetes	Lama pasien menderita diabetes (5-20)	Numerik	Riwayat Medis
17	Riwayat_Keluarga	Memiliki riwayat keluarga penderita diabetes (0= Tidak; 1= Ya)	Kategorikal yang dinumerikkan	Demografi
18	Output	Normal; Naik		

Fitur-fitur dataset pada tabel 2 menunjukkan terdapat 17 fitur sebagai prediktor, dan fitur Output sebagai kelas atau target prediksi. Pada fitur Output sebanyak 574 data masuk dalam kelas normal, dan 197 data masuk ke dalam kelas naik. Tipe data bernilai numerik sebanyak 10 fitur, dan 6 fitur bertipe data kategorikal yang dinumerikkan.

Hasil pengumpulan data didapatkan domain fitur pada dataset, pengelompokkan fitur-fitur ke dalam domain fitur ditampilkan pada tabel 3.

Tabel 3 Pengelompokkan Fitur Dataset

No	Domain Fitur	Fitur	Jumlah Fitur
1	Demografi	Usia, Jenis Kelamin, Riwayat Keluarga	3
2	Riwayat Medis	BMI, Trestbps, Kandungan_Obat_Diabetes, Tipe_Diabetes, Lipid Profile, HbA1c, Durasi_Diabetes	7
3	Gaya Hidup	Merokok, Frekuensi Kehadiran, Konsumsi Buah atau Sayur, Frekuensi Konsumsi Minuman Berpemanis, Mengonsumsi Alkohol, Frekuensi_Konsumsi_Karbohidrat, Durasi Tidur	7

3.2 Analisis Data

3.2.1. Data preprocessing

Hasil deteksi dari dataset yang telah didapatkan, terdapat fitur-fitur yang memiliki missing values. Fitur-fitur tersebut ditampilkan pada tabel 4.

Tabel 4. Missing Values

No	Fitur	Missing Values
1	Usia	0
2	Jenis_Kelamin	0
3	BMI	0
4	Merokok	15
5	Trestbps	0
6	Frekuensi_Kehadiran	0
7	Konsumsi_Sayur_Buah	0
8	Frekuensi_Minuman_Berpemanis	0
9	Konsumsi Alkohol	116
10	Kandungan_Obat_Diabetes	25
11	Tipe_Diabetes	0
12	Lipid Profile	0

No	Fitur	Missing Values
13	HbA1c	0
14	Frekuensi_Konsumsi_Karbohidrat	10
15	Durasi_Tidur	18
16	Durasi_Diabetes	0
17	Riwayat_Keluarga	0

Melalui tabel 4 dapat diketahui bahwa terdapat 5 fitur yang terdapat *missing values*. Pada fitur yang terdapat *missing values* digunakan fungsi *replace missing values*. Metode yang digunakan untuk *replace missing values* adalah menggunakan teknik *average*. Teknik *average* digunakan untuk mengisi nilai yang tidak terisi atau hilang.

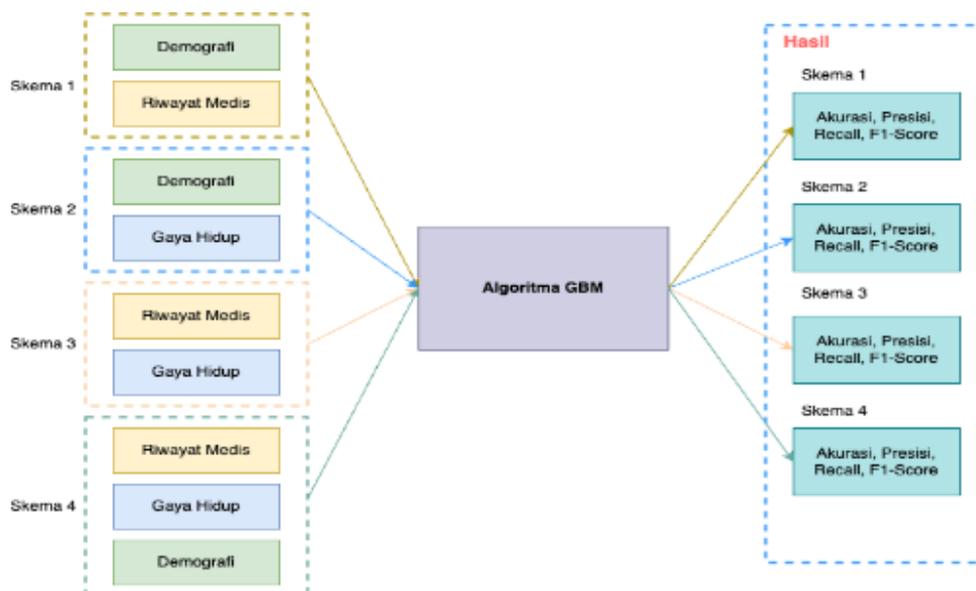
Setelah dataset dilakukan peningkatan kualitas, maka dilakukan proses pemecahan dataset menjadi dua bagian. Dataset dibagi menjadi data latih dan data uji, jumlah data latih sebanyak 540, dan 231 record untuk data uji, seperti pada tabel 5.

Tabel 5 Jumlah Dataset Latih dan Uji

No	Kelompok	Jumlah
1	Data Latih	540
2	Data Uji	231

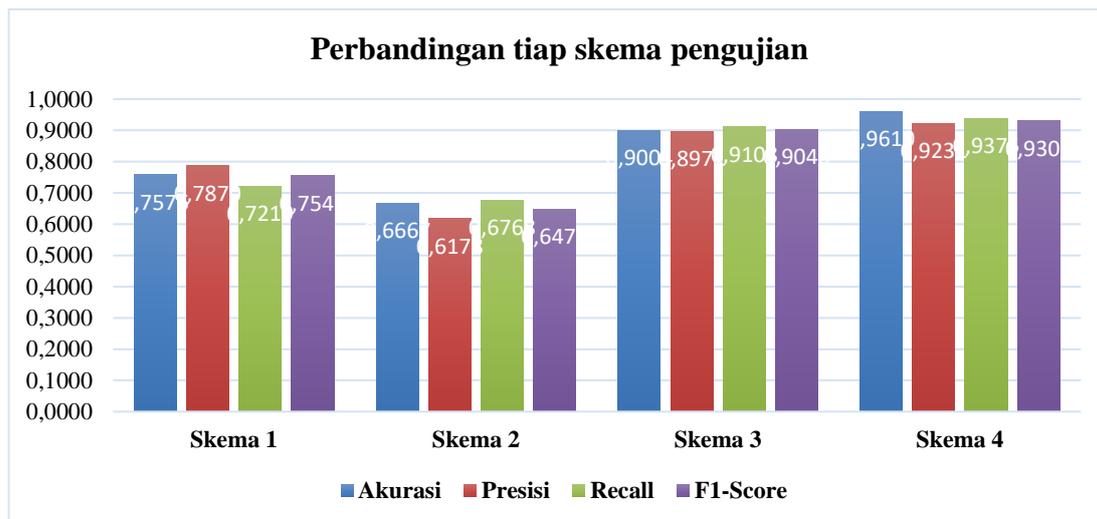
3.2.2. Pengujian Algoritma GBM

Pengujian model algoritma GBM dengan menggunakan pendekatan berbasis domain fitur pada dataset menghasilkan skema seperti pada gambar 3.



Gambar 3. Skema pengujian algoritma GBM

Pada gambar 3 terlihat terdapat 4 skema yang digunakan untuk menguji algoritma GBM, dari masing-masing skema diukur tingkat akurasi, presisi, *recall*, dan *F1-score*. Hasil evaluasi menunjukkan dengan menggunakan pengujian skema 4 menghasilkan tingkat akurasi paling tinggi, hasil pengujian ditunjukkan pada gambar 4.



Gambar 4 Grafik perbandingan akurasi, presisi, recall, dan F1-Score setiap skema

Pada grafik gambar 4 menunjukkan bahwa tingkat akurasi terbaik didapatkan oleh skema 4, skema 4 menggunakan 3 domain fitur, yaitu: data medis, gaya hidup, dan demografi. Nilai akurasi yang didapatkan oleh skema 4 adalah 0,96 atau 96%, nilai akurasi paling rendah didapatkan oleh skema 2 yaitu 0,67 atau 67%.

Tabel 6 Perbandingan Tingkat Akurasi Setiap Skema

No.	Skema	Domain Fitur	Akurasi	Presisi	Recall	F1-Score
1	Skema 1	Demografi	0,7576	0,7870	0,7219	0,7545
		Riwayat Medis				
2	Skema 2	Demografi	0,6667	0,6178	0,6768	0,6473
		Gaya Hidup				
3	Skema 3	Riwayat Medis	0,9004	0,8971	0,9108	0,9040
		Gaya Hidup				
4	Skema 4	Riwayat Medis	0,9610	0,9231	0,9375	0,9303
		Gaya Hidup				
		Demografi				

Hasil pada tabel 6 menunjukkan bahwa pada skema 1 dan 2 mengalami penurunan tingkat akurasi yang disignifikan dibandingkan dengan skema 3 dan 4. Skema 1 menggunakan domain fitur demografi dan riwayat medis, sedangkan skema 2 menggunakan domain fitur demografi dan gaya hidup. Dari tabel 6 juga dapat dicermati bahwa skema yang tidak menggunakan domain fitur riwayat medis memiliki tingkat akurasi paling rendah, dan skema 1 yang tidak menggunakan domain fitur gaya hidup memiliki tingkat akurasi paling rendah ke-2. Sehingga domain fitur riwayat medis dan gaya hidup memiliki pengaruh yang signifikan terhadap hasil pengujian menggunakan algoritma GBM. Skema 3 yang menggunakan domain fitur riwayat medis dan gaya hidup masih dapat menghasilkan tingkat akurasi sebesar 90%, sehingga didapatkan tingkat kepentingan domain fitur pada tabel 7.

Tabel 7 Tingkat Kepentingan Domain Fitur Dataset

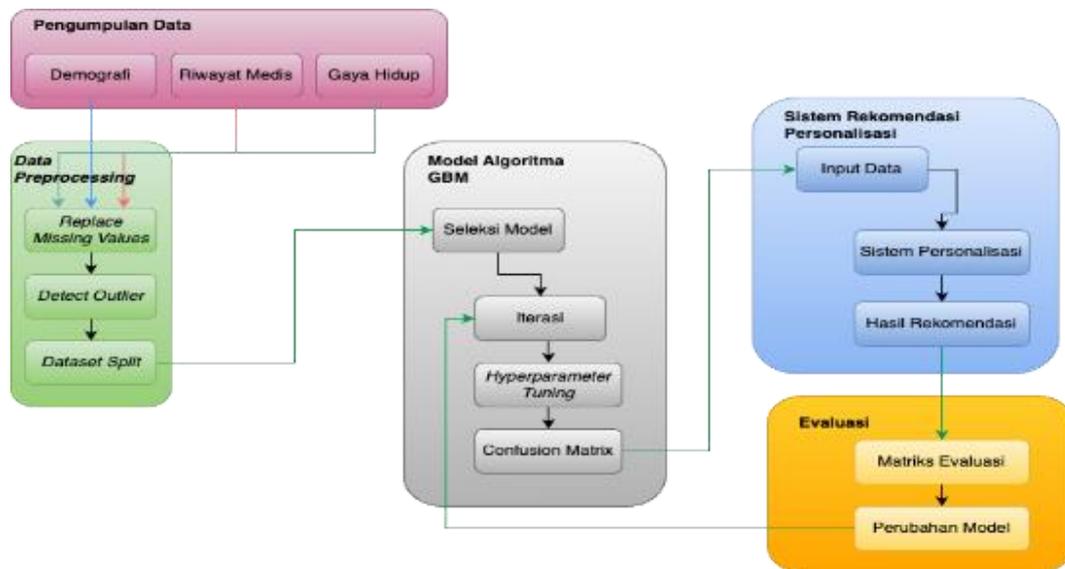
No.	Domain Fitur	Digunakan pada Skema	Rata-rata Tingkat Akurasi
1	Riwayat Medis	Skema 1, Skema 3, Skema 4	0,873
2	Gaya Hidup	Skema 2, Skema 3, Skema 4	0,8427
3	Demografi	Skema 1, Skema 2, Skema 4	0,7951

Pada tabel 7 menunjukkan bahwa domain fitur riwayat medis memiliki pengaruh signifikan terhadap hasil deteksi kenaikan kadar gula dalam darah, dengan rata-rata tingkat akurasi sebesar 87,3%. Fitur-fitur yang terdapat pada domain riwayat medis adalah BMI, *Trestbps*, *Kandungan_Obat_Diabetes*, *Tipe_Diabetes*, *Lipid_Profile*, *HbA1c*, dan *Durasi_Diabetes*. Pada penelitian [10] [11] [12] tingkat akurasi terbaik dihasilkan dari fitur yang

sudah dilakukan proses seleksi fitur, namun hasil dari penelitian ini menunjukkan perbedaan. Perbedaan tersebut adalah pada penelitian ini tingkat akurasi paling baik dihasilkan ketika menggunakan semua domain fitur, ketika salah satu domain fitur tidak digunakan, maka tingkat akurasi menjadi turun. Hasil dari pengujian ini membuktikan bahwa dalam penanganan pasien diabetes secara personal membutuhkan fitur pada dataset yang lebih bervariasi, karena faktor penyebab kenaikan kadar gula dalam darah pasien diabetes dapat berbeda-beda [9].

3.3. Diagram Arsitektur

Diagram arsitektur secara visual menyajikan sistem personalisasi untuk pasien diabetes yang mengikuti program Prolanis, serta menampilkan komponen-komponennya dan bagaimana sistem berinteraksi dengan pengguna. Diagram pada gambar 4 menunjukkan kolaborasi antara antarmuka pengguna, *backend*, basis data, dan model algoritma GBM, untuk menghasilkan sistem yang dapat membantu tenaga medis. Setiap komponen memiliki peran penting dalam memastikan fungsionalitas sistem, sehingga output yang dihasilkan oleh sistem dapat memprediksi kenaikan kadar gula dalam darah. Diagram arsitektur memberikan gambaran umum yang komprehensif mengenai struktur dan fungsionalitas sistem personalisasi, yang menggambarkan bagaimana sistem dapat memenuhi kebutuhan pasien diabetes dengan memberikan rekomendasi yang dipersonalisasi terkait kandungan gula dalam darah pasien menjadi terkontrol.



Gambar 5 Diagram arsitektur

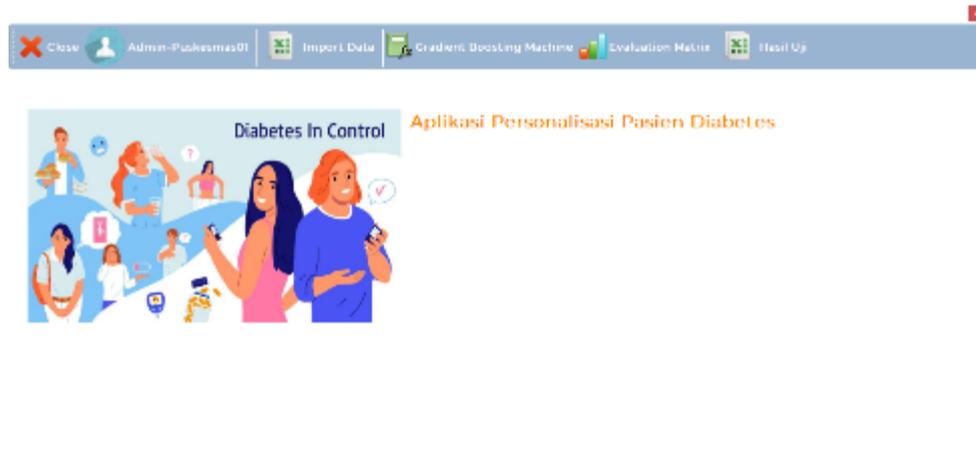
Sistem personalisasi pasien diabetes yang mengikuti program Prolanis memiliki alur proses sebagai berikut:

- a) Interaksi pengguna
Pengguna berinteraksi dengan antarmuka aplikasi, memberikan masukan data pasien diabetes.
- b) Komunikasi antarmuka pengguna dan *backend*
Antarmuka aplikasi mengirimkan input pengguna ke backend sistem untuk diproses, interaksi menggunakan sebuah pesan yang menunjukkan transfer data.
- c) Pemrosesan *backend*
Backend yang di dalamnya memiliki model algoritma GBM, memproses masukan pengguna untuk menghasilkan deteksi kemungkinan kenaikan kadar gula dalam darah, dan sistem memberikan rekomendasi personalisasi.
- d) Interaksi *backend* dan basis data
Backend sistem berinteraksi dengan basis data untuk menyimpan data hasil deteksi dan rekomendasi personalisasi.

- e) Respon basis data
Basis data merespons permintaan backend sistem, mengonfirmasi penyimpanan data telah berhasil atau memberikan informasi yang diminta oleh pengguna.
- f) Respo pengguna
Pengguna aplikasi mendapatkan respon berupa informasi dari sistem terkait hasil deteksi dan rekomendasi personalisasi pasien diabetes.

3.4 Antarmuka Aplikasi

Berikut hasil implementasi algoritma GBM untuk personalisasi pasien diabetes program Prolanis ke dalam sebuah aplikasi, gambar 6 merupakan antarmuka aplikasi pada fitur menu.



Gambar 6 Fitur Menu pada aplikasi

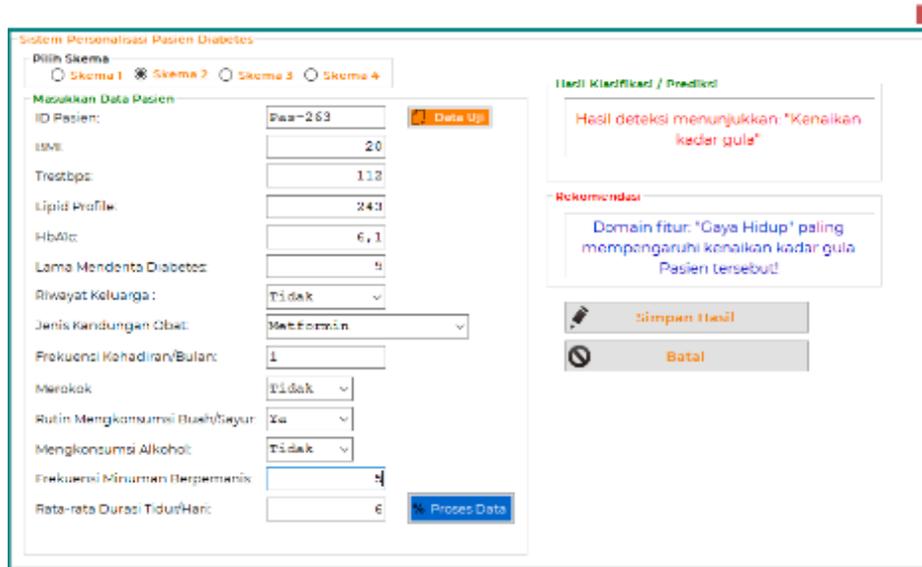
Pada fitur menu aplikasi, pengguna dapat mengakses fitur-fitur lain, seperti: import data, personalisasi menggunakan algoritma GBM, evaluasi hasil pengujian, dan ekspor data hasil pengujian. Pada gambar 7 merupakan fitur aplikasi untuk import data dari file Microsoft Excel.

Usia	Jenis	BMI	Merokok	Terdas	Frekuensi_Konsumsi_Sayur_Buah	Frekuensi_Minum_Mengonsumsi_L	Kandungan_Gl	Tipe Diabetes	Lipid
31	0	17.5	0	161	3	0	3	0	294
36	1	17	1	154	3	1	2	0	273
38	1	25	1	155	1	0	3	1	253
35	1	22	1	127	1	1	3	1	279
35	0	32	0	132	2	1	3	0	212
37	0	22	0	111	1	0	2	0	294
37	1	33	1	171	2	0	4	1	190
37	1	12	1	136	2	0	3	0	317
37	1	22	1	127	1	1	3	1	279
39	0	22	0	159	2	1	2	0	179
43	1	30	1	137	2	1	2	0	278
43	1	22	1	153	2	1	5	1	280
45	1	35	1	209	3	0	1	1	191
45	1	35	1	111	2	1	3	0	280
47	0	15	0	117	3	0	5	0	293
47	0	34	0	115	3	0	5	0	218
49	0	31.6	0	171	3	1	3	0	178
49	1	25	1	152	2	1	2	1	280

Gambar 7 Fitur Import Data pada aplikasi

Fitur Import Data digunakan untuk memasukkan data dengan format Excel, kemudian pengguna dapat melakukan penyimpanan data secara permanen ke dalam tabel di basisdata yang terhubung dengan antarmuka aplikasi. Pengguna dapat menentukan data yang telah di import disimpan ke dalam tabel data latih atau uji. Aplikasi juga menyediakan fitur utama yaitu

personalisasi pasien diabetes dengan menggunakan algoritma GBM, fitur tersebut dapat dilihat pada gambar 8.



Gambar 8 Fitur Personalisasi pada aplikasi

Pada fitur Personalisasi, pengguna dapat melakukan proses deteksi untuk pasien diabetes, deteksi yang dihasilkan berupa potensi kenaikan kadar gula dalam darah atau normal pada pasien dan rekomendasi lain berupa domain fitur yang paling berpengaruh terhadap pasien. Sebelum aplikasi mengeluarkan hasil deteksi dan rekomendasi, maka pengguna terlebih dahulu memilih skema domain fitur. Langkah berikutnya pengguna memasukkan data pasien sesuai kolom yang disediakan, setelah itu dengan menekan tombol Proses Data, maka aplikasi akan mengeluarkan hasil deteksi dan rekomendasi.

3.5 Pengujian Aplikasi

Hasil pengujian terhadap aplikasi yang telah dikembangkan menunjukkan hasil yang sama dengan hasil yang didapat oleh Python Library. Persamaan hasil deteksi disajikan pada tabel 8.

Tabel 8 Persamaan Hasil Deteksi Aplikasi dengan Python

No.	Skema	Domain Fitur	Akurasi	Presisi	Recall	F1-Score
1	Aplikasi yang dibuat	Demografi	0,7576	0,7870	0,7219	0,7545
		Riwayat Medis				
		Demografi	0,6667	0,6178	0,6768	0,6473
		Gaya Hidup				
		Riwayat Medis	0,9004	0,8971	0,9108	0,9040
		Gaya Hidup				
		Riwayat Medis	0,9610	0,9231	0,9375	0,9303
		Gaya Hidup				
2	Menggunakan tool Python Library	Demografi	0,7576	0,7870	0,7219	0,7545
		Riwayat Medis				
		Demografi	0,6667	0,6178	0,6768	0,6473
		Gaya Hidup				
		Riwayat Medis	0,9004	0,8971	0,9108	0,9040
		Gaya Hidup				
		Riwayat Medis	0,9610	0,9231	0,9375	0,9303
		Gaya Hidup				

Dari tabel 8 diketahui tingkat persamaan hasil deteksi antara aplikasi yang telah dibuat dengan Python adalah 100%, sehingga implementasi algoritma GBM ke dalam bahasa pemrograman

telah berhasil. Melalui pengujian dengan pendekatan domain fitur, nilai akurasi, presisi, recall, dan F1-Score menghasilkan nilai yang sama antara aplikasi yang dibuat dengan *tool Python Library*.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil evaluasi terhadap algoritma GBM menggunakan pendekatan berbasis domain fitur pada dataset didapatkan bahwa dengan menggunakan tiga domain fitur, yaitu: Riwayat Medis, Gaya Hidup, dan Demografi menghasilkan tingkat akurasi paling baik, yaitu sebesar 96%. Domain fitur yang memiliki pengaruh paling besar terhadap kondisi pasien diabetes adalah riwayat medis, sedangkan domain fitur demografi memiliki tingkat pengaruh paling rendah. Aplikasi yang dibuat dapat membantu tenaga medis dalam melakukan deteksi dan memberikan rekomendasi kepada pasien terkait fitur apa saja yang mempengaruhi kondisi pasien.

Pada penelitian lebih lanjut potensi pemanfaatan fitur demografi dapat ditingkatkan dengan penambahan jumlah fitur, hal ini dikarenakan jumlah fitur pada domain demografi lebih sedikit dibanding domain fitur riwayat medis dan gaya hidup.

DAFTAR PUSTAKA

- [1] Erlin, Y. N. Marlim, Junadhi, L. Suryati, and N. Agustina, "Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression", *Jurnal Nasional Teknik Elektro dan Teknologi Informasi*, vol. 11, no. 2, 2022, doi: 10.22146/jnteti.v11i2.3586.
- [2] R. A. Abdoul Raheem and A. K. M. Al-Qurabat, "Developing a Predictive Health Care System for Diabetes Diagnosis as a Machine Learning-Based Web Service", *Journal Of University Of Babylon for Pure and Applied Sciences*, 2022, doi: 10.29196/jubpas.v30i1.4061.
- [3] K. Fujihara and H. Sone, "Machine Learning Approach to Drug Treatment Strategy for Diabetes Care", 2023. doi: 10.4093/dmj.2022.0349.
- [4] C. Duckworth *et al.*, "Explainable Machine Learning for Real-Time Hypoglycemia and Hyperglycemia Prediction and Personalized Control Recommendations", *J Diabetes Sci Technol*, vol. 18, no. 1, 2024, doi: 10.1177/19322968221103561.
- [5] M. R. Islam, S. Banik, K. N. Rahman, and M. M. Rahman, "A comparative approach to alleviating the prevalence of diabetes mellitus using machine learning", *Computer Methods and Programs in Biomedicine Update*, vol. 4, Jan. 2023, doi: 10.1016/j.cmpbup.2023.100113.
- [6] Y. Xiong *et al.*, "A personalized prediction model for urinary tract infections in type 2 diabetes mellitus using machine learning", *Front Pharmacol*, vol. 14, 2023, doi: 10.3389/fphar.2023.1259596.
- [7] J. Zhang *et al.*, "Machine learning for post-acute pancreatitis diabetes mellitus prediction and personalized treatment recommendations", *Sci Rep*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-31947-4.
- [8] K. Vivek Balaji and R. Sugumar, "Harnessing the Power of Machine Learning for Diabetes Risk Assessment: A Promising Approach", in *2023 International Conference on Data Science, Agents and Artificial Intelligence, ICDSAAI 2023*, 2023. doi: 10.1109/ICDSAAI59313.2023.10452476.
- [9] T. A. Fortuna, H. Karuniawati, D. Purnamasari, and D. E. Purlinda, "Faktor-Faktor yang Mempengaruhi Komplikasi pada Pasien Diabetes Mellitus di RSUD Dr. Moewardi", 2023. [Online]. Available: <http://journals.ums.ac.id/index.php/pharmacon>
- [10] M. Yusa, F. F. Coastera, and M. R. Yandika, "Reduksi Dimensi Data menggunakan Metode Wrapper Sequential Feature Selection untuk Peningkatan Performa Algoritma Naïve Bayes terhadap Dataset Medis", *Jurnal Edukasi dan Penelitian Informatika (JEPIN)*, vol. 8, no. 2, 2022, doi: 10.26418/jp.v8i2.54328.

- [11] A. Rahmadeyan and M. Mustakim, "Seleksi Fitur pada Supervised Learning: Klasifikasi Prestasi Belajar Mahasiswa Saat dan Pasca Pandemi COVID-19", *Jurnal Nasional Teknologi dan Sistem Informasi*, vol. 9, no. 1, 2023, doi: 10.25077/teknosi.v9i1.2023.21-32.
- [12] A. García-Domínguez *et al.*, "Diabetes Detection Models in Mexican Patients by Combining Machine Learning Algorithms and Feature Selection Techniques for Clinical and Paraclinical Attributes: A Comparative Evaluation", *J Diabetes Res*, vol. 2023, doi: 10.1155/2023/9713905.
- [13] S. Di, J. Petch, H. C. Gerstein, R. Zhu, and D. Sherifali, "Optimizing Health Coaching for Patients With Type 2 Diabetes Using Machine Learning: Model Development and Validation Study", *JMIR Form Res*, vol. 6, no. 9, 2022, doi: 10.2196/37838.
- [14] S. Park, J. H. Kim, Y. K. Cha, M. J. Chung, J. H. Woo, and S. Park, "Application of Machine Learning Algorithm in Predicting Axillary Lymph Node Metastasis from Breast Cancer on Preoperative Chest CT", *Diagnostics*, vol. 13, no. 18, 2023, doi: 10.3390/diagnostics13182953.
- [15] E. Fauziah, "Pemanfaatan Program Pengelolaan Penyakit Kronis", *HIGEIA (Journal of Public Health Research and Development)*, vol. 4, no. 4, 2020.
- [16] J. Ha, M. Kambe, and J. Pe, *Data Mining: Concepts and Techniques*. 2011. doi: 10.1016/C2009-0-61819-5.
- [17] M. Glučina, A. Lorencin, N. Anđelić, and I. Lorencin, "Cervical Cancer Diagnostics Using Machine Learning Algorithms and Class Balancing Techniques", *Applied Sciences (Switzerland)*, vol. 13, no. 2, 2023, doi: 10.3390/app13021061.
- [18] W. Nugraha, "Prediksi Penyakit Jantung Cardiovascular Menggunakan Model Algoritma Klasifikasi", *Jurnal Manajemen dan Informatika*, vol. 9, no. 2, 2021.
- [19] M. Nagassou, R. W. Mwangi, and E. Nyarige, "A Hybrid Ensemble Learning Approach Utilizing Light Gradient Boosting Machine and Category Boosting Model for Lifestyle-Based Prediction of Type-II Diabetes Mellitus", *Journal of Data Analysis and Information Processing*, vol. 11, no. 04, pp. 480–511, 2023, doi: 10.4236/jdaip.2023.114025.
- [20] Y. Wang, R. Hou, B. Ni, Y. Jiang, and Y. Zhang, "Development and validation of a prediction model based on machine learning algorithms for predicting the risk of heart failure in middle-aged and older US people with prediabetes or diabetes", *Clin Cardiol*, vol. 46, no. 10, 2023, doi: 10.1002/clc.24104.
- [21] Y. M. Brekmans Darkel, L. Ermilinda, G. Kurniawan Al Yulianto, and C. Fransiska Pacolinus, "Implementasi Model Algoritma C4.5 Untuk Klasifikasi Status Stunting Di Kabupaten Sikka", *Jurnal Techno.Com*, vol. 23, no. 04, 2024, doi: <https://doi.org/10.62411/tc.v23i4.11663>
- [22] F. Zhapa-Camacho, M. Kulmanov, and R. Hoehndorf, "mOWL: Python library for machine learning with biomedical ontologies", *Bioinformatics*, vol. 39, no. 1, 2023, doi: 10.1093/bioinformatics/btac811.
- [23] N. Baharun, N. F. M. Razi, S. Masrom, N. A. M. Yusri, and A. S. A. Rahman, "Auto Modellingfor Machine Learning: A Comparison Implementation between Rapid Miner and Python", *International Journal of Emerging Technology and Advanced Engineering*, vol. 12, no. 5, pp. 15–27, May 2022, doi: 10.46338/ijetae0522_03.