JIMU: Jurnal Ilmiah Multi Disiplin
Vol. 02, N0. 04, Tahun 2024, Hal. 803-812, ISSN: 3031-9498 (Online)

Prediksi Dini Penyakit Diabetes Pada Ibu Hamil Dengan Algoritma Random Forest

Resa Budi Prasetyo
Teknik Informatika, Universitas Muhammadiyah Ponorogo
Email: resaprasetyo17@gmail.com

Article Info

Article history:

Received May 25, 2024
Revised May 29, 2024
Accepted June 12, 2024

Keywords:
diabetes; ibu hamil; random forest; machine learning

ABSTRACT

Diabetes in pregnant women, otherwise known as gestational diabetes, is a serious health condition that can negatively affect the mother and fetus. Early detection and appropriate intervention are essential to reduce the risk of complications during pregnancy and childbirth. This study aims to develop a prediction model that can detect the risk of diabetes in pregnant women using the Random Forest algorithm. The data used in this study included relevant clinical and demographic parameters of pregnant women, such as age, body mass index, family history, and laboratory test results. The Random Forest model was chosen because of its ability to handle complex data and provide accurate prediction results. The results of this study show that the model developed has an accuracy rate of 98% in predicting the risk of gestational diabetes. In addition, important features that contribute to diabetes risk prediction have also been identified, providing additional insights for medical practitioners in conducting risk evaluations. The implementation of this model is expected to help in efforts to prevent and manage gestational diabetes, thereby improving maternal and infant health.

This is an open access article under the CC BY-SA license.

Article Info

Article history:

Received May 25, 2024
Revised May 29, 2024
Accepted June 12, 2024

Keywords:

diabetes; ibu hamil; random forest; machine learning

ABSTRACT

Diabetes pada ibu hamil, atau dikenal sebagai diabetes gestasional, merupakan kondisi kesehatan serius yang dapat berdampak negatif pada ibu dan janin. Deteksi dini dan intervensi yang tepat sangat penting untuk mengurangi risiko komplikasi selama kehamilan dan persalinan. Penelitian ini bertujuan untuk mengembangkan model prediksi yang dapat mendeteksi risiko diabetes pada ibu hamil menggunakan algoritma Random Forest. Data yang digunakan dalam penelitian ini mencakup parameter klinis dan demografis yang relevan dari ibu hamil, seperti usia, indeks massa tubuh, riwayat keluarga, dan hasil tes laboratorium. Model Random Forest dipilih karena kemampuannya dalam menangani data yang kompleks dan memberikan hasil prediksi yang akurat. Hasil dari penelitian ini menunjukkan bahwa model yang dikembangkan memiliki tingkat akurasi sebesar 98% dalam memprediksi risiko diabetes gestasional. Selain itu, fitur penting yang berkontribusi terhadap prediksi risiko diabetes

juga telah diidentifikasi, memberikan wawasan tambahan bagi praktisi medis dalam melakukan evaluasi risiko. Implementasi model ini diharapkan dapat membantu dalam upaya pencegahan dan manajemen diabetes gestasional, sehingga meningkatkan kesehatan ibu dan bayi.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Nama penulis: Resa Budi Prasetyo
Teknik Informatika, Universitas Muhammadiyah Ponorogo
Email: resaprasetyo17@gmail.com

Pendahuluan

Diabetes Mellitus Gestasional adalah salah satu bentuk diabetes yang khusus terjadi selama kehamilan. Hal ini ditandai dengan peningkatan kadar glukosa (hiperglikemia) dan dapat menyebabkan komplikasi seperti makrosomia, hipoglikemia janin, kebutuhan perawatan intensif pada bayi baru lahir, dan bahkan kematian neonatal[1]. Diabetes melitus gestasional adalah komplikasi kehamilan yang paling umum. Salah satu perubahan hormon yang terjadi selama kehamilan adalah kondisi ini, dan itu harus diantisipasi agar tidak ber-dampak buruk pada ibu, penyakit diabetes ini meningkat 7% per tahun[2].

Aryo Pinandito dkk. (2019) melakukan penelitian mengenai Implementasi Machine Learning untuk prediksi penyakit Diabetes Melitus dengan metode yang digunakan adalah K-Nearest Neighbor (KKN). Metode K-Nearest Neighbor (KNN) dapat digunakan untuk mendeteksi risiko kehamilan diabetes mellitus pada ibu hamil dengan menggunakan perangkat lunak Weka. Dataset yang digunakan terdiri dari 248 data pemeriksaan ibu hamil yang dikumpulkan dari Puskesmas Lawang dari Januari hingga Agustus 2021, dengan 16 atribut penting yang digunakan dalam klasifikasi. Hasil pengujian menunjukkan bahwa akurasi tertinggi dicapai pada skenario pengujian ketiga dengan nilai k=4 sebesar 78.38%, menggunakan rasio data latih dan uji sebesar 7:3. Namun, penelitian ini menghadapi keterbatasan berupa ketidakseimbangan data dalam dataset, yang mempengaruhi hasil akurasi. Oleh karena itu, metode KNN dalam penelitian ini hanya dapat digunakan sebagai alat bantu deteksi dini dan bukan sebagai pengganti diagnosis medis.

Berbeda dengan penelitian sebelumnya yang menggunakan algoritma KKN, penelitian ini menggunakan algoritma Random Forest dengan nilai akurasi pengujian yang dilakukan lebih tinggi, yaitu sebesar 98%. Sumber dataset yang digunakan berasal dari kaggle yang diunggah pada tahun 2018 berisi data pemeriksaan ibu hamil yang berada di Rumah Sakit Frankrut, Jerman. Algoritma Random Forest, yang merupakan metode ensemble learning dengan menggabungkan hasil dari beberapa pohon keputusan untuk meningkatkan akurasi prediksi, menunjukkan keunggulannya dalam menangani dataset yang digunakan. Dengan kemampuan untuk menangani data yang tidak seimbang dan mengurangi risiko overfitting,

Random Forest memberikan hasil yang lebih konsisten dan andal. Hal ini menunjukkan bahwa algoritma Random Forest mungkin lebih efektif dalam mendeteksi risiko kehamilan diabetes melitus pada ibu hamil dibandingkan dengan metode K-Nearest Neighbor (KNN) yang digunakan dalam penelitian sebelumnya.

Secara keseluruhan penelitian ini akan melibatkan beberapa tahapan utama: Pengumpulan data, preprocessing data, pelatihan model, evaluasi kinerja model, dan desain aplikasi web. Hasil akhir dari penelitian ini diharapkan dapat membantu tenaga medis dalam pengambilan keputusan yang lebih cepat dan tepat, memungkinkan intervensi dini dan pengelolaan yang lebih baik terhadap pasien berisiko. Selain itu, dengan adanya sistem ini, diharapkan dapat terjadi pengurangan beban kerja tenaga medis dan peningkatan efisiensi dalam proses screening

Tinjauan Pustaka

a. Diabetes Gestasional

Diabetes melitus adalah kelainan metabolik yang disebabkan oleh beberapa faktor penentu, termasuk kurangnya insulin atau ketidakmampuan tubuh untuk memanfaatkan insulin, juga dikenal sebagai insulin resistance. Gejalanya termasuk hiperglikemia kronis dan gangguan metabolisme karbohidrat, lemak, dan protein karena sekresi hormon insulin yang rendah atau aktivitas insulin yang rendah.[3] Salah satu jenis diabetes melitus adalah Diabetes mellitus gestasional yang terjadi pada ibu hamil.

Diabetes mellitus gestasional (DMG) adalah gangguan toleransi karbohidrat yang muncul selama kehamilan atau diketahui sebagai premakalis. Kondisi ini biasanya muncul selama 24 minggu kehamilan dan kembali normal setelah persalinan, menimbulkan masalah besar bagi masyarakat karena berdampak langsung pada kesehatan ibu dan janin[2].

Seorang Ibu yang menderita diabetes melitus gestasional memiliki risiko lebih tinggi untuk mengalami gangguan hipertensi (preeklampsia), janin besar, keguguran, partus lama, bayi lahir prematur, dan persalinan secara sectio caesarea. Selain itu, ibu yang menderita diabetes gestasional memiliki risiko lebih tinggi untuk mengalami penyakit kardiovaskular pada wanita yang memiliki riwayat diabetes melitus tipe II, serta lebih sedikit waktu yang diberikan untuk melahirkan bayi.[1]

b. Random Forest

Untuk klasifikasi dan regresi, Random Forest adalah sebuah kumpulan metode pembelajaran yang dibangun dan dikombinasikan dengan pohon keputusan sebagai basis klasifikasi. Tiga komponen penting dari metode Random Forest adalah sebagai berikut: (1) mengambil sampel bootstrap untuk membuat pohon prediksi; (2) memprediksi masing-masing pohon keputusan dengan prediktor acak; dan (3) membuat prediksi dengan meng-gabungkan hasil dari setiap pohon keputusan dengan suara mayoritas untuk klasifikasi atau rata-rata untuk regresi[4]. Selain itu, Random Forest memiliki proses seleksi fitur yang memungkinkannya mengambil fitur terbaik, yang dapat meningkatkan kinerjanya terhadap model klasifikasi.

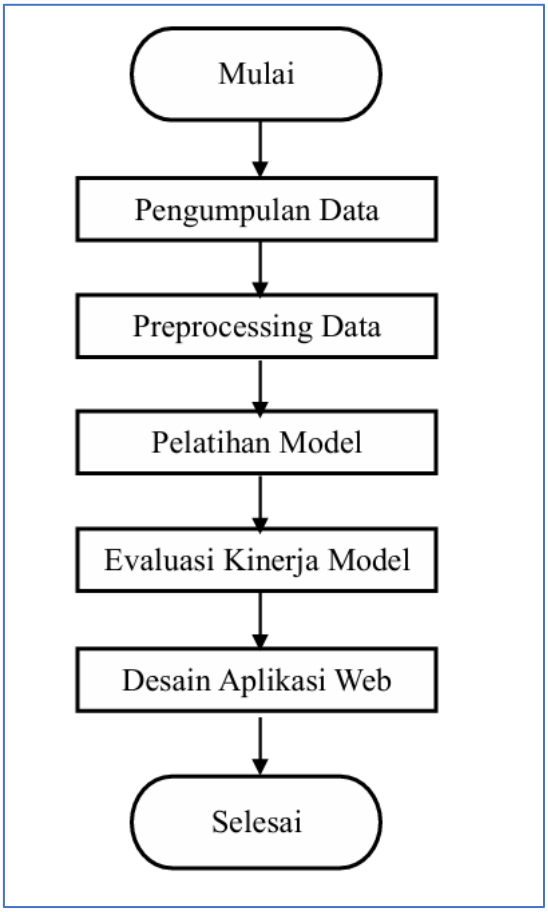
Dengan bantuan proses seleksi fitur ini, Random Forest dapat bekerja secara efektif dengan data besar dengan parameter yang kompleks[5].

c. Machine Learning

Cabang ilmu Kecerdasan Buatan (Intelligence) yang dikenal sebagai Machine Learning menggunakan bahasa pemrograman untuk membuat komputer berperilaku cerdas serupa dengan manusia. Apabila data tersedia, pengajaran mesin berfungsi untuk menemukan pola tertentu untuk digunakan dalam analisis kumpulan data besar. Pendidikan mesin terdiri dari pengawasan, pengawasan tanpa pengawasan, dan penguatan.[6]

Metodologi

Penelitian ini bertujuan untuk melakukan deteksi dini penyakit diabetes gestasional pada ibu hamil menggunakan Random Forest berbasis Machine Learning. Metodologi yang digunakan dalam penelitian ini meliputi beberapa tahapan : Pengumpulan data, preprocessing data, pelatihan model, evaluasi kinerja model dan desain aplikasi web. Berikut flowchart dari metodologi yang akan dilakukan:



Gambar 1 Flowchart Metodologi

a. Pengumpulan Data

Studi literatur saat ini dilakukan sebagai teknik pengumpulan data untuk menemukan teori-teori baru di bidang pembuatan sistem. Ini termasuk membaca literatur yang berkaitan dengan topik penelitian dan mengeksplorasi internet, termasuk mencari jurnal-jurnal yang relevan dan mengunjungi situs web yang terkait dengan topik penelitian. Penelitian ini mengumpulkan data dari ibu hamil yang menderita diabetes dengan berbagai parameter penting. Parameter tersebut termasuk jumlah kehamilan, kadar glukosa, tekanan darah, ketebalan kulit, kadar insulin, indeks massa tubuh, fungsi pedigree diabetes, usia, dan indeks glukosa tubuh. Data ini diharapkan dapat memberikan gambaran yang luas tentang risiko yang terkait dengan diabetes pada ibu hamil.

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	2	138	62	35	0	33.6	0.127	47	1
1	0	84	82	31	125	38.2	0.233	23	0
2	0	145	0	0	0	44.2	0.630	31	1
3	0	135	68	42	250	42.3	0.365	24	1
4	1	139	62	41	480	40.7	0.536	21	0
5	0	173	78	32	265	46.5	1.159	58	0
6	4	99	72	17	0	25.6	0.294	28	0
7	8	194	80	0	0	26.1	0.551	67	0
8	2	83	65	28	66	36.8	0.629	24	0
9	2	89	90	30	0	33.5	0.292	42	0
10	4	99	68	38	0	32.8	0.145	33	0
11	4	125	70	18	122	28.9	1.144	45	1
12	3	80	0	0	0	0.0	0.174	22	0
13	6	166	74	0	0	26.6	0.304	66	0
14	5	110	68	0	0	26.0	0.292	30	0
15	2	81	72	15	76	30.1	0.547	25	0
16	7	195	70	33	145	25.1	0.163	55	1
17	6	154	74	32	193	29.3	0.839	39	0
18	2	117	90	19	71	25.2	0.313	21	0
19	3	84	72	32	0	37.2	0.267	28	0

Gambar 2 Dataset

b. Preprocessing Data

Dalam preprocessing data ini, pertama-tama nilai nol pada beberapa fitur penting seperti Glucose, BloodPressure, SkinThickness, Insulin, dan BMI digantikan dengan NaN untuk menandai data yang hilang. Kemudian, jumlah NaN dihitung untuk setiap kolom tersebut. Untuk menangani nilai NaN, nilai-nilai tersebut diisi dengan rata-rata kolom yang bersangkutan, sehingga data menjadi lengkap dan siap untuk analisis lebih lanjut. Selanjutnya, dilakukan normalisasi fitur dengan menggunakan MinMaxScaler untuk menskalakan semua fitur ke rentang 0 hingga 1, yang kemudian disimpan dalam DataFrame baru. Fitur-fitur yang dipilih untuk model adalah Glucose, Insulin, BMI, dan Age, sedangkan variabel target adalah

Outcome. Data dibagi menjadi set pelatihan dan pengujian dengan proporsi 80:20, menggunakan metode stratifikasi untuk memastikan distribusi kelas yang seimbang.

```
dataset_new = dataset

# Replacing zero values with NaN
dataset_new[["Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI"]] = dataset_new[["Glucose", "BloodPressure", "SkinThickness", "Insulin", "BMI"]].replace(0, np.NaN)

# Count of NaN
dataset_new.isnull().sum()

Pregnancies                 0
Glucose                     5
BloodPressure              35
SkinThickness             227
Insulin                   374
BMI                        11
DiabetesPedigreeFunction    0
Age                         0
Outcome                     0
dtype: int64

# Replacing NaN with mean values
dataset_new["Glucose"].fillna(dataset_new["Glucose"].mean(), inplace = True)
dataset_new["BloodPressure"].fillna(dataset_new["BloodPressure"].mean(), inplace = True)
dataset_new["SkinThickness"].fillna(dataset_new["SkinThickness"].mean(), inplace = True)
dataset_new["Insulin"].fillna(dataset_new["Insulin"].mean(), inplace = True)
dataset_new["BMI"].fillna(dataset_new["BMI"].mean(), inplace = True)
```

Gambar 3 Preprocessing data

```
# Statistical summary
dataset_new.describe().T

Show hidden output

# Feature scaling using MinMaxScaler
from sklearn.preprocessing import MinMaxScaler
sc = MinMaxScaler(feature_range = (0, 1))
dataset_scaled = sc.fit_transform(dataset_new)

dataset_scaled = pd.DataFrame(dataset_scaled)

# Selecting features - [Glucose, Insulin, BMI, Age]
X = dataset_scaled.iloc[:, [1, 4, 5, 7]].values
Y = dataset_scaled.iloc[:, 8].values

# Splitting X and Y
from sklearn.model_selection import train_test_split
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size = 0.20, random_state = 42, stratify = dataset_new['Outcome'] )

# Checking dimensions
print("X_train shape:", X_train.shape)
print("X_test shape:", X_test.shape)
print("Y_train shape:", Y_train.shape)
print("Y_test shape:", Y_test.shape)

X_train shape: (614, 4)
X_test shape: (154, 4)
Y_train shape: (614,)
Y_test shape: (154,)
```

Gambar 4 Preprocessing data – split data latih dan uji

c. Pelatihan Model

Model Random Forest diterapkan dengan parameter n_estimators sebesar 11 dan kriteria 'entropy' untuk membagi node. Model ini dilatih menggunakan data pelatihan yang tersedia, yaitu X_train dan Y_train. Setelah proses pelatihan selesai, model digunakan untuk memprediksi hasil pada data uji, yang memungkinkan evaluasi performa model dengan berbagai metrik seperti akurasi, precision, recall, dan F1-score. Dengan hanya menggunakan Random Forest, fokus utama adalah pada pengaturan parameter yang optimal dan analisis performa model yang digunakan

d. Evaluasi Kinerja Model

Evaluasi kinerja model Random Forest dilakukan dengan menggunakan data uji setelah model dilatih. Metrik utama yang digunakan untuk menilai kinerja adalah akurasi, yang mengukur seberapa sering model memberikan prediksi yang benar dibandingkan dengan total prediksi. Selain akurasi, metrik lain seperti precision, recall, dan F1-score juga dievaluasi untuk memberikan gambaran yang lebih komprehensif mengenai performa model. Precision mengukur seberapa akurat prediksi positif model, recall mengukur kemampuan model dalam menangkap semua kasus positif yang ada, dan F1-score memberikan keseimbangan antara precision dan recall. Analisis hasil dari metrik ini memungkinkan penilaian efektivitas model Random Forest dalam mendeteksi risiko penyakit diabetes, serta memberikan insight tentang kekuatan dan kelemahan model. Evaluasi ini memastikan bahwa model tidak hanya efektif dalam klasifikasi, tetapi juga dapat diandalkan untuk aplikasi klinis dan deteksi dini.

e. Desain Aplikasi Web

Desain aplikasi web untuk model deteksi risiko diabetes dibuat agar pengguna dapat memasukkan data kesehatan pribadi seperti jumlah kehamilan, kadar glukosa, tekanan darah, ketebalan kulit, kadar insulin, indeks massa tubuh (BMI), fungsi pedigree diabetes, dan usia kedalam kolom input diaplikasi web. Setelah data dimasukkan, aplikasi menggunakan model Random Forest yang telah dilatih untuk memprediksi risiko diabetes. Hasil prediksi akan ditampilkan kepada pengguna dalam bentuk yang mudah dipahami, menunjukkan apakah mereka berisiko terdiagnosis diabetes atau tidak.

Hasil dan Pembahasan

a. Analisis Model

Hasil analisis model ketika dijalankan menunjukkan bahwa algoritma Random Forest mampu mendeteksi risiko diabetes mellitus pada ibu hamil dengan akurasi akhir sebesar 98%. Model ini memiliki kemampuan yang luar biasa untuk menangani data yang kompleks dan tidak seimbang dan menunjukkan kinerja yang luar biasa dalam klasifikasi. Jumlah akurasi yang sangat tinggi ini menunjukkan bahwa Random Forest adalah alat yang efektif untuk membantu diagnosis risiko diabetes dan memberikan prediksi yang andal dan konsisten. Hasil ini tidak hanya menunjukkan bahwa Random Forest cukup andal untuk memproses dan menganalisis data kesehatan, tetapi juga menunjukkan bahwa itu dapat digunakan untuk sistem deteksi dini, yang dapat sangat membantu mengelola risiko kesehatan ibu hamil.

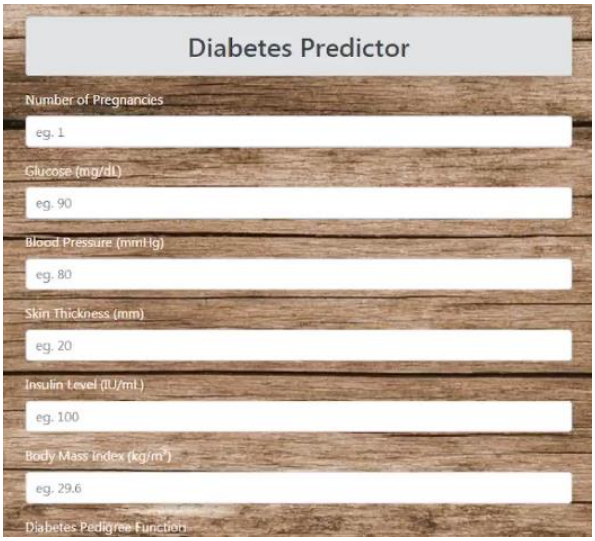
```
#Random forest
from sklearn.ensemble import RandomForestClassifier

clf=RandomForestClassifier()
clf.fit(train_X,train_y)

y_pred=clf.predict(test_X)
#find accuracy
ac=accuracy_score(test_y,y_pred)
acc.append(ac)

#find the ROC_AOC curve
rc=roc_auc_score(test_y,y_pred)
roc.append(rc)
print("\nAccuracy {0} ROC {1}".format(ac,rc))

#cross val score
result=cross_validate(clf,train_X,train_y,scoring=scoring,cv=10)
display_result(result)

#display predicted values uncomment below line
#pd.DataFrame(data={'Actual':test_y,'Predicted':y_pred}).head()
```

```
Accuracy 0.9879154078549849 ROC 0.9859793290410032
TP:  [40 39 39 38 40 41 40 37 38 38]
TN:  [86 90 90 90 89 89 89 89 91 86]
FN:  [2 2 2 3 1 0 1 4 3 4]
FP:  [5 1 1 1 2 2 2 2 0 4]
```

Gambar 5 Akurasi model

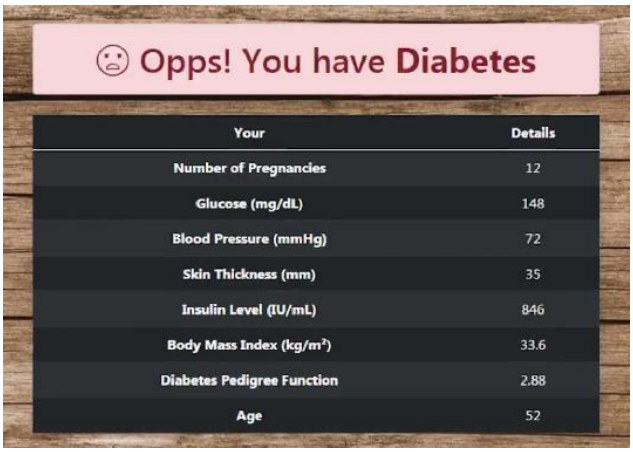
b. Tampilan Aplikasi Web

Terdapat dua halaman utama dalam aplikasi web ini yang dimaksudkan untuk mendukung deteksi dini risiko diabetes mellitus pada ibu hamil. Halaman pertama adalah halaman input, di mana ibu hamil dapat memasukkan riwayat medis, tekanan darah, gula darah, dan parameter kesehatan lainnya. Pengguna dapat mengirimkan data untuk diproses oleh model setelah data dimasukkan. Halaman kedua adalah halaman hasil, yang men-unjukkan apakah ibu hamil memiliki risiko diabetes atau tidak. Dengan menggunakan data yang telah dimasukkan, sistem memberikan hasil prediksi yang jelas dan mudah dipahami tentang kemungkinan risiko diabetes mellitus. Aplikasi ini bertujuan untuk membantu menjaga kesehatan ibu hamil dengan membantu mengidentifikasi masalah dengan cepat melalui antarmuka yang mudah dipahami dan mudah digunakan. Berikut adalah tampilan dari aplikasi web:



Gambar 6 Input data

Gambar 7 Negatif Diabetes



Gambar 8 Positif Diabetes

Simpulan

Berdasarkan analisis dan implementasi model dalam penelitian ini, dapat disimpulkan bahwa model Random Forest terbukti efektif dalam mendeteksi risiko diabetes mellitus pada ibu hamil dengan mencapai akurasi yang tinggi. Penelitian ini menunjukkan bahwa Random Forest mampu memberikan hasil prediksi yang andal dan konsisten, dengan akurasi akhir mencapai 98%. Evaluasi kinerja model mengungkapkan kelebihan utama dari Random Forest, seperti kemampuan untuk mengatasi overfitting dan menangani data yang tidak seimbang dengan baik. Namun, model ini juga memiliki kelemahan, termasuk kompleksitas yang tinggi dan keterbatasan dalam interpretabilitas serta potensi dampak ketidakseimbangan data yang ekstrem. Penggunaan model Random Forest dalam aplikasi web untuk deteksi dini risiko diabetes memberikan manfaat yang signifikan, meskipun perlu diingat bahwa akurasi model dapat bervariasi tergantung pada kualitas dan keseimbangan dataset yang digunakan. Dengan demikian, penelitian ini memberikan kontribusi penting dalam pengembangan alat bantu deteksi dini yang berguna untuk ibu hamil, sambil menyoroti kebutuhan untuk pemantauan dan penyesuaian lebih lanjut dalam implementasi praktis.

Daftar Referensi

[1] Rianti Nurpalah, Meti Kusmiati, Meri Meri, Hendro Kasmanto, dan Dina Ferdiani, "Deteksi Dini Diabetes Melitus Gestasional (Dmg) Melalui Pemeriksaan Glukosa Darah Sebagai Upaya Pencegahan Komplikasi Pada Ibu Hamil," *J-ABDI J. Pengabdi. Kpd. Masy.*, vol. 2, no. 9, hlm. 6425–6432, Jan 2023, doi: 10.53625/jabdi.v2i9.4880.

[2] A. Umiyah, "Analisis kejadian diabetes melitus gestasional di wilayah kerja Puskesmas Banyuputih," *J. Ilmu Kesehat. Bhakti Husada Health Sci. J.*, vol. 14, no. 02, hlm. 317–323, Des 2023, doi: 10.34305/jikbh.v14i02.824.

[3] A. Pinandito, S. A. Wicaksono, dan S. H. Wijoyo, "Implementasi Machine Learning Dalam Deteksi Risiko Tinggi Diabetes Melitus Pada Kehamilan".

[4] A. Primajaya dan B. N. Sari, "Random Forest Algorithm for Prediction of Precipitation," *Indones. J. Artif. Intell. Data Min.*, vol. 1, no. 1, hlm. 27, Mar 2018, doi: 10.24014/ijaidm.v1i1.4903.

[5] R. Supriyadi, W. Gata, N. Maulidah, dan A. Fauzi, "Penerapan Algoritma Random Forest Untuk Menentukan Kualitas Anggur Merah," *E-Bisnis J. Ilm. Ekon. Dan Bisnis*, vol. 13, no. 2, hlm. 67–75, Nov 2020, doi: 10.51903/e-bisnis.v13i2.247.

[6] P. P. P. A. N. W. F. I. R.H. Zer, B. H. Hayadi, dan A. R. Damanik, "Pendekatan Machine Learning Menggunakan Algoritma C4.5 Berbasis Pso Dalam Analisa Pemahaman Pemrograman Website," *J. Inform. Dan Tek. Elektro Terap.*, vol. 10, no. 3, Agu 2022, doi: 10.23960/jitet.v10i3.2700.