

Deteksi Dini Risiko Kesehatan Ibu Hamil Menggunakan Algoritma XGBoost dan Analisis *Feature Importance*

Nurhayati^{1*}, Agung Suryadi², Yunita Wisda Tumarta Arif³, Aril Rizal Saputra⁴

¹Rekam Medik dan Informasi
Kesehatan
Universitas Duta Bangsa,
Surakarta
^{1*} nurhayati@udb.ac.id

²Rekam Medik dan Informasi
Kesehatan
Universitas Duta Bangsa,
Surakarta
² agungsuryadi@udb.ac.id

²Rekam Medik dan Informasi
Kesehatan
Universitas Duta Bangsa,
Surakarta
² yunita_wisda@udb.ac.id

⁴Rekam Medik dan Informasi
Kesehatan
Universitas Duta Bangsa,
Surakarta
^{4*} arilrizalsaputra83@gmail.com

Abstrak— Risiko kesehatan pada ibu hamil merupakan salah satu kondisi yang perlu dideteksi sejak dini karena berkaitan dengan kesehatan ibu dan janin. Penelitian ini bertujuan untuk membangun model klasifikasi risiko kesehatan ibu hamil menggunakan algoritma *Extreme Gradient Boosting* (XGBoost) berdasarkan data kondisi kesehatan ibu hamil. *Dataset* yang digunakan terdiri dari 12 variabel klinis, yaitu Usia (*Age*), Tekanan Darah Sistolik (*Systolic Blood Pressure*), Tekanan Darah Diastolik (*Diastolic Blood Pressure*), Kadar Gula Darah (*Blood Sugar*), Suhu Tubuh (*Body Temperature*), Indeks Massa Tubuh (*Body Mass Index*), Riwayat Komplikasi (*Previous Complications*), Riwayat Diabetes Sebelumnya (*Preexisting Diabetes*), Diabetes Kehamilan (*Gestational Diabetes*), Kesehatan Mental (*Mental Health*), Denyut Jantung (*Heart Rate*), Dan Kategori Risiko (*Risk Level*). Target Klasifikasi Terdiri Dari Dua Tingkat Risiko Kehamilan, Yaitu *Low Risk* (Risiko Rendah) Dan *High Risk* (Risiko Tinggi). Tahapan penelitian meliputi pengumpulan data, *preprocessing* data, pembagian *dataset*, pemodelan menggunakan XGBoost, evaluasi model, dan analisis *feature importance*. Hasil penelitian menunjukkan bahwa model XGBoost menghasilkan performa klasifikasi yang sangat baik dengan *accuracy* sebesar 99.16%, *precision* sebesar 0.979, *recall* sebesar 0.986, dan *F1-Score* sebesar 0.99. Hasil analisis *feature importance* menunjukkan bahwa variabel *Blood Sugar*, *Gestational Diabetes*, dan *Preexisting Diabetes* merupakan faktor yang paling berpengaruh terhadap tingkat risiko kesehatan ibu hamil.

Berdasarkan hasil penelitian, algoritma XGBoost layak digunakan sebagai bagian dari sistem deteksi dini risiko kesehatan ibu hamil untuk membantu proses pengambilan keputusan dan pencegahan komplikasi secara lebih cepat dan tepat.

Kata kunci— Deteksi Dini, Kehamilan, Machine Learning, XGBoost, Feature Importance

Abstract— Health risks in pregnant women are a condition that requires early detection because they are related to the health of the mother and fetus. This study aims to develop a maternal health risk classification model using the Extreme Gradient Boosting (XGBoost) algorithm based on maternal health data. The dataset used consists of 12 clinical variables: age, systolic blood pressure, diastolic blood pressure, blood sugar level, body temperature, body mass index, history of complications, preexisting diabetes, gestational diabetes, mental health, heart rate, and risk level. The classification targets consist of two levels of pregnancy risk: Low Risk and High Risk. The research stages include data collection, data preprocessing, dataset division, modeling using XGBoost, model evaluation, and feature importance analysis. The results showed that the XGBoost model produced excellent classification performance with an accuracy of 99.16%, a precision of 0.979, a recall of 0.986, and an F1-score of 0.99. Feature importance analysis revealed that Blood Sugar, Gestational Diabetes, and Preexisting Diabetes were the most influential factors affecting maternal health risk levels. Based on the research results, the XGBoost algorithm is suitable for use as part of an early detection system for maternal health risks to facilitate faster and more accurate decision-making and prevent complications.

Keywords— Early Detection, Pregnancy, Machine Learning, XGBoost, Feature Importance

I. PENDAHULUAN

Kehamilan memiliki risiko komplikasi yang membahayakan kesehatan ibu maupun janin. Komplikasi kehamilan di Indonesia masih cukup tinggi, selama rentang waktu 2019-2024 berkisar 20-25% ibu hamil mengalami minimal satu komplikasi. Jumlah ini berkontribusi pada

meningkatnya angka kematian ibu di Indonesia[1].

Risiko kesehatan ibu hamil dipengaruhi beberapa faktor diantaranya anemia, hipertensi, diabetes kehamilan, status gizi buruk dan kurangnya kunjungan *antenatal care*[2]. Kondisi tersebut menunjukkan penting adanya

deteksi dini risiko kehamilan.

Deteksi dini risiko kehamilan menjadi sangat penting mulai dari awal kehamilan, seperti diabetes gestasional, tekanan darah dan potensi kelahiran prematur. Keterlambatan deteksi risiko pada ibu hamil dapat meningkatkan risiko komplikasi kesehatan dan kematian[3]. Perlu adanya inovasi deteksi dini, melalui algoritma *machine learning* yang mampu menganalisis data kesehatan ibu hamil secara menyeluruh.

Machine learning banyak digunakan untuk membantu diagnosis dan klasifikasi penyakit menggunakan kecerdasan buatan seperti untuk prediksi diabetes[4], prediksi serangan jantung[5] dan klasifikasi kesehatan[6]. Salah satu algoritma yang handal digunakan dalam klasifikasi pada *machine learning* adalah algoritma XGBoost.

XGBoost adalah algoritma pada *machine learning* berbasis *ensemble learning* yang menggunakan teknik *boosting* dan *decision tree* untuk meningkatkan performa prediksi. Cara kerjanya adalah dengan membangun *decision tree* secara bertahap. Keunggulan XGBoost adalah akurasi klasifikasi yang tinggi untuk berbagai masalah prediksi data[7]. XGBoost juga memiliki kemampuan *feature importance* yang berdampak pada hasil prediksi.

Feature importance merupakan suatu metode pada *machine learning* yang berguna untuk mengetahui tingkat pengaruh setiap variabel terhadap hasil prediksi model. Setiap variabel akan diberikan nilai berdasarkan kontribusinya pada proses klasifikasi atau prediksi. Fitur ini membantu untuk mengenali variabel yang paling dominan pada model sehingga berdampak pada faktor yang paling mempengaruhi hasil prediksi/klasifikasi. Proses analisis menjadi lebih cepat karena hubungan antara variabel input dan target output dapat lebih mudah dipahami. Semakin besar nilai *feature importance* maka akan semakin besar pula pengaruh variabel tersebut terhadap hasil prediksi[8].

Beberapa penelitian terdahulu yang mengangkat penerapan beberapa algoritma *machine learning* diantaranya:

- a) Metode *naive bayes* dimanfaatkan untuk klasifikasi kesehatan karyawan dengan tingkat akurasi 70%[6], klasifikasi risiko hipertensi dengan tingkat akurasi 89%[9] dan klasifikasi penyakit jantung dengan tingkat akurasi 80%[10].
- b) *Random forest* digunakan prediksi diabetes dan penyakit jantung dan kanker dengan akurasi 89% dan memiliki performa klasifikasi yang lebih baik dibandingkan metode yang lain pada data klinis[11].
- c) *Logistic regression* digunakan untuk memprediksi risiko penyakit berdasarkan faktor kesehatan pasien dengan akurasi 85%[12].
- d) XGBoost menghasilkan performa terbaik untuk memprediksi risiko diabetes tipe 2 dengan akurasi 96,07%[13], mampu memprediksi preeklamsia berdasarkan data klinis dengan akurasi 92,1%[14], mampu mengklasifikasi risiko kehamilan berdasarkan faktor klinis penting dengan akurasi 94,1%[15].

Berdasarkan penelitian terdahulu menunjukkan bahwa metode XGBoost terbukti mampu menghasilkan akurasi yang sangat tinggi, performa yang sangat baik dan kemampuan prediksi yang tinggi.

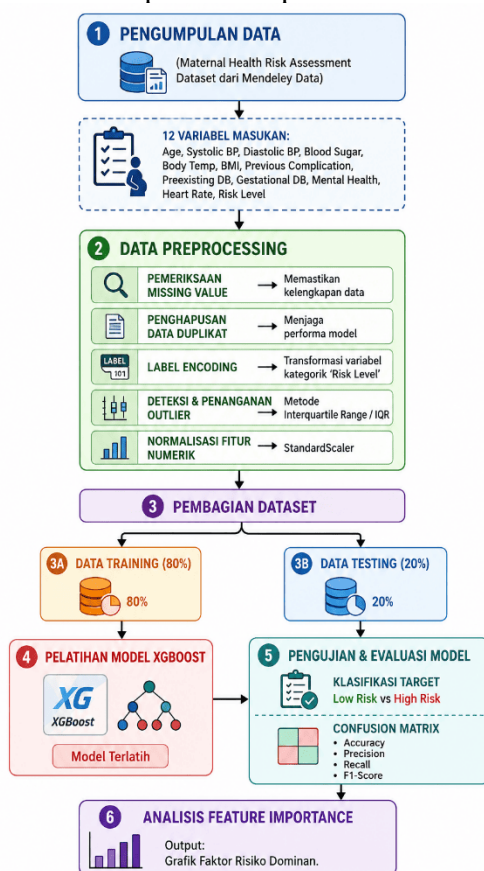
Penelitian ini bermaksud untuk menerapkan algoritma *Extreme Gradient Boosting* (XGBoost) yang bertujuan deteksi dini risiko kesehatan ibu hamil berdasarkan variabel kondisi kesehatan ibu hamil. Penelitian ini bertujuan melakukan klasifikasi risiko kesehatan ibu hamil menggunakan algoritma XGBoost, menganalisis performa XGBoost berdasarkan *accuracy*, *precision*, *recall* dan *F1 Score*, mengidentifikasi variabel yang paling berpengaruh terhadap prediksi risiko kesehatan ibu hamil menggunakan *feature importance*.

II. METODOLOGI PENELITIAN

Penelitian ini merupakan penelitian kuantitatif dengan metode eksperimen menggunakan

algoritma *Extreme Gradient Boosting* (XGBoost) yang digunakan untuk melakukan deteksi dini kesehatan ibu hamil. Penelitian berfokus pada pembuatan model klasifikasi dan analisis *feature importance* untuk mengetahui faktor yang berpengaruh terhadap risiko kesehatan ibu hamil.

Metode penelitian meliputi beberapa tahapan mulai dari pengumpulan data hingga analisis hasil akhir. Alur arsitektur model dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Arsitektur Model Deteksi Dini Risiko Kesehatan Ibu Hamil menggunakan XGBoost.

Adapun penjelasan dari masing-masing tahapan pada Gambar 1 adalah sebagai berikut:

a) Pengumpulan data

Dataset yang digunakan dalam penelitian ini merupakan data kesehatan ibu hamil yang diperoleh dari *Maternal Health Risk Assessment Dataset* yang tersedia secara publik pada platform Mendeley Data[16].

Dataset terdiri atas 12 variabel klinis ibu hamil yaitu usia (*Age*), tekanan darah sistolik (*Systolic Blood Pressure*), tekanan darah diastolik (*Diastolic Blood Pressure*), kadar gula

darah (*Blood Sugar*), suhu tubuh (*Body Temperatur*), Massa Indeks Tubuh (*Body Mass Index*), Riwayat Komplikasi (*Previous Complication*), Riwayat diabetes sebelumnya (*Preexisting Diabetes*), Diabetes Kehamilan (*Gestational Diabetes*), Kesehatan Mental (*Mental Health*), Denyut Jantung (*Heart Rate*) dan Kategori Risiko (*Risk Level*). Target klasifikasi 2 tingkat risiko kehamilan yaitu *Low Risk* (Risiko Rendah) dan *High Risk* (Risiko Tinggi).

b) Preprocessing Data

Tahap *preprocessing* dilakukan untuk meningkatkan kualitas data sebelum proses pemodelan menggunakan algoritma XGBoost. Proses ini meliputi pemeriksaan *missing value* untuk memastikan kelengkapan data, penghapusan data duplikat agar tidak memengaruhi performa model, serta transformasi variabel kategorik pada kolom *Risk Level* menjadi bentuk numerik menggunakan *label encoding*. Berikutnya dilakukan deteksi dan penanganan *outlier* pada variabel numerik seperti tekanan darah, gula darah, suhu tubuh, dan denyut jantung menggunakan metode *Interquartile Range* (IQR). Seluruh fitur numerik dinormalisasi menggunakan *StandardScaler* agar memiliki skala yang seimbang sebelum digunakan pada proses pelatihan model.

c) Pembagian Dataset

Dataset yang telah melalui tahap *preprocessing* kemudian dibagi menjadi data *training* sebesar 80% dan data *testing* sebesar 20% menggunakan metode *train-test split* dengan teknik *stratified split* untuk menjaga pembagian kelas risiko yang seimbang.

d) Pemodelan Menggunakan XGBoost

Pada tahap pemodelan ini, algoritma XGBoost digunakan untuk mengklasifikasikan risiko kesehatan ibu hamil berdasarkan variabel usia, tekanan darah sistolik, tekanan darah diastolik, kadar gula darah, suhu tubuh, *body mass index*, riwayat komplikasi, *preexisting diabetes*, *gestational diabetes*, kesehatan mental, dan denyut jantung. XGBoost dipilih karena mampu

memberikan hasil klasifikasi yang baik dan dapat mengolah data dengan efektif. Model dilatih menggunakan data *training* dan diuji menggunakan data *testing*.

e) Evaluasi Model

Kinerja model dievaluasi menggunakan *confusion matrix* dengan parameter:

- 1) *Accuracy*: untuk mengukur ketepatan prediksi model.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

- 2) *Precision*: untuk mengukur ketepatan prediksi masing-masing kelas.

$$Precision = \frac{TP}{TP + FP}$$

- 3) *Recall*: untuk mengukur kemampuan model dalam mendeteksi setiap kelas Risiko.

$$Recall = \frac{TP}{TP + FN}$$

Keterangan:

TP (*True Positive*) = data positif yang diprediksi benar,

TN (*True Negative*) = data negatif yang diprediksi benar,

FP (*False Positive*) = data negatif tetapi diprediksi positif,

FN (*False Negative*) = data positif tetapi diprediksi negative.

- 4) *F1-Score*: untuk mengukur keseimbangan *precision* dan *recall*.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

- 5) *Confusion matrix*: untuk mengukur jumlah prediksi benar dan salah pada masing masing kategori risiko.

f) Analisis *Feature Importance*

Analisis *feature importance* dilakukan setelah model XGBoost dibuat, analisis ini digunakan untuk mengetahui fitur yang paling berpengaruh dalam prediksi risiko kesehatan ibu hamil. Analisis dilakukan dengan melihat pengaruh masing masing variabel diantaranya usia, tekanan darah sistolik, tekanan darah diastolik, gula darah, suhu tubuh, *body mass index*, *previous complications*, *preexisting diabetes*, *gestational diabetes*, *mental health*, dan denyut

jantung terhadap hasil prediksi model. Hasil analisis tersebut diolah menjadi grafik untuk memudahkan dalam memahami faktor dominan dalam menentukan tingkat risiko kehamilan.

III. HASIL DAN PEMBAHASAN

a) Hasil *Preprocessing*

Dataset awal terdiri atas 1205 data, setelah dilakukan pemeriksaan *missing value* dan data duplikat maka sebanyak 57 data hilang, jumlah data menjadi 1148 data yang siap digunakan.

Dilanjutkan tahapan transformasi data untuk mengubah variabel target *risk level* ke dalam bentuk numerik dengan tujuan label *encoding*. *Low Risk* dikode 0, dan *High Risk* dikode 1, tujuannya untuk memudahkan data diproses oleh algoritma *machine learning*.

Tahapan selanjutnya memastikan semua variabel bertipe numerik untuk dapat diproses XGBoost. Variabel x adalah seluruh variabel input, variabel y adalah variabel target (*risk level*). Tahap normalisasi data menggunakan *standar scaler* bertujuan untuk menyamakan rentang seluruh fitur, mengurangi perbedaan rentang nilai antar variabel serta model menjadi lebih stabil.

Dataset terbagi atas 80% data *training* dan 20% data *testing* menggunakan metode *stratified split*.

Proses pemodelan menggunakan algoritma XGBoost telah berhasil dilakukan. Tahap awal dilakukan dengan menginstal library XGBoost, kemudian model XGBoost dibuat menggunakan XGBClassifier. Selanjutnya model dilatih menggunakan data *training* melalui fungsi model fit (*x_train*, *y_train*). Hasil output menunjukkan bahwa proses pelatihan model berjalan dengan baik dan sistem menampilkan parameter default dari algoritma XGBoost.

Proses prediksi menggunakan data *testing* melalui fungsi model predict (*x_test*) untuk menghasilkan prediksi tingkat risiko kesehatan ibu hamil yang akan digunakan pada tahap evaluasi model.

b) Hasil Evaluasi Model

Hasil evaluasi model tersaji pada tabel 1.

Tabel 1. Matrik Konfusi Klasifikasi

Kelas	TP	TN	FP	FN
0 Low Risk	141	96	0	2
1 High Risk	95	141	2	0

Berdasarkan tabel 1 diketahui rincian perhitungan evaluasi kinerja model sebagai berikut:

1) Accuracy

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

$$Accuracy = \frac{141 + 95}{141 + 95 + 2 + 0}$$

$$Accuracy = \frac{236}{238} = 0.9916 = 99.16\%$$

Hasil tersebut menunjukkan bahwa model XGBoost telah mampu menghasilkan klasifikasi yang sangat baik dengan tingkat akurasi 99,16%.

2) Precision

Perhitungan *precision* dilakukan dengan mengukur ketepatan model dalam memberikan prediksi positif, perhitungan dilakukan per kelas, sebagai berikut:

i) Precision Kelas 0 (Low Risk)

Diketahui nilai TP 141 dan nilai FP 0, maka:

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{141}{141 + 0} = 1.00$$

Hasil tersebut menandakan bahwa seluruh data yang telah diprediksi sebagai kelas *low risk* adalah benar, sedangkan tidak ada data kategori *high risk* yang salah diprediksi menjadi *low risk*.

ii) Precision Kelas 1 (High Risk)

Diketahui nilai TP 95, nilai FP 2, maka:

$$Precision = \frac{TP}{TP + FP}$$

$$Precision = \frac{95}{95 + 2} = 0.979$$

Hasil tersebut menandakan bahwa dari 97 data yang diprediksi sebagai *high risk*, 95 data benar benar masuk kategori *high risk*, sedangkan 2 data

salah prediksi. Hal tersebut menunjukkan model XGBoost mampu melakukan prediksi kelas *high risk* dengan ketepatan yang sangat tinggi 97,9%.

3) Recall

Recall digunakan untuk mengukur kemampuan model untuk mengenali seluruh data pada suatu kelas. Perhitungan *recall* dilakukan perkelas yaitu:

i) Recall Kelas 0 (Low Risk)

Diketahui nilai TP 141, nilai FN 2, maka:

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{141}{141 + 2} = \frac{141}{143} = 0.986$$

Hasil tersebut menandakan bahwa model mampu mengenali hampir seluruh data *low risk*, hanya 2 data *low risk* yang gagal dikenali dan salah diprediksi sebagai kategori *high risk*.

ii) Recall Kelas 1 (High Risk)

Diketahui nilai TP 95, nilai FN 0, maka:

$$Recall = \frac{TP}{TP + FN}$$

$$Recall = \frac{95}{95 + 0} = 1.00$$

Hasil tersebut menandakan bahwa seluruh data *high risk* telah berhasil diklasifikasikan secara benar.

4) F1 Score

F1 Score digunakan untuk mengukur keseimbangan antara *precision* dengan *recall*. Perhitungan *F1 Score* dilakukan perkelas yaitu:

i) F1 Score Kelas 0 (Low Risk)

Diketahui nilai *precision* 1.00, nilai *recall* 0.986, maka:

$$F1 = 2 \times \frac{precision \times recall}{precision + recall}$$

$$F1 = 2 \times \frac{1.00 \times 0.986}{1.00 + 0.986}$$

$$F1 = 2 \times \frac{0.986}{1.986}$$

$$F1 = 2 \times 0.496$$

$$F1 = 0.993$$

Berdasarkan perhitungan tersebut maka dapat dinilai bahwa model

memiliki keseimbangan yang sangat baik antara kemampuan ketepatan untuk memprediksi dan kemampuan mengenali data *low risk*, karena hasil mendekati 1.

ii) *F1 Score* Kelas 1 (*High Risk*)

Diketahui nilai *precision* 0.979, nilai *recall* 1.00, maka:

$$F1 = 2 \times \frac{0.979 \times 1.00}{0.979 + 1.00}$$

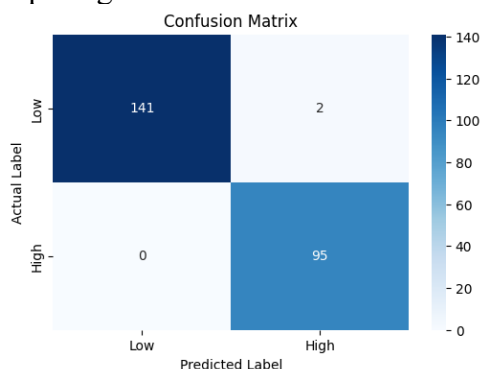
$$F1 = 2 \times \frac{0.979}{1.979}$$

$$F1 = 2 \times 0.495$$

$$F1 = 0.989$$

Berdasarkan hasil tersebut menandakan bahwa model memiliki kemampuan untuk mempertahankan keseimbangan yang sangat baik antara *precision* dan *recall* pada kelas *high risk*, karena hasil mendekati angka 1.00.

Hasil grafik matrik konfusi tersaji pada gambar 2:



Gambar 2 Hasil Matrik Konfusi Model

Berdasarkan gambar 2 menunjukkan model XGBoost mampu mengklasifikasikan sebagian besar data dengan benar. Model berhasil memprediksi 141 data *Low Risk* dan 95 data *High Risk* secara tepat, hanya 2 kesalahan prediksi pada kelas *Low Risk*. Hal ini menunjukkan bahwa model memiliki performa klasifikasi yang sangat baik.

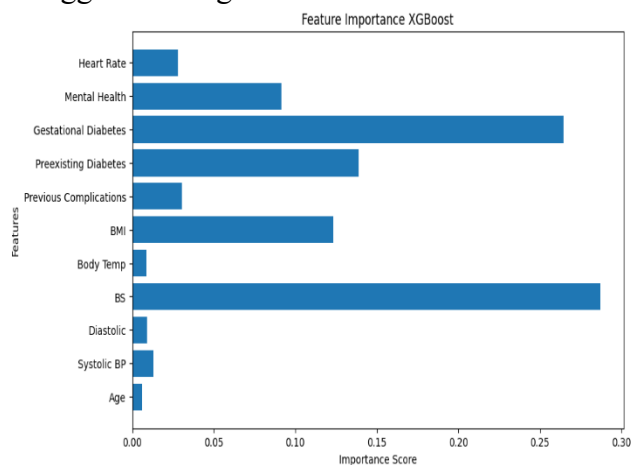
c) *Feature Importance*

Nilai *importance* tersaji pada gambar 3:

```
Age : 0.0057877805
Systolic BP : 0.01304724
Diastolic : 0.008892881
BS : 0.2868989
Body Temp : 0.008516766
BMI : 0.123460144
Previous Complications : 0.030626653
Preexisting Diabetes : 0.13901289
Gestational Diabetes : 0.2643753
Mental Health : 0.09142486
Heart Rate : 0.027956532
```

Gambar 3 Nilai Importance

Hasil analisis *feature importance* menunjukkan bahwa variabel gula darah (*Blood Sugar*) memiliki pengaruh yang paling besar terhadap hasil prediksi risiko kesehatan ibu hamil dengan nilai *importance* 0,286. Sedangkan variabel *gestational diabetes* dan *preexisting diabetes* juga memberikan pengaruh yang cukup tinggi terhadap model yaitu 0,264 dan 0,139. Hal ini berbanding terbalik dengan variabel umur (*Age*) dan *body temperature* yang memiliki nilai *importance* yang rendah yaitu 0,005 dan 0,008, sehingga dua variabel ini memiliki kontribusi yang kecil terhadap proses klasifikasi. Dapat disimpulkan jika faktor gula darah dan riwayat diabetes menjadi variabel penentu dalam proses prediksi risiko kesehatan ibu hamil menggunakan algoritma XGBoost.



Gambar 4 Grafik *Feature Importance*

Gambar 4 menunjukkan bahwa model XGBoost mampu menunjukkan variabel mana yang paling berpengaruh pada proses prediksi risiko kesehatan ibu hamil.

Sumbu horisontal menunjukkan nilai *importance score*, sumbu vertikal menunjukkan variabel-variabel yang digunakan dalam penelitian. Semakin panjang batang pada grafik, maka semakin besar pula pengaruh variabel tersebut terhadap hasil prediksi dari model.

Berdasarkan gambar 4 menunjukkan bahwa variabel *blood sugar/BS* (gula darah) memiliki batang paling panjang dengan nilai *feature importance* yang paling tinggi yaitu 0,286, diikuti oleh *gestational diabetes* dan *preexisting diabetes* yang memiliki panjang batang urutan ke-2 dan ke-3.

Berdasarkan hasil penelitian ini diketahui bahwa model XGBoost memiliki akurasi 99,16% yang memandakan bahwa model XGBoost merupakan model yang sangat akurat.

Hal tersebut sesuai dengan beberapa penelitian yang menyatakan bahwa algoritma XGBoost memiliki kemampuan yang baik dalam melakukan klasifikasi dan prediksi data kesehatan karena mampu mempelajari hubungan antar variabel secara efektif. XGBoost bekerja dengan menggabungkan banyak *decision tree* untuk memperbaiki kesalahan prediksi sebelumnya sehingga menghasilkan akurasi yang tinggi[17]. Algoritma ini juga memiliki keunggulan dalam mengurangi *overfitting*, memproses data dengan cepat, serta mampu menentukan variabel yang paling berpengaruh melalui *feature importance*[18]. Penelitian lain menunjukkan bahwa XGBoost memiliki performa klasifikasi yang lebih baik dibandingkan beberapa metode *ensemble* lainnya seperti *Random Forest* dan *Gradient Boosting Machine*[19]. XGBoost juga banyak digunakan dalam berbagai bidang karena mampu menangani data kompleks, menghasilkan prediksi yang akurat, dan memiliki kemampuan generalisasi yang baik[20].

Hasil analisis *feature importance* menunjukkan bahwa variabel *blood sugar*, *gestational diabetes* dan *preexisting diabetes* menempati tiga nilai tertinggi. Hasil tersebut menandakan bahwa kondisi gula darah, riwayat diabetes sangat mempengaruhi risiko kesehatan ibu hamil.

Hal tersebut sesuai dengan beberapa penelitian yang menyatakan bahwa kadar gula darah yang tinggi akan meningkatkan risiko komplikasi kesehatan ibu hamil. Penelitian sebelumnya menunjukkan bahwa kadar gula darah yang tinggi selama kehamilan dapat

meningkatkan risiko komplikasi pada ibu maupun janin. Kadar gula darah yang tidak terkontrol berhubungan dengan meningkatnya risiko *diabetes gestasional*, *preeklamsia*, hipertensi kehamilan, persalinan prematur, hingga gangguan pertumbuhan janin[21]. Kadar gula darah yang tinggi pada awal kehamilan juga diketahui dapat menyebabkan perubahan pertumbuhan janin dan meningkatkan risiko komplikasi kelahiran[22]. *Diabetes gestasional* merupakan salah satu komplikasi kehamilan yang umum terjadi akibat peningkatan kadar gula darah selama kehamilan dan dapat berdampak buruk terhadap kesehatan ibu dan bayi apabila tidak terdeteksi sejak dini[23].

Model XGBoost mampu mengenali tiga variabel tersebut sebagai faktor penentu yang paling berpengaruh pada proses klasifikasi risiko kesehatan ibu hamil. Hasil penelitian menunjukkan bahwa prediksi berbasis *machine learning* mampu membantu tenaga medis untuk melakukan deteksi dini dan melakukan pencegahan dan penanganan terhadap kemungkinan komplikasi kehamilan.

Hasil penelitian sesuai dengan beberapa penelitian yang menyatakan bahwa deteksi dini menggunakan *machine learning* dapat membantu tenaga kesehatan melakukan pencegahan dan penanganan penyakit lebih cepat dan lebih tepat, mampu mengenali pola risiko penyakit sejak awal sehingga tindakan medis dapat dilakukan sebelum kondisi pasien menjadi lebih parah[24].

Deteksi dini dengan *machine learning* membantu meningkatkan kualitas pelayanan kesehatan serta mengurangi keterlambatan diagnosis[25], meningkatkan akurasi deteksi penyakit, mempercepat proses diagnosis, serta membantu pengambilan keputusan medis secara lebih efisien[26].

IV. KESIMPULAN

Penelitian ini berhasil menerapkan algoritma *Extreme Gradient Boosting* (XGBoost) untuk mengklasifikasi risiko kesehatan ibu hamil berdasarkan variabel usia, tekanan darah diastolik, tekanan darah sistolik, gula darah, suhu tubuh, *body mass index*, riwayat

komplikasi, diabetes kehamilan, riwayat diabetes sebelumnya, kesehatan mental dan denyut jantung. Model XGBoost memiliki performa yang sangat baik dengan *accuracy* 99,16%, *precision*, *recall* dan *F1 Score* yang sangat baik pada semua kelas.

Hasil *feature importance* menunjukkan variabel gula darah (*Blood Sugar*), diabetes kehamilan (*Gestational Diabetes*), riwayat diabetes sebelumnya (*Preexisting diabetes*) menjadi faktor yang paling mempengaruhi tingkat risiko kesehatan ibu hamil. Secara garis besar kondisi gula darah dan riwayat diabetes memiliki keterkaitan yang kuat terhadap risiko komplikasi kehamilan.

Algoritma XGBoost memiliki peluang yang sangat baik sebagai bagian dari sistem deteksi dini risiko kesehatan ibu hamil yang membantu tenaga kesehatan untuk mempercepat pengambilan keputusan medis.

REFERENSI

- [1] B. P. Statistik, "Profil Kesehatan Ibu dan Anak 2024," Badan Pusat Statistik, Jakarta, 2024. [Daring]. Tersedia pada: https://www.bps.go.id/id/publication/2024/12/31/a919c55a72b74e33d011b0dc/profil-kesehatan-ibu-dan-anak-2024.html?utm_source=copilot.com
- [2] S. T. Salsabila, R. Widiastih, dan L. Mamuroh, "Jurnal Keperawatan Karya Bhakti DETERMINAN TERJADINYA KOMPLIKASI KEHAMILAN: ANALISIS DATA INDONESIAN DEMOGRAPHIC AND HEALTH SURVEYS (IDHS) 2017 Pendahuluan," *J. Keperawatan Karya Bhakti*, vol. 11, no. 1, hal. 67–79, 2025, [Daring]. Tersedia pada: <https://journal.akperkbn.com/index.php/jkkb/article/view/194>
- [3] M. R. Sarker, A. E. Skeith, H. Bacheller, A. B. Caughey, dan A. M. Valent, "585: Impact of delayed initiation of prenatal care on maternal outcomes," *Am. J. Obstet. Gynecol.*, vol. 220, no. 1, hal. S388–S389, 2019, doi: 10.1016/j.ajog.2018.11.607.
- [4] W. Apriliah, I. Kurniawan, M. Baydhowi, dan T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *Sist. Sist. Inf.*, vol. 10, hal. 163–171, 2021, [Daring]. Tersedia pada: <https://sistemasi.ftik.unisi.ac.id/index.php/stmsi/article/view/1129/355>
- [5] S. Dharmawan, V. Fernandes, dan H. Halim, "PREDIKSI SERANGAN JANTUNG DENGAN MENGGUNAKAN METODE LOGISTIC REGRESSION CLASSIFIER DAN ADABOOST," *Comput. J. Comput. Sci. Inf. Syst.*, vol. 8, no. 1, hal. 96–103, 2024, doi: <https://doi.org/10.24912/computatio.v8i1.15176>.
- [6] Z. Dafianto, Y. Yunefti, dan A. Digdaya, "Model Klasifikasi Kesehatan Karyawan Berbasis Naïve Bayes Classifier dan K-Nearest Neighbor," in *Prosiding-Seminar Nasional Teknologi Informasi & Ilmu Komputer (SEMESTER)*, 2024, vol. 3, no. 1, hal. 277–288. [Daring]. Tersedia pada: <https://journal.unilak.ac.id/index.php/Semaster/article/view/24126>
- [7] T. Chen dan C. Guestrin, "XGBoost: A Scalable Tree Boosting System," Washington, 2016. doi: <https://doi.org/10.1145/2939672.2939785>.
- [8] H. Kaneko, "Interpretation of Machine Learning Models for Data Sets with Many Features Using Feature Importance," *ACS OMEGA*, vol. 8, no. 25, hal. 23218–23225, 2023, doi: 10.1021/acsomega.3c03722.
- [9] M. N. Zanis dan G. Ghufro, "IMPLEMENTASI METODE NAIVE BAYES UNTUK KLASIFIKASI RISIKO PENYAKIT HIPERTENSI," *J. Nusantara Eng.*, vol. 8, no. 2, hal. 409–416, 2025, [Daring]. Tersedia pada: <https://ojs.unpkediri.ac.id/index.php/noe/article/view/25207>
- [10] A. D. A. S. Surya dan Y. Yamasari, "Penerapan Algoritma Naïve Bayes (NB) untuk Klasifikasi Penyakit Jantung," *J. Informatics Comput. Sci.*, vol. 05, no. 3, hal. 447–455, 2024, doi: <https://doi.org/10.26740/jinacs.v5n03.p447-455>.
- [11] V. Jackins, S. Vimal, M. Kaliappan, dan M. Y. Lee, "AI-based smart prediction of clinical disease using random forest classifier and Naive Bayes," *J. Supercomput.*, vol. 77, no. 5, hal. 5198–5219, 2021, doi: 10.1007/s11227-020-03481-x.
- [12] K. J. Olowe, N. L. Edoh, S. J. C. Zouo, dan J. Olamijuwon, "Comprehensive review of logistic regression techniques in predicting health outcomes and trends," *World J. Adv. Pharm. Life Sci.*, vol. 7, no. 2, hal. 016–026, 2024, [Daring]. Tersedia pada: <https://zealjournals.com/wjapls/content/comprehensive-review-logistic-regression-techniques-predicting-health-outcomes-and-trends>
- [13] Z. Rafie, M. S. Talab, B. E. Z. Koor, A. Garavand, C. Salehnasab, dan M. Ghaderzadeh, "Leveraging XGBoost and explainable AI for accurate prediction of type 2 diabetes," *BMC Public Health*, vol. 25, no. 3688, hal. 1–17, 2025, doi: <https://doi.org/10.1186/s12889-025-24953-w>.
- [14] I. Shyu, C.-F. Liu, Y.-C. Tsai, Y.-S. Ma, T.-N. Kuo, dan S. Y.-Ling, "Machine learning predictive system to predict the risk of developing pre-eclampsia," *BMJ Heal. Care Informatics*, vol. 32, no. 1, hal. 1–8, 2025, doi: 10.1136/bmjhci-2024-101151.
- [15] Y. Wahyuni, H. Hadiyanto, R. Sanjaya, dan N. Herdianto, "A Comparative Analysis and Interpretation of Random Forest , XGBoost , and SVM (RBF) Models with Explainable AI for Pregnancy Risk Classification," *SISFORMA J. Inf. Syst.* (, vol. 12, no. 2, hal. 191–204, 2025, doi: 10.24167/sisforma.v12i2.14292.
- [16] M. U. Mojumdar et al., "Maternal Health Risk Assessment Dataset," Mendeley Data, 2024. doi: 10.17632/p5w98dvbbk.1.
- [17] R. Mitchell dan E. Frank, "Accelerating the XGBoost algorithm using GPU computing," *PeerJ Comput. Sci.*, 2017, doi: 10.7717/peerj-cs.127.
- [18] Y. Zhu, "Stock Price Prediction based on LSTM and XGBoost Combination Model," *Trans. Comput. Sci. Intell. Syst. Res.*, vol. 1, no. 1, hal. 1–16, 2023, [Daring]. Tersedia pada: <https://wepub.org/index.php/TCSISR/article/view/90>
- [19] E. K. Sahin, "Assessing the predictive capability of ensemble tree methods for landslide susceptibility mapping using XGBoost , gradient boosting machine , and random forest," *SN Appl. Sci.*, vol. 2, no. 7, hal. 1–17, 2020, doi: 10.1007/s42452-020-3060-1.
- [20] S. Fatima et al., "XGBoost and Random Forest Algorithms : An In-Depth Analysis," *Pakistan J. Sci. Res.*, vol. 3, no. 1, hal. 26–31, 2023, [Daring]. Tersedia pada: <https://paas-pk.org/index.php/pjost/article/view/1876>
- [21] C. A. Negrato, R. Mattar, dan M. B. Gomes, "Adverse pregnancy outcomes in women with diabetes," *Diabetol. Metab. Syndr.*, vol. 4, no. 41, hal. 1–6, 2012, [Daring]. Tersedia pada: <https://pubmed.ncbi.nlm.nih.gov/articles/PMC3514247/>
- [22] M. L. Geurtsen, E. E. L. Van Soest, E. Voerman, E. A. P. Steegers, V. W. V. Jaddoe, dan R. Gaillard, "High maternal early-pregnancy blood glucose levels are associated with altered fetal growth and increased risk of adverse birth outcomes," *Diabetologia*, vol. 62, hal. 1880–1890, 2019, [Daring]. Tersedia pada: <https://pubmed.ncbi.nlm.nih.gov/31392381/>
- [23] F. K. Adli, "DIABETES MELITUS GESTASIONAL: DIAGNOSIS DAN FAKTOR RISIKO," *J. Med. Utama*, vol. 03, no. 01, hal. 1545–1551, 2021, [Daring]. Tersedia pada: <https://jurnalmedikahutama.com/index.php/JMH/article/view/3>

- 12
- [24] O. Zapata-cortes, M. D. Arango-Serna, Andres, J. A. Zapata-Cortes, dan J. A. Restrepo-Carmona, "Machine Learning Models and Applications for Early Detection," *Sensors*, vol. 24, no. 14, hal. 1–20, 2024, [Daring]. Tersedia pada: <https://www.mdpi.com/1424-8220/24/14/4678>
- [25] R. Islam, M. Kabir, M. F. Mridha, S. Alfarhood, M. Safran, dan D. Che, "Deep Learning-Based IoT System for Remote Monitoring and Early Detection of Health Issues in Real-Time," *Sensors*, vol. 23, no. 11, hal. 1–18, 2023, [Daring]. Tersedia pada: <https://www.mdpi.com/1424-8220/23/11/5204>
- [26] M. M. Ahsan, S. A. Luna, dan Z. Siddique, "Machine-Learning-Based Disease Diagnosis: A Comprehensive Review," vol. 10, no. 3, hal. 1–30, 2022, [Daring]. Tersedia pada: <https://www.mdpi.com/2227-9032/10/3/541>