Evaluasi Metode SHAP dan LIME untuk Menganalisis Faktor Risiko Diabetes

Raihan Fhadilah[1], Jondri[2], Indwiarti[3]

Fakultas Informatika, Universitas Telkom, Bandung
[1]mrrxxnzz@students.telkomuniversity.ac.id, [2]jondri@telkomuniversity.ac.id,
[3]indwiarti@telkomuniversity.ac.id,

Abstrak

Diabetes adalah penyakit yang angka penderitanya terus meningkat setiap tahunnya, menjadikannya salah satu masalah kesehatan utama yang dihadapi dunia. Meskipun ada berbagai metode untuk mendeteksi diabetes, prediksi berbasis kecerdasan buatan semakin populer dalam mendiagnosis penyakit diabetes dengan tingkat akurasi yang tinggi. Namun, model-model *AI* sering kali bersifat *black-box*, sehingga sulit untuk menginterpretasikan faktor-faktor yang memengaruhi hasil prediksi. Penelitian ini bertujuan untuk membandingkan metode *Explainable AI* (*XAI*), yaitu metode SHAP dan LIME, dalam memberikan penjelasan terhadap hasil dari prediksi model. Penelitian ini menggunakan *dataset* diabetes yang mencakup fitur-fitur seperti kehamilan, kadar glukosa, tekanan darah, ketebalan kulit, insulin, indeks massa tubuh, keturunan, usia, dan *output* (kelas diabetes dan kelas non-diabetes). Dengan menggunakan metode SHAP dan LIME, penelitian ini memuat hasil berupa penjelasan keputusan dari prediksi yang dibuat oleh model *XGBoost*. Hasil menunjukkan bahwa SHAP memberikan interpretasi yang lebih stabil, konsisten, dan dapat dipercaya dibandingkan LIME, serta lebih direkomendasikan untuk digunakan dalam mendukung pengambilan keputusan medis terkait diagnosis diabetes.

Kata kunci: Explainable AI, SHAP, LIME, XGBoost, Diabetes, Black-Box.

1. Pendahuluan

Latar Belakang

Diabetes terjadi ketika kadar glukosa berlebihan, misalnya karena mengonsumsi makanan atau minuman yang memiliki kadar gula tinggi. *Glukosa* merupakan salah satu sumber energi utama yang dibutuhkan tubuh. Tubuh secara alami membentuk glukosa dari makanan maupun minuman yang dikonsumsi setiap harinya. Sebelum glukosa disalurkan menjadi energi yang dibutuhkan tubuh, glukosa akan masuk ke dalam sel dengan bantuan hormon *insulin* yang diproduksi oleh pankreas. Insulin membantu glukosa masuk ke dalam sel tubuh. Pengidap diabetes memiliki kelainan pada hormon insulin yang tidak mencukupi untuk mengalirkan glukosa ke dalam sel tubuh, atau pankreas sama sekali tidak memproduksi hormon insulin. Akibatnya, glukosa tidak dapat diserap oleh sel-sel tubuh melainkan akan tetap berada dalam darah [2].

Secara umum, diabetes dibagi menjadi beberapa tipe: diabetes tipe 1, diabetes tipe 2, dan *diabetes gestasional*. Diabetes tipe 1 merupakan suatu gangguan yang disebabkan oleh kerusakan pada sel pankreas, yaitu sel penghasil insulin, sehingga insulin tidak dapat diproduksi sama sekali. Diabetes tipe 2 adalah penyakit yang terjadi ketika tubuh tidak dapat menggunakan insulin dengan baik atau tidak dapat menghasilkan insulin yang cukup untuk mengubah glukosa menjadi energi. Sementara itu, *diabetes gestasional* adalah diabetes yang terjadi selama masa kehamilan. Umumnya, penderita diabetes gestasional akan sembuh setelah melahirkan. Namun, diabetes gestasional dapat meningkatkan risiko terkena diabetes tipe 2 di masa depan [2].

Penyebab diabetes sangat beragam, mulai dari faktor genetik, gaya hidup, hormon, gangguan *autoimun*, hingga faktor lainnya. Faktor genetik memainkan peran utama dalam diabetes. Individu dengan riwayat keluarga yang memiliki penyakit diabetes memiliki kemungkinan lebih tinggi untuk mengalami penyakit ini [2,3]. Selain itu, gaya hidup juga memengaruhi risiko diabetes. Pola makan tidak sehat, kurangnya aktivitas fisik atau olahraga, serta obesitas merupakan penyebab umum diabetes tipe 2. Pada diabetes gestasional, perubahan hormon seperti hormon *human placental lactogen* (HPL) dan hormon plasenta meningkatkan *resistensi insulin* untuk memastikan glukosa masuk ke janin ibu hamil. Di sisi lain, gangguan autoimun dapat menyebabkan diabetes tipe 1. Dalam kondisi ini, sistem kekebalan tubuh menyerang sel-sel pankreas yang memproduksi insulin [2].

Diabetes memiliki pengaruh besar terhadap kesehatan, baik dalam jangka pendek maupun panjang. Dalam banyak kasus, diabetes dapat merusak saraf tubuh yang kemudian memengaruhi fungsi ginjal, penglihatan, dan organ lainnya. Diabetes juga berdampak pada kualitas hidup individu, misalnya

menyebabkan kelelahan dan rasa haus berlebihan. Dalam jangka panjang, diabetes dapat menyebabkan kerusakan organ tubuh, mulai dari saraf hingga organ vital lainnya, yang dapat memperpendek usia penderita [2].

Seiring berkembangnya teknologi, kecerdasan buatan atau *Artificial Intelligence (AI)* kini memiliki peran penting dalam membantu diagnosis berbagai penyakit, termasuk diabetes. Penerapan teknologi ini memiliki tantangan tersendiri, salah satunya adalah konsep *black-box* atau kotak hitam. Konsep ini mengacu pada algoritma atau mekanisme dalam model AI yang sulit dijelaskan kepada pengguna, sehingga sulit untuk memahami cara kerja model dalam menghasilkan prediksi [1,10].

Black box merupakan suatu model yang sulit untuk dijelaskan, salah satunya adalah *XGBoost*. Model *XGBoost* merupakan singkatan dari *Extreme Gradient Boosting*, adalah algoritma *pembelajaran mesin* berbasis *boosting* yang sering kali digunakan karena kemampuannya untuk menangani data tabular dengan sangat baik dan menghasilkan performa yang unggul dalam berbagai kompetisi pembelajaran mesin [6]. Namun, sifat kompleks dari model ini, yang menggabungkan banyak pohon keputusan dalam proses iteratif untuk mengoptimalkan fungsi objektif, membuatnya sulit untuk diinterpretasikan secara langsung oleh manusia.

Model *XGBoost* menggunakan kombinasi antara *regularisasi*, *subsampling*, dan teknik penguatan untuk meningkatkan akurasi prediksi, tetapi hal ini juga meningkatkan sifat *black-box*-nya. Struktur model yang terdiri dari ribuan aturan keputusan pada setiap tahap prediksi menyebabkan sulitnya memahami bagaimana suatu prediksi akhir dibuat. Oleh karena itu, diperlukan metode *Explainable AI (XAI)*, seperti SHAP dan LIME, untuk memberikan wawasan tentang kontribusi setiap fitur terhadap hasil prediksi yang dihasilkan oleh model [6].

Untuk mengatasi masalah tersebut, dikembangkan konsep *Explainable AI* (XAI) dengan berbagai metode, seperti *Local Interpretable Model-agnostic Explanations* (LIME) dan *SHapley Additive exPlanations* (SHAP). Keduanya menjadi metode yang banyak digunakan untuk menjelaskan hasil prediksi dari model AI [1]. LIME bekerja dengan menciptakan model lokal yang sederhana untuk menjelaskan fitur-fitur tertentu yang berpengaruh terhadap hasil prediksi. Metode ini memungkinkan pengguna untuk memahami hubungan antara fitur data, seperti kadar glukosa atau indeks massa tubuh, dengan hasil diagnosis [10]. Sementara itu, SHAP menggunakan pendekatan berbasis teori nilai *Shapley* untuk menghitung kontribusi setiap fitur terhadap hasil prediksi, baik secara individual maupun secara keseluruhan data [1,10].

Penerapan metode LIME dan SHAP dalam diagnosis diabetes memberikan dampak positif, terutama dalam menjelaskan faktor-faktor yang memengaruhi kondisi pasien. Dengan metode ini, tenaga medis dapat menjelaskan faktor kunci yang berkontribusi terhadap diabetes, sekaligus merancang langkah-langkah yang lebih akurat untuk setiap pasien. Selain itu, metode tersebut dapat membantu tenaga medis memverifikasi faktor-faktor yang perlu diperbaiki dalam pengelolaan diabetes [2].

Rumusan Masalah

1. Bagaimana metode SHAP dan LIME memberikan penjelasan terhadap prediksi model klasifikasi pada kasus deteksi diabetes?
2. Seberapa konsisten hasil interpretasi yang diberikan oleh SHAP dan LIME terhadap data individu yang berbeda?
3. Apa saja kelebihan dan kekurangan dari masing-masing metode (SHAP dan LIME) dalam memberikan interpretasi model?
4. Fitur-fitur apa yang paling berpengaruh dalam menentukan apakah seorang individu diklasifikasikan sebagai penderita diabetes atau tidak menurut hasil interpretasi model?

Tujuan

1. Menganalisis bagaimana metode SHAP dan LIME memberikan penjelasan terhadap prediksi yang dihasilkan oleh model klasifikasi, khususnya dalam konteks deteksi diabetes.
2. Mengevaluasi tingkat konsistensi penjelasan yang dihasilkan oleh SHAP dan LIME terhadap individu yang diabetes maupun non-diabetes.
3. Mengidentifikasi kelebihan dan kekurangan masing-masing metode.
4. Menentukan fitur-fitur paling berpengaruh yang berkontribusi terhadap klasifikasi seorang individu sebagai penderita diabetes atau tidak.

2. Studi Terkait

1. SHAP (Shapley Additive Explanations)

SHAP *(Shapley Additive Explanations)* adalah metode yang digunakan untuk menjelaskan keputusan dari model, dengan dasar teori nilai Shapley dari teori permainan. *SHAP* memungkinkan kita untuk menghitung kontribusi setiap fitur terhadap hasil prediksi model, baik secara lokal maupun global, yang membantu meningkatkan transparansi dalam mode [10].

$$\emptyset i = \sum S \sqsubseteq \{1, \dots, p\}\{i\} \frac{|S|!(p-|S|-1)!}{p!} \times [Val(S \cup \{i\}) - Val(S)] \quad \textbf{(1)}$$

Untuk $\emptyset i$ merupakan suatu nilai shapley yang diperoleh dari suatu fitur pada hasil dari prediksi. $Val(S)$

merupakatn suatu luaran dari model suatu *Machine Learning* yang akan dijelaskan menggunakan fitur dari dataset S, dan p merupakan total dari keseluruhan dari fitur dataset [14].

Proses pengambilan nilai SHAP dimulai menentukan kontribusi masing-masing fitur terhadap hasil prediksi dari model. Untuk mengetahui seberapa besar kontribusi sebuah fitur, SHAP membandingkan prediksi model dalam berbagai kondisi saat fitur tersebut disertakan maupun tidak disertakan dalam proses prediksi [9].

SHAP akan mencoba mengamati apa yang terjadi terhadap prediksi model jika sebuah fitur diaktifkan atau di non-aktifkan, lalu membandingkan perbedaannya. Namun, SHAP akan melakukan perbandingan tidak dilakukan dalam satu skenario saja, melainkan pada semua kemungkinan kombinasi fitur yang dapat terjadi. SHAP kemudian menghitung rata-rata kontribusi fitur tersebut di seluruh kombinasi tersebut [9].

Setelah nilai kontribusi dihitung, SHAP akan menyusun penjelasan akhir dengan menjumlahkan kontribusi dari setiap fitur. Hasilnya adalah sebuah gambaran yang menunjukkan seberapa besar setiap fitur menaikkan atau menurunkan hasil prediksi. Nilai-nilai ini memiliki makna semakin besar nilai SHAP dari suatu fitur, semakin besar pula pengaruh fitur tersebut terhadap prediksi [9].

2. LIME (Local Interpertable Model Agnostic Explanation)

LIME *(Local Interpertable Model Agnostic Explanation)* merupakan suatu metode yang mempelajari model yaang dijelaskan secara lokal di sekitar prediksi oleh model, dan merupakan metode yang dapat menjelaskan suatu model yang sudah dibentuk dari model tertentu[11]. Metode ini menjelaskan suatu model dengan cara memodifikasi data yang masuk dan mengamati hasil dari suatu model untuk memahami bagaimana proses prediksi berubah ketika dilakukan perubahan pada data yang masuk pada saat proses selanjutnya. Jika terdapat suatu perubahan pada suatu model ketika dilakukan perubahan pada nilai data, maka LIME akan menggap fitur yang diubah tersebut memiliki nilai [16].

$$\xi(x) = argminL(f, g, \prod_x) + \Omega(g), g \in G \qquad \textbf{(2)}$$

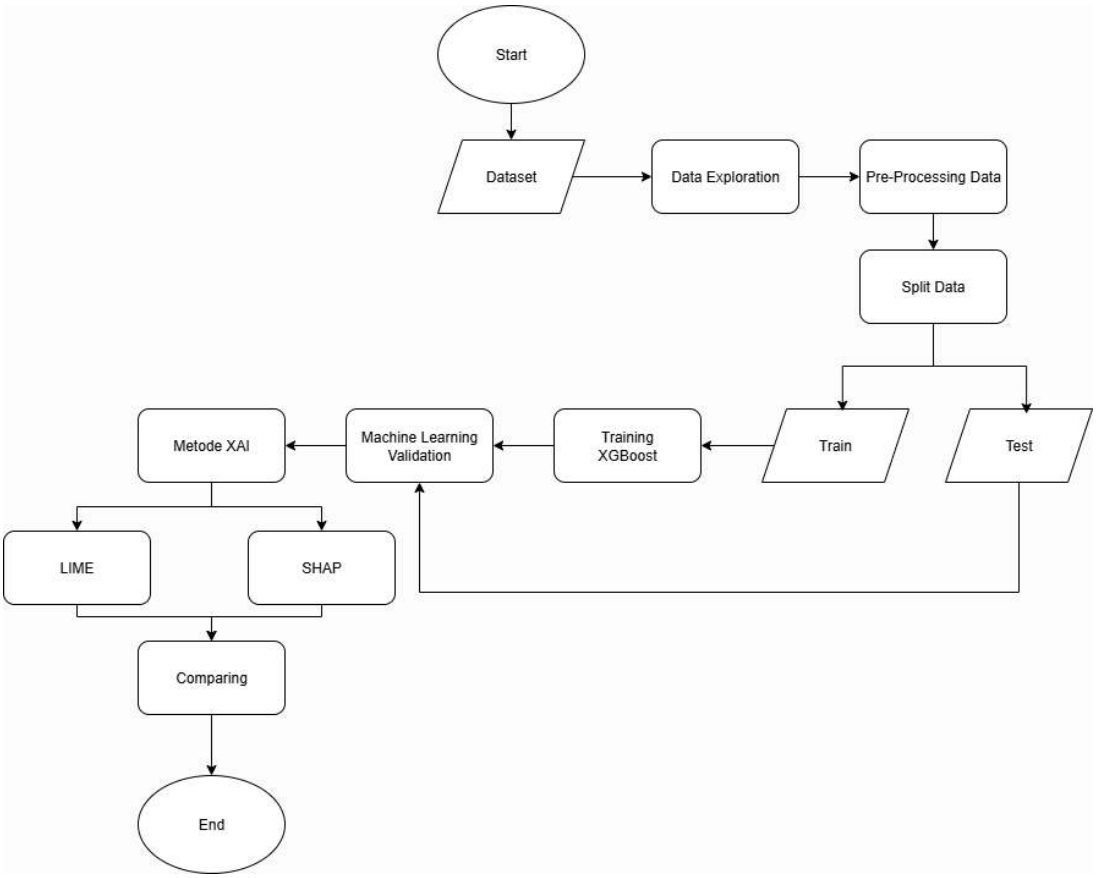
Variable G merupakan suatu himpunan model yang dapat dijelaskan seperti, *Decision trees* atau *model linear* lainnya. Untuk setiap model didalam G dilambangkan dengan g. Kompleksitas dari model g dinyatakan sebagai $\Omega(g)$ yang merupakan jumlah dari pohon keputusan tersebut. Untuk f melambangkan model *black-box* itu sendiri. Akurasi diantara G dan f dinyakatakn oleh fungsi L, yang akan dievalusi dalam nilai lokal $\prod_x$ yang merupakan nilai di sekitar titik x. Metode LIME sendiri bertujuan untuk meminimalkan nilai dari fungsi L dan $\Omega(g)$ [16].

Langkah pertama dimulai dengan memilih satu *instance* (data masuk) yang prediksinya ingin dijelaskan. LIME kemudian menciptakan data baru (*Random Samples*) melalui *perturbasi* yaitu mengubah fitur-fitur dari *instance* tersebut untuk menghasilkan variasi input. Setiap data hasil *perturbasi* dimasukkan ke model asli untuk memperoleh *output*. LIME lalu menghitung jarak antara data hasil *perturbasi* dengan *instance* asli dan memberikan *bobot* yang lebih besar pada data yang lebih mirip [9].

Selanjutnya, LIME melatih model sederhana seperti *regresi linear* menggunakan data yang telah diperturbasi dan dibobot. Model ini berfungsi sebagai *surrogate model* yang meniru perilaku model asli secara lokal. Terakhir, *output* dari model lokal yang digunakan sebagai penjelasan, menunjukkan kontribusi masing-masing fitur terhadap prediksi. Dengan demikian, LIME memberikan penjelasan yang mudah dipahami tanpa harus mengakses struktur dalam model kompleks [9].

3. Sistem yang Dibangun

1. FlowChart



Gambar 1. FlowChart

Flowchart penelitian menggambarkan alur kerja dalam membandingkan dua metode interpretasi *XAI* (Explainable Artificial Intelligence), yaitu *LIME* dan *SHAP*, pada klasifikasi diabetes menggunakan algoritma *XGBoost*. Proses dimulai dari pengumpulan dan eksplorasi dataset yang berisi fitur klinis pasien, dilanjutkan dengan praproses berupa pembersihan, normalisasi, serta pembagian data menjadi *train* dan *test*.

Model kemudian dibangun menggunakan *XGBoost* dan divalidasi untuk mengevaluasi performanya. Selanjutnya, hasil prediksi dijelaskan dengan *LIME* (berbasis data lokal) dan *SHAP* (berbasis kontribusi fitur global). Kedua metode dibandingkan guna menilai kelebihan dan kekurangannya, dengan tujuan menentukan pendekatan interpretasi yang paling sesuai untuk konteks medis, khususnya analisis diabetes.

2. Data Eksplorasi

Pada tahap eksplorasi data (*Exploratory Data Analysis*), langkah awal yang dilakukan adalah mengidentifikasi tipe data dari setiap fitur yang terdapat dalam *dataset* diabetes. Adapun tipe data dari masing-masing fitur dalam *dataset* tersebut adalah sebagai berikut:

Table 1. Tipe Data

Attribute	Tipe Data
Pregnancies	Integer
Glucose	Integer
BloodPressure	Integer
SkinThickness	Integer
Insulin	Integer
BMI	Float
DiabetesPedigreeFunction	Float
Age	Integer

Setelah dilakukan pengecekan terhadap tipe data, tahap selanjutnya adalah melakukan analisis statistik

deskriptif pada *dataset*. Berikut merupakan ringkasan hasil statistik deskriptif :

Table 2. Statistik Deskriptif

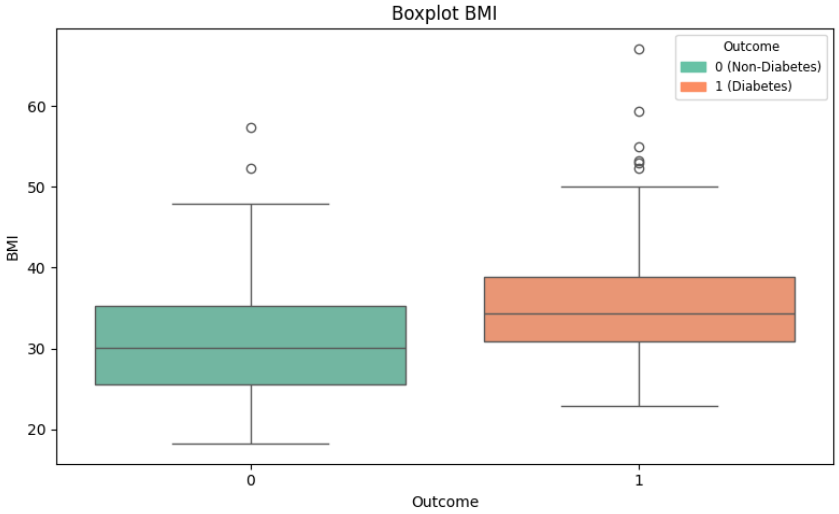
	Pregnancies	Glucose	Blood Pressure	Skin Thickness	Insulin	BMI	Diabetes Pedigree Function	Age	Outcome
Count	768	768	768	768	768	768	768	768	768
Mean	3.85	120.9	69.11	20.54	79.7995	31.99	0.471876	33.2	0.35
Std	3.37	31.97	19.36	15.95	115.244	7.884	0.331329	11.8	0.48
Min	0	0	0	0	0	0	0.078	21	0
25%	1	99	62	0	0	27.3	0.24375	24	0
50%	3	117	72	23	30.5	32	0.3725	29	0
75%	6	140.3	80	32	127.25	36.6	0.62625	41	1
Max	17	199	122	99	846	67.1	2.42	81	1

Karena Pada dasarnya nila dari *Glucose, BMI, Insulin, BloodPressure*, dan *SkinThickness* tidak mungkin terdapat nilai 0 maka akan dilakukan pengecekan kembali jumlah nilai 0 yang terdapat pada attribut tersebut :

Table 3. Jumlah Nilai 0

Attribut	Jumlah
Glucose	5
BMI	11
Insulin	374
BloodPressure	35
SkinThickness	227

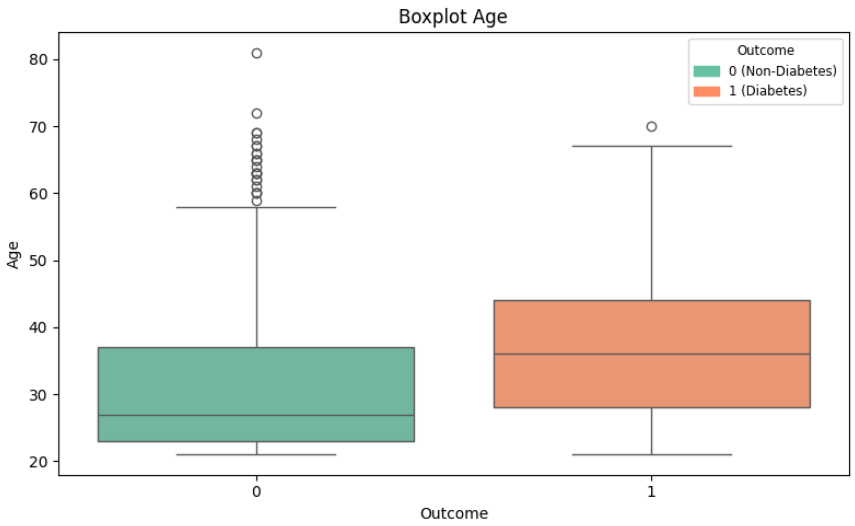
Dari banyaknya jumlah nilai 0 yang terdapat pada atribut *Glucose*, *BMI*, *Insulin*, *BloodPressure*, dan *SkinThickness*, maka dilakukan penggantian nilai 0 tersebut menjadi *null* atau dianggap tidak memiliki nilai. Penggantian ini dilakukan untuk mencegah gangguan dalam proses eksplorasi data lebih lanjut yang nantinya dapat memengaruhi tahap pengambilan kesimpulan.



Gambar 2. Boxplot BMI

Kelompok diabetik (*Outcome* = 1) memiliki nilai *BMI* yang cenderung lebih tinggi dibandingkan dengan kelompok non-diabetik. Hal ini ditunjukkan oleh nilai median *BMI* kelompok diabetik yang berada di sekitar 34, sedangkan kelompok non-diabetik memiliki median sekitar 30. Selain itu, kelompok diabetik menunjukkan jumlah *outlier* yang lebih banyak, terutama pada nilai *BMI* ekstrem di atas 50, bahkan mendekati 70. Kondisi ini mengindikasikan adanya potensi risiko komplikasi serius pada individu dengan diabetes.

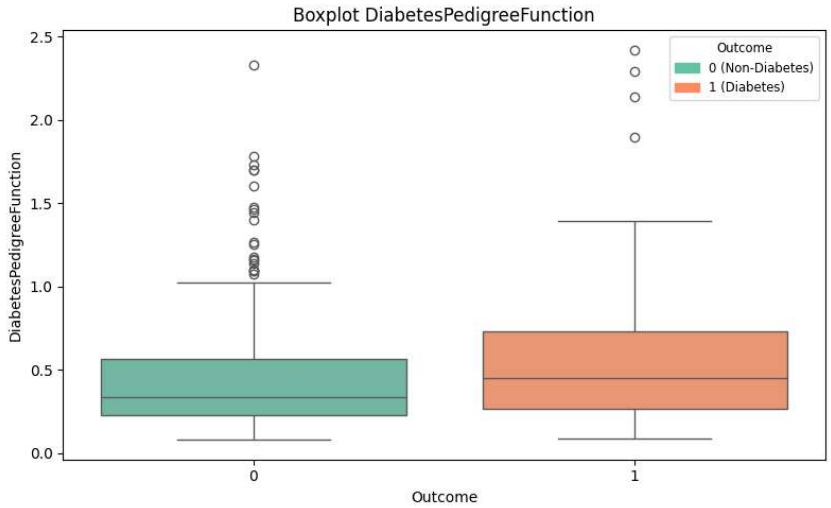
Sebaran data pada kelompok diabetik juga tampak lebih lebar, yang mencerminkan variasi berat badan yang lebih besar di antara individu dalam kelompok tersebut. Temuan ini menunjukkan bahwa individu dengan berat badan berlebih atau indeks massa tubuh (*Body Mass Index*/*BMI*) yang tinggi memiliki risiko yang lebih besar untuk menderita diabetes.



Gambar 3. Boxplot Age

Kelompok non-diabetik (*Outcome* = 0) umumnya terdiri dari individu yang berusia lebih muda, dengan kuartil pertama (*Q1*) berada pada awal usia 20-an, nilai median sekitar 27 tahun, dan kuartil ketiga (*Q3*) di pertengahan usia 30-an. Meskipun terdapat beberapa *outlier* hingga usia mendekati 80 tahun, jumlahnya relatif sedikit.

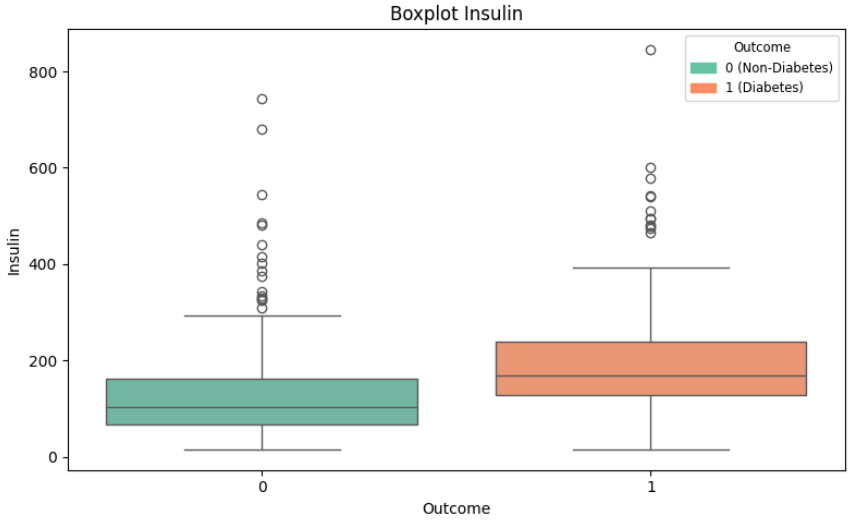
Sebaliknya, kelompok diabetik (*Outcome* = 1) memiliki nilai median usia yang lebih tinggi, yaitu berada pada pertengahan hingga akhir usia 30-an, serta menunjukkan sebaran yang lebih luas. Jumlah *outlier* pada kelompok ini lebih sedikit, meskipun tetap terdapat individu dengan usia hingga 70-an. Temuan ini mengindikasikan bahwa dalam *dataset* yang dianalisis, risiko diabetes cenderung meningkat seiring dengan bertambahnya usia.



Gambar 4. Boxplot DiabetesPedigreeFunction

Pada kelompok non-diabetik (*Outcome* = 0), nilai *Diabetes Pedigree Function* umumnya berada pada rentang antara 0,2 hingga 0,6, dengan nilai median sekitar 0,35. Meskipun terdapat beberapa *outlier* dengan nilai di atas 2,0, sebagian besar nilai tetap berada di bawah 1,0, yang mengindikasikan riwayat keluarga terhadap diabetes yang relatif lemah.

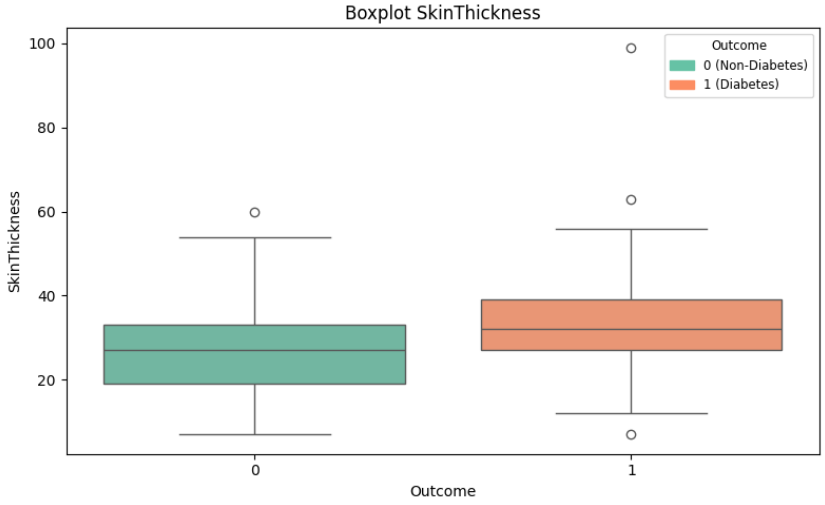
Sementara itu, pada kelompok diabetik (*Outcome* = 1), distribusi *Diabetes Pedigree Function* bergeser ke arah kanan dengan nilai median yang lebih tinggi, yakni di kisaran 0,45, serta rentang interkuartil yang lebih lebar. Beberapa *outlier* bahkan tercatat melebihi angka 2,0. Temuan ini menunjukkan bahwa semakin kuat riwayat diabetes dalam keluarga, semakin tinggi pula risiko individu tersebut untuk menderita diabetes.



Gambar 5. Boxplot Insulin

Kelompok non-diabetik (*Outcome* = 0) memiliki median *insulin* sekitar 100, lebih rendah dibandingkan kelompok diabetik (*Outcome* = 1) dengan median mendekati 180. Sebaran pada kelompok diabetik lebih lebar dengan banyak *outlier*, termasuk nilai ekstrem di atas 800, sedangkan kelompok non-diabetik cenderung lebih rapat dan simetris.

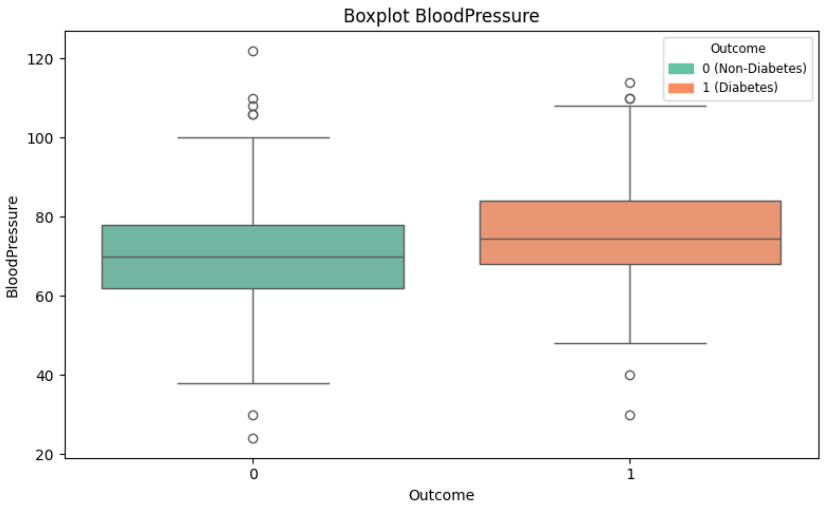
Perbedaan ini mengindikasikan adanya hubungan antara tingginya kadar *insulin* dengan kejadian diabetes. Namun, keberadaan nilai *insulin* nol atau mendekati nol perlu dicermati karena berpotensi mencerminkan data hilang atau tidak tercatat.



Gambar 6. Boxplot SkinThickness

Kelompok diabetik (*Outcome* = 1) memiliki median *skin thickness* lebih tinggi (±32) dibandingkan kelompok non-diabetik (*Outcome* = 0) yang berada di kisaran 27. Distribusi pada kelompok diabetik juga menunjukkan lebih banyak *outlier* ekstrem, termasuk nilai mendekati 100.

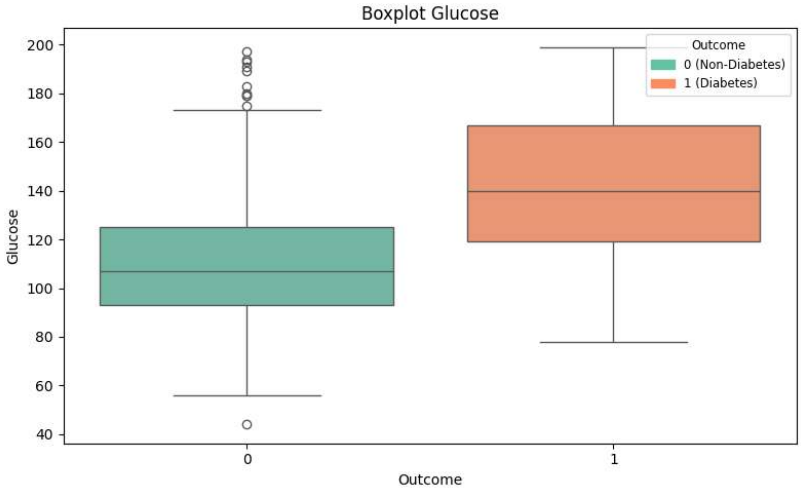
Sebaliknya, kelompok non-diabetik memiliki distribusi lebih stabil dengan rentang sempit. Temuan ini mengindikasikan bahwa ketebalan lipatan kulit dapat menjadi indikator potensial risiko diabetes.



Gambar 7. Boxplot BloodPressure

Kelompok diabetik (*Outcome* = 1) memiliki median *blood pressure* sekitar 75, sedikit lebih tinggi dibandingkan kelompok non-diabetik (*Outcome* = 0) dengan median 70. Sebaran tekanan darah pada kelompok non-diabetik lebih luas dan mengandung lebih banyak *outlier* ekstrem (di bawah 40 maupun di atas 120).

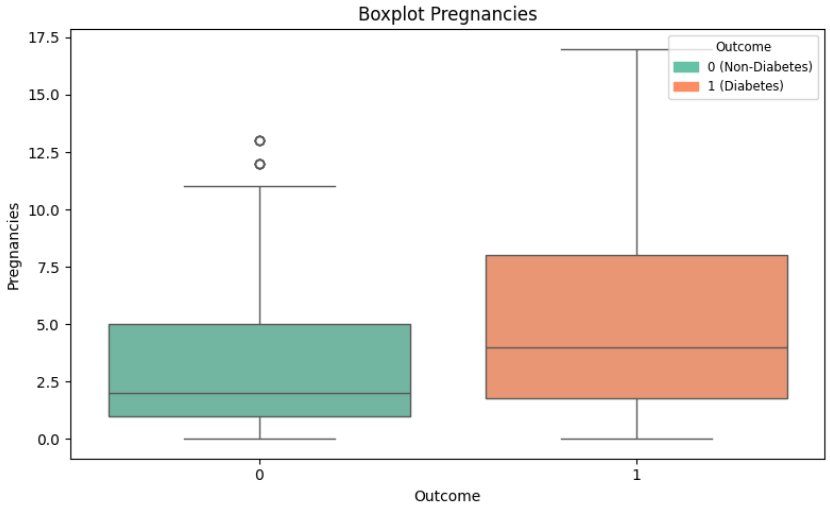
Sebaliknya, distribusi pada kelompok diabetik lebih rapat dengan *outlier* lebih sedikit. Hal ini menunjukkan bahwa meskipun tekanan darah cenderung lebih tinggi pada penderita diabetes, variabilitas ekstrem justru lebih sering muncul pada kelompok non-diabetik.



Gambar 8. Boxplot Glucose

Kelompok diabetik (*Outcome* = 1) memiliki median *glucose* sekitar 140, jauh lebih tinggi dibandingkan kelompok non-diabetik (*Outcome* = 0) dengan median ±105. Sebaran *glucose* pada kelompok diabetik juga lebih tinggi (78–200), sedangkan kelompok non-diabetik berada pada rentang lebih rendah dengan beberapa *outlier* ekstrem.

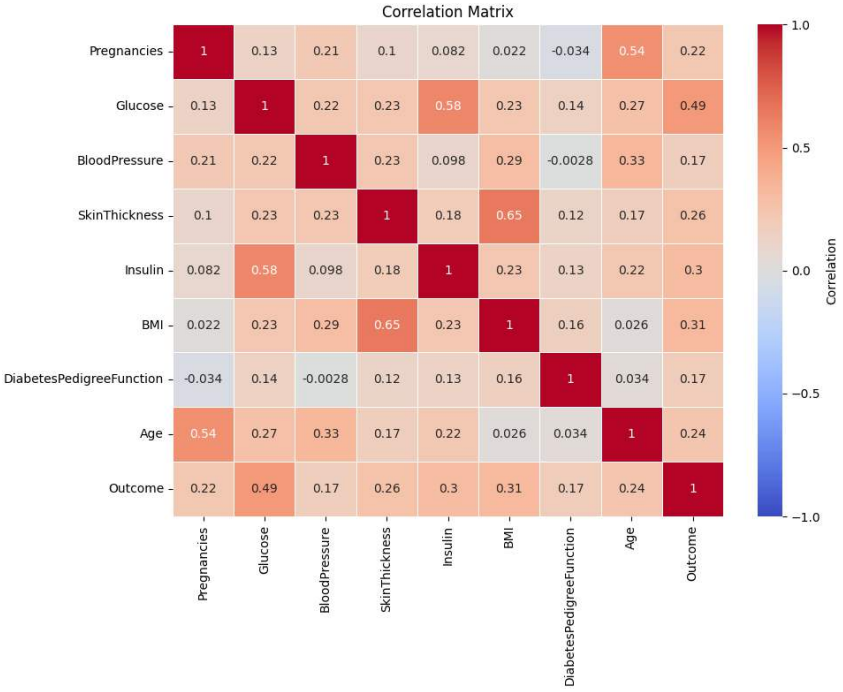
Perbedaan ini menegaskan bahwa *glucose* merupakan indikator utama dalam membedakan individu dengan dan tanpa diabetes, serta menjadi variabel kunci dalam proses klasifikasi.



Gambar 9. Boxplot Pregnancies

Kelompok diabetik (*Outcome* = 1) memiliki median jumlah *pregnancies* lebih tinggi (±4) dibandingkan kelompok non-diabetik (*Outcome* = 0) dengan median sekitar 2. Sebaran pada kelompok diabetik lebih luas, dengan rentang interkuartil 2–8, sedangkan kelompok non-diabetik lebih sempit (1–5).

Meskipun terdapat *outlier* ekstrem (hingga 17 pada diabetik dan 13 pada non-diabetik), distribusi data menunjukkan bahwa individu dengan diabetes cenderung memiliki jumlah kehamilan lebih banyak dibandingkan non-diabetik.



Gambar 10. Heatmap Correlation Matrix

Hasil *heatmap* matriks korelasi menunjukkan bahwa *glucose* memiliki hubungan terkuat dengan *outcome* 0,49, menegaskan perannya sebagai indikator utama diabetes. Variabel *BMI* 0,31 dan *insulin* 0,30 juga berkorelasi cukup signifikan, sementara *skin thickness*, *age*, dan *pregnancies* menunjukkan korelasi lebih rendah namun tetap relevan. Sebaliknya, *blood pressure* dan *diabetes pedigree function* memiliki korelasi lemah sehingga kontribusinya terhadap klasifikasi relatif kecil.

3. Pre-processing

Pada tahap *pre-processing*, dilakukan proses pembersihan data guna memastikan kualitas data yang digunakan dalam pelatihan model berada dalam kondisi optimal. Salah satu langkah penting dalam tahap ini adalah pengecekan terhadap keberadaan nilai *null* atau kosong. Nilai *null* dapat menimbulkan gangguan dalam proses pelatihan model, karena mencerminkan adanya informasi yang hilang dan berpotensi memengaruhi

akurasi serta kestabilan model yang dibangun.

Table 4. Jumlah nilai null

Attribute	Jumlah
BMI	11
DiabetesPerigreeFunction	0
Age	0
Pregnancies	0
BloodPressure	35
SkinThickness	227
Insulin	374
Outcome	0

Terdapat beberapa Attribute yang memiliki nilai *null* seperti *BMI*, *BloodPressure*, *SkinThickness*, dan *Insulin*. Untuk mengatasi hal tersebut dilakukan penghapusan data yang memliki nilai *null* dikarenakan akan mengganggu pada proses pelatihan model.

4. Training Model XGBoost

Proses pelatihan model dilakukan menggunakan algoritma XGBoost (*Extreme Gradient Boosting*), yang dikenal memiliki performa tinggi dalam tugas klasifikasi. Dalam tahap ini, *dataset* dibagi menjadi dua bagian, yaitu 80% data digunakan sebagai data pelatihan (*training set*) dan 20% sisanya digunakan sebagai data pengujian (*test set*).

5. Machine Learning Validation

Tahap ini akan menampilkan performa dari model yang sudah dilatih dengan XGBoost. Performa akan diukur dengan menggunakan nilai *Recall*, *Accuracy*, *F1-Score*, dan juga *Precission*

Table 5. Machine Learning Validation

	Precission	Recall	F1-Score	Support
0	0,81	0,83	0,82	52
1	0,65	0,63	0,64	27
Accuracy			0,76	79
Macro Avg	0,73	0,73	0,73	79
Weighted Avg	0,76	0,76	0,76	79

Model menunjukkan akurasi keseluruhan sebesar 76%, yang berarti 76% dari seluruh prediksi sesuai dengan label sebenarnya. Untuk kelas non-diabetik (*0*), performa model tergolong baik dengan *precision* 0,81, *recall* 0,83, dan *F1-score* 0,82, menandakan bahwa model cukup andal dalam mengenali individu yang tidak menderita diabetes. Kemudian, untuk kelas diabetik (*1*), performa model menurun, dengan *precision* 0,65, *recall* 0,63, dan *F1-score* 0,64. Ini menunjukkan bahwa model masih kurang optimal dalam mendeteksi kasus diabetes, yang bisa disebabkan oleh ketidakseimbangan data atau kompleksitas pola pada kelompok ini. Nilai *macro average* dan *weighted average* masing-masing berada di angka 0,73–0,76, yang menunjukkan bahwa meskipun model condong lebih baik pada kelas mayoritas.

4. Evaluasi

1. Pengujian Performa LIME Sebelum Tuning

Pada tahap ini, metode LIME diterapkan menggunakan kode standar yang umum digunakan sesuai dengan dokumentasi resmi LIME. Terdapat perbedaan hasil dari beberapa kali eksekusi ditunjukkan pada tabel berikut:

Table 6. Hasil uji LIME sebelum Tuning

Hasil	BMI	Diabetes Pedigree Function	Glucose	Age	Pregnancies	Blood Pressure	Skin Thickness	Insulin
1	0,1	-0,04	0,03	-0,12	-0,01	0,08	0,12	-0,03
2	0,11	-0,04	0,04	-0,1	0,01	0,08	0,08	-0,03
3	0,08	-0,04	0,04	-0,11	-0,01	0,08	0,12	0,04
4	0,08	-0,05	0,05	-0,12	0,01	0,08	0,11	-0,04

Hasil dari LIME pada eksekusi 1 hingga 4 menunjukkan bahwa fitur *BMI* secara konsisten memberikan kontribusi positif terhadap prediksi diabetes. Namun, fitur-fitur lain seperti *Glucose*, *Age*, dan *Skin Thickness* memperlihatkan ketidakkonsistenan dalam kontribusinya, baik dari sisi arah (positif/negatif) maupun besarnya nilai. Sebagai contoh, *Pregnancies* dan *Diabetes Pedigree Function* memberikan kontribusi negatif di sebagian besar observasi, namun dengan nilai yang tidak stabil. Fitur seperti *Insulin* dan *BloodPressure* pun tampak memiliki peran yang berubah-ubah antar observasi, bahkan terkadang berlawanan arah.

2. Pengujian Performa LIME Sesudah Tuning num_samples

Perbedaan utama dengan hasil sebelumnya dan yang telah dimodifikasi terletak pada penambahan parameter *num_samples*. Pada pengujian awal, parameter *num_samples* tidak dituliskan secara langsung, sehingga menggunakan nilai *default*, yaitu 5.000. Nilai tersebut sangat kecil dan tidak cukup untuk menghasilkan hasil yang kuat terhadap model. Dengan menaikkan nilai num_samples hingga 1.000.000, jumlah sampel acak yang dihasilkan oleh LIME menjadi jauh lebih banyak. Berikut merupakan hasil dari beberapa kali eksekusi setelah dilakukan *tuning*:

Table 7. Hasil uji LIME setelah tuning num_samples

Hasil	BMI	Diabetes Pedigree Function	Glucose	Age	Pregnancies	Blood Pressure	Skin Thickness	Insulin
1	0,09	-0,05	0,03	-0,1	0	0,8	0,1	-0,03
2	0,09	-0,05	0,03	-0,1	0	0,8	0,1	-0,03
3	0,09	-0,05	0,03	-0,1	0	0,8	0,1	-0,03
4	0,09	-0,05	0,03	-0,1	0	0,8	0,1	-0,03

Setelah parameter *num_samples* ditingkatkan menjadi 1.000.000, interpretasi *LIME* pada eksekusi 1–4 menunjukkan konsistensi yang jauh lebih baik. Seluruh uji coba menghasilkan kontribusi fitur yang sama, menandakan model lokal lebih stabil.

3. Pengujian Performa Lokal metode SHAP

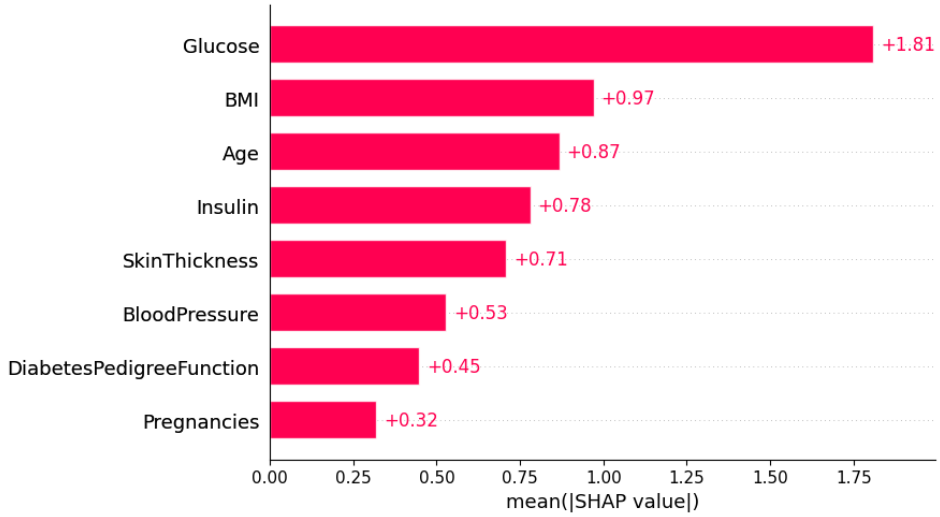
Uji coba SHAP dilakukan sebanyak 4 kali untuk melihat konsistensi dari metode tersebut. Setelah dilakukan uji coba sebanyak 4 kali, hasil dari uji coba tergolong konsiten yang memiliki hasil sebagai berikut:

Table 8. Hasil uji SHAP

Hasil	BMI	Diabetes Pedigree Function	Glucose	Age	Pregnancies	Blood Pressure	Skin Thickness	Insulin
1	0,22	-1,28	-2,17	-0,4	0,33	-0,72	0,74	0,69
2	0,22	-1,28	-2,17	-0,4	0,33	-0,72	0,74	0,69
3	0,22	-1,28	-2,17	-0,4	0,33	-0,72	0,74	0,69
4	0,22	-1,28	-2,17	-0,4	0,33	-0,72	0,74	0,69

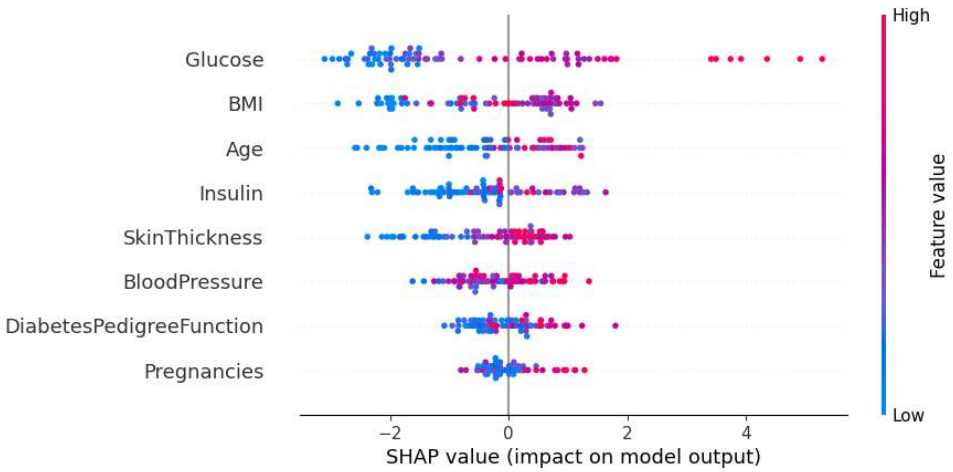
Hasil interpretasi penggunaan SHAP menunjukkan konsistensi antar eksekusi, dengan kontribusi fitur yang identik pada data 1 hingga 4. Hal ini mengindikasikan bahwa model membuat prediksi berdasarkan pola yang seragam terhadap fitur-fitur *input* yang diberikan.

4. Pengujian Global metode SHAP



Gambar 11. Kontribusi fitur metode SHAP

Selain penjelasan lokal, SHAP dapat menjelaskan model secara global terkait pengaruh atribut pada model. Hasil analisis SHAP memperkuat bahwa *glukosa* merupakan fitur paling dominan dalam mempengaruhi prediksi diabetes, dengan nilai SHAP rata-rata tertinggi (+1,81). Ini menunjukkan bahwa kadar *glukosa* tinggi secara konsisten meningkatkan kemungkinan seseorang diklasifikasikan sebagai diabetik oleh model. Disusul oleh *BMI* (+0,97) dan *usia* (+0,87), yang juga memiliki kontribusi signifikan terhadap prediksi. Keduanya dikenal sebagai faktor risiko utama dalam literatur medis. Fitur seperti *insulin*, ketebalan kulit (*SkinThickness*), dan tekanan darah memberikan pengaruh sedang, sementara riwayat keturunan (*Diabetes Pedigree Function*) dan jumlah kehamilan (*Pregnancies*) memiliki dampak paling rendah terhadap model.

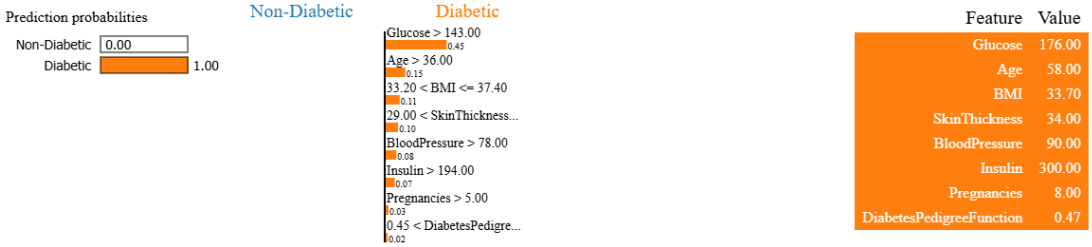


Gambar 12. Summaryplot metode SHAP

Hasil visualisasi SHAP menunjukkan bahwa *glukosa* merupakan fitur berpengaruh dalam memprediksi diabetes. Nilai SHAP yang tinggi pada kadar *glukosa* tinggi (warna merah) memberikan kontribusi positif signifikan terhadap prediksi diabetes. Hal ini menegaskan bahwa *glukosa* adalah indikator utama dalam diagnosis. Fitur *BMI* dan *usia* juga memberikan pengaruh besar. Nilai *BMI* dan *usia* yang tinggi cenderung meningkatkan prediksi diabetes, terlihat dari sebaran nilai SHAP merah di sisi positif.

Fitur lain seperti *insulin* dan *SkinThickness* memberikan pengaruh terhadap model, meskipun tidak sekuat *glukosa* dan *BMI*. Sementara itu, fitur seperti *BloodPressure*, *Diabetes Pedigree Function*, dan *Pregnancies* menunjukkan dampak yang lebih rendah, namun tetap memberikan kontribusi kecil terhadap prediksi.

5. Analisis Hasil Individu Diabetes LIME



Gambar 13. Hasil Individu Diabetes dengan metode LIME

Analisis *LIME* menunjukkan bahwa *glucose* merupakan fitur paling dominan dengan nilai 176, jauh melampaui ambang 143, sehingga kuat mendorong prediksi diabetes. Usia pasien (58) yang melebihi ambang 36, serta *BMI* 33,7 (kategori obesitas), juga berkontribusi signifikan.

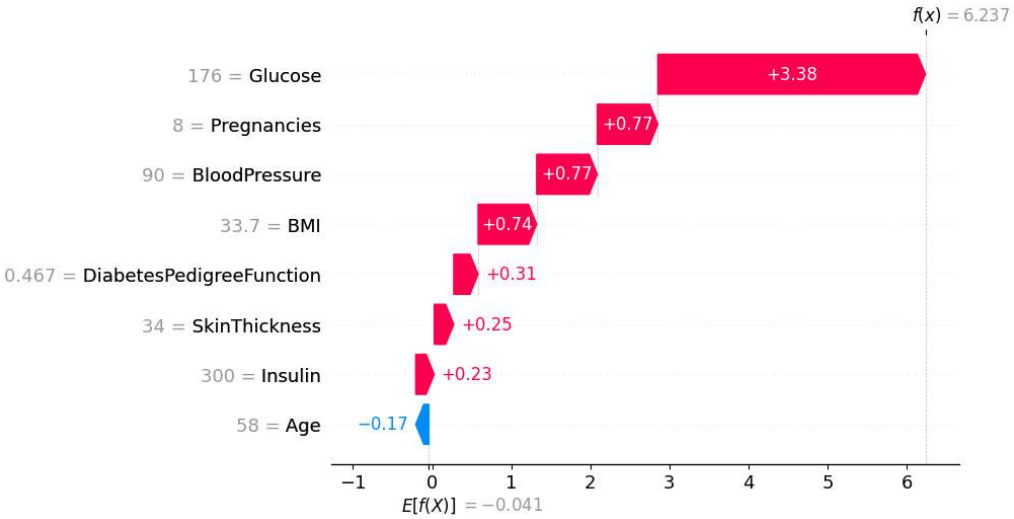
Fitur lain seperti *insulin* (300), *skin thickness* (34), dan *blood pressure* (90) memiliki pengaruh sedang karena berada di atas ambang model. Sementara itu, *pregnancies* (8) dan *diabetes pedigree function* (0,47) memberikan kontribusi paling rendah, meskipun tetap informatif dalam klasifikasi.

6. Analisis Hasil Individu Non-Diabetes LIME

Gambar 14. Hasil Individu Non-Diabetes dengan metode LIME

Analisis LIME menunjukkan bahwa seluruh fitur pasien mendukung prediksi non-diabetes dengan probabilitas 93%. Faktor utama yang memperkuat klasifikasi ini adalah kadar glukosa sebesar 124, di bawah ambang 143, sehingga dianggap masih aman. Usia pasien (26 tahun) juga berada di bawah batas risiko 36 tahun, sedangkan nilai *Diabetes Pedigree Function* (0,30) serta *insulin* (130) relatif rendah dibanding ambang model, sehingga memperkuat prediksi non-diabetes. Jumlah *pregnancies* (3) juga di bawah batas 5. Sementara itu, fitur *blood pressure* (80), *skin thickness* (33), dan *BMI* (33,2) sedikit melampaui ambang model, namun kontribusinya tergolong sedang hingga rendah dan tidak mengubah hasil akhir klasifikasi.

7. Analisis Hasil Individu Diabetes SHAP

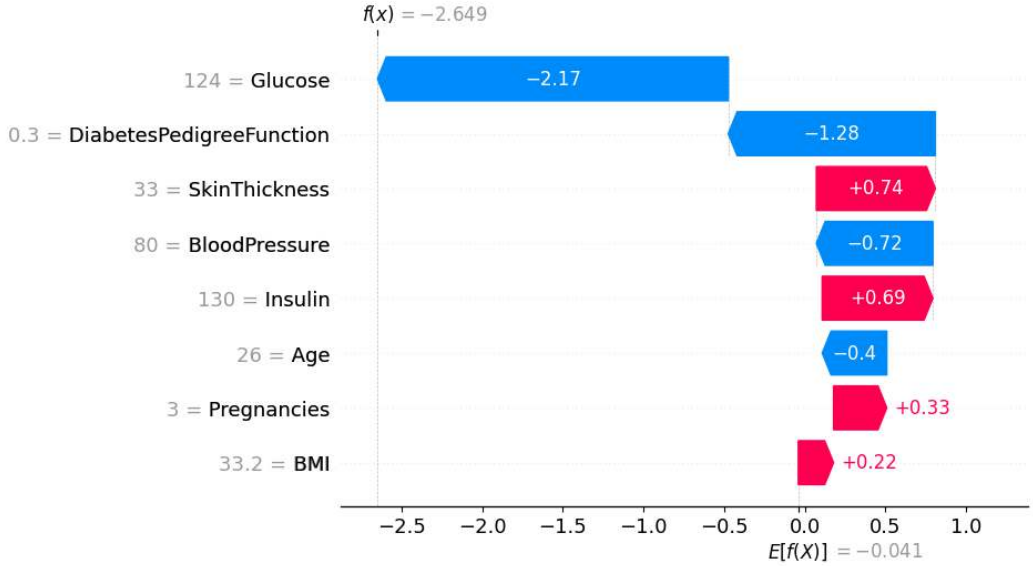


Gambar 15. Hasil Individu Diabetes dengan metode SHAP

Analisis *SHAP* menunjukkan bahwa *glucose* merupakan fitur paling dominan dengan nilai *SHAP* +3,38. Nilai *glucose* pasien sebesar 176, jauh di atas rata-rata model, sehingga kuat mendorong prediksi diabetes. Fitur *pregnancies* (8) dan *blood pressure* (90) juga berkontribusi signifikan dengan nilai *SHAP* +0,77. Nilai *BMI* 33,7 menambah pengaruh positif (+0,74), menegaskan obesitas sebagai faktor risiko utama.

Fitur lain seperti *diabetes pedigree function* (0,467), *skin thickness* (34), dan *insulin* (300) memberikan kontribusi lebih kecil (+0,31, +0,25, +0,23), namun tetap mendukung prediksi positif. Sebaliknya, usia pasien (58) justru memberi efek negatif (-0,17), menunjukkan bahwa dalam model ini usia bukan faktor dominan dibandingkan fitur metabolik lainnya.

8. Analisis Hasil Individu Non-Diabetes SHAP



Gambar 16 Hasil Individu Non-Diabetes dengan metode SHAP

Analisis *SHAP* menunjukkan bahwa *glucose* merupakan faktor paling dominan yang menurunkan kemungkinan diabetes, dengan nilai 124 dan kontribusi *SHAP* -2,17. Fitur *diabetes pedigree function* (0,3) juga memberi kontribusi negatif kuat (-1,28), menandakan riwayat keturunan tidak signifikan terhadap risiko. Selain itu, *blood pressure* (80) dan usia (26) turut menurunkan prediksi dengan nilai *SHAP* -0,72 dan -0,40.

Sebaliknya, beberapa fitur memberi kontribusi positif, meski relatif kecil, seperti *skin thickness* (33, +0,74), *insulin* (130, +0,69), *pregnancies* (3, +0,33), dan *BMI* (33,2, +0,22). Kombinasi seluruh fitur menghasilkan prediksi *SHAP* -2,65, jauh di bawah rata-rata model (-0,04), sehingga profil pasien ini kuat mengarah pada klasifikasi non-diabetes.

5. Kesimpulan

Dari hasil interpretasi terhadap data pasien diabetes dan non-diabetes, kedua metode sepakat bahwa *glukosa* merupakan fitur paling dominan yang mempengaruhi klasifikasi, dengan nilai tinggi yang secara signifikan mendorong model mengarah pada prediksi diabetes. Fitur lain seperti *BMI*, jumlah kehamilan (*Pregnancies*), dan tekanan darah (*BloodPressure*) juga memberikan kontribusi positif yang cukup kuat terhadap risiko diabetes. Di sisi lain, *usia*, *insulin*, dan ketebalan kulit (*SkinThickness*) menunjukkan kontribusi yang bervariasi, tergantung pada masing-masing individu.

Namun demikian, hasil LIME menunjukkan adanya kelemahan dari segi konsistensi interpretasi antar individu, terutama saat parameter seperti *num_samples* yang tidak dioptimalkan. LIME terbukti sensitif terhadap konfigurasi dan distribusi data lokal di sekitar titik prediksi, sehingga interpretasi yang dihasilkan dapat berubah-ubah. Oleh karena itu, agar interpretasi dari LIME lebih stabil dan akurat, perlu dilakukan penyesuaian parameter, seperti meningkatkan jumlah sampel yang dianalisis.

Sebaliknya, SHAP mampu menjelaskan kontribusi setiap fitur terhadap *output* model secara teoritis berdasarkan nilai *Shapley* dan memberikan hasil yang konsisten, transparan, serta dapat dipercaya. Kemampuannya dalam menangkap pengaruh fitur baik pada level individu maupun populasi menjadikannya lebih unggul untuk digunakan dalam aplikasi yang membutuhkan tingkat akurasi dan *interpretability* tinggi.

Dengan demikian, SHAP lebih direkomendasikan untuk digunakan dalam pengujian dan analisis model klasifikasi di bidang kesehatan, khususnya untuk pengidap penyakit diabetes, karena kemampuannya memberikan nilai kontribusi fitur yang lebih konsisten dan terukur.

Daftar Pustaka

[1] Johannes Allgaier, Lena Mulansky, Rachel Lea Draelos, and Rüdiger Pryss. “How does the model make predictions? A systematic literature review on the explainability power of machine learning in healthcare.” *Artificial Intelligence in Medicine*, 2023.

[2] S. Ahmed, M. S. Kaiser, M. Shahadat Hossain and K. Andersson. "A Comparative Analysis of LIME and SHAP Interpreters With Explainable ML-Based Diabetes Predictions." *IEEE Access*, vol. 13, pp. 37370–37388, 2025. https://doi.org/10.1109/ACCESS.2024.3422319.

[3] Nur Arminarahmah and Galih Mahalisa. "Implementasi model machine learning pada klasifikasi status penyakit diabetes berbasis Streamlit." *Smart Comp: Jurnalnya Orang Pintar Komputer*, vol. 13, no. 3, pp. 470–475, 2024. https://doi.org/10.30591/smartcomp.v13i3.5866.

[4] Serg Masis. *Interpretable Machine Learning with Python: Build Explainable, Fair, and Robust High-Performance Models*. Packt Publishing, 2023.

[5] Mubaraqah, Annisa Nurul Puteri, and A. Sumardin. "Comparison of Random Forest and XGBoost for Diabetes Classification with SHAP and LIME Interpretation." *JTERA (Jurnal Teknologi Rekayasa)*, vol. 9, no. 2, pp. 121–130, 2024. https://doi.org/10.31544/jtera.v9.i1.2024.121-130.

[6] Andhika Brahmandjati, Abd Mizwar A. Rahim, and Firman Asharudin. "Optimasi prediksi diabetes dengan algoritma XGBoost dan teknik preprocessing data." *LOGIC: Jurnal Ilmu Komputer dan Pendidikan*, vol. 3, no. 1, pp. 116–125, 2025.

[7] Muhammad Surono, Muhammad Fadli, Dian Sri Purwamti, and Erliyan Redy Susanto. "Hybrid XGBoost-SVM model untuk sistem pendukung keputusan dalam prediksi penyakit diabetes." *INSOLOGI: Jurnal Sains dan Teknologi*, vol. 4, no. 3, pp. 443–454, 2025. https://doi.org/10.55123/insologi.v4i3.5410.

[8] Wildan Hidayatulloh. "Analisis Prediksi Kinerja Akademik Mahasiswa Menggunakan Metode Explainable Artificial Intelligence Berbasis SHAP dan LIME." 1 Juli 2025. https://doi.org/10.13140/RG.2.2.18836.62084.

[9] C. Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*, 1st ed., Leanpub, 2019.

[10] Varad Vishwarupe, Prachi M. Joshi, Nicole Mathias, Shrey Maheshwari, Shweta Mhaisalkar, and Vishal Pawar. "Explainable AI and interpretable machine learning: A case study in perspective." *Procedia Computer Science*, vol. 204, pp. 869–876, 2022.

[11] Ninda Rizky Nuraeda, Muhaza Liebenlito, and Taufik Edy Sutanto. "Explainable sentiment analysis pada ulasan aplikasi Shopee menggunakan Local Interpretable Model-agnostic Explanations." *Indonesian Journal of Computer Science*, 2024.

[12] Y. Wu, L. Zhang, U. A. Bhatti, and M. Huang. "Interpretable machine learning for personalized medical recommendations: A LIME-based approach." *Diagnostics*, vol. 13, no. 16, p. 2681, 2023. https://doi.org/10.3390/diagnostics13162681.

[13] Z. Li. "Extracting spatial effects from machine learning model using local interpretation method: An example of SHAP and XGBoost." *Computers, Environment and Urban Systems*, vol. 96, p. 101845, 2022. https://doi.org/10.1016/j.compenvurbsys.2022.101845.

[14] Mohammad Teddy Syamkalla, Siti Khomsah, and Yohani Setiya Rafika Nur. "Implementasi algoritma CatBoost dan Shapley Additive Explanations (SHAP) dalam memprediksi popularitas game indie pada platform Steam." *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)*, vol. 11, no. 4, pp. 771–776, 2024. https://doi.org/10.25126/jtiik.1148503.

[15] M. M. Islam, H. R. Rifat, M. S. B. Shahid, A. Akhter, M. A. Uddin, and K. M. M. Uddin. "Explainable machine learning for efficient diabetes prediction using hyperparameter tuning, SHAP analysis, partial dependency, and LIME." *Engineering Reports*, vol. 7, e13080, 2025. https://doi.org/10.1002/eng2.13080.

[16] Mesut Toğaçar, Nedim Muzoğlu, Burhan Ergen, Bekir Sıddık Binboğa Yarman, and Ahmet Mesrur Halefoğlu. "Detection of COVID-19 findings by the local interpretable model-agnostic explanations method of types-based activations extracted from CNNs." *Biomedical Signal Processing and Control*, vol. 71, part A, article no. 103128, 2022. https://doi.org/10.1016/j.bspc.2021.103128.

[17] Tsehay Admassu Assegie. "Evaluation of Local Interpretable Model-Agnostic Explanation and Shapley Additive Explanation for Chronic Heart Disease Detection." *Progress in Engineering and Technology Innovation*, 2023. https://doi.org/10.46604/peti.2023.10101.