

# Komparasi Perfoma Model Algoritma Random Forest dan K-Nearest Neighbors Pada Diagnosis Gestational Diabetes Mellitus (GDM)

Dede Brahma Arianto<sup>1\*</sup>, Iksan Sapani<sup>2</sup>, Muhamad Wahyu Ramdan<sup>3</sup>, Aldi<sup>4</sup>

<sup>1</sup> Program Studi Informatika, Fakultas Sains dan Teknik, Universitas Faletehan, dedebrahma@uf.ac.id

<sup>2</sup> Program Studi Informatika, Fakultas Sains dan Teknik, Universitas Faletehan, iksansapani@gmail.com

<sup>3</sup> Program Studi Informatika, Fakultas Sains dan Teknik, Universitas Faletehan, wahyuramdan04@gmail.com

<sup>4</sup> Program Studi Informatika, Fakultas Sains dan Teknik, Universitas Faletehan, aldi010403@gmail.com

\* Corresponding Author : dedebrahma@uf.ac.id

## Abstrak

Diabetes melitus gestasional merupakan suatu kondisi dimana kadar glukosa darah berada di atas normal selama masa kehamilan. Kondisi ini dapat berdampak pada ibu dan bayi yang dikandungnya, sehingga deteksi dini terhadap GDM menjadi sangat penting. Penelitian ini bertujuan untuk membandingkan penggunaan algoritma machine learning, K-Nearest Neighbors (KNN) dan Random Forest, dalam mengklasifikasi dan memprediksi GDM. Data yang digunakan dalam penelitian ini berupa rekam medis dari 133 ibu hamil yang diobservasi sejak minggu ke-20 kehamilan hingga melahirkan. Dataset terdiri dari 14 variabel, dengan variabel dm gestasional digunakan sebagai variabel dependen. Perbandingan kinerja didasarkan pada precision, recall, f1-score, dan accuracy. Hasil penelitian menunjukkan bahwa akurasi algoritma KNN sebesar 93%, sedangkan akurasi algoritma Random Forest sebesar 96%. Dari hasil tersebut, dapat disimpulkan bahwa algoritma Random Forest lebih unggul dibandingkan dengan algoritma KNN dalam melakukan klasifikasi dan prediksi diabetes melitus gestasional pada ibu hamil.

Kata kunci: machine learning, k-nearest neighbors, random forest, diabetes mellitus gestasional.

## *Comparative Performance Models of Random Forest and K-Nearest Neighbours Algorithms in The Diagnosis Of Gestational Diabetes Mellitus*

### *Abstract*

*Gestational diabetes mellitus (GDM) is a condition where the blood glucose level is above normal during pregnancy. This condition can have an impact on both the mother and the unborn baby, making early detection of GDM crucial. This study aims to compare the use of machine learning algorithms, K-Nearest Neighbors (KNN) and Random Forest, in classifying and predicting GDM. The data used in this study consists of medical records from 133 pregnant women observed from the 20th week of pregnancy until delivery. The dataset comprises 14 variables, with the gestational dm variable used as the dependent variable. Performance comparison is based on precision, recall, f1-score, and accuracy. The results show that the accuracy of the KNN algorithm is 93%, while the accuracy of the Random Forest algorithm is 96%. With this difference, it can be concluded that the K-Nearest Neighbors algorithm is superior to the Random Forest algorithm in the classification and prediction of gestational diabetes mellitus in pregnant women.*

*Keywords: machine learning, k-nearest neighbors, random forest, diabetes mellitus gestasional.*

## PENDAHULUAN

*Gestational Diabetes Mellitus* (GDM) adalah salah satu jenis diabetes yang dapat terjadi selama kehamilan dimana nilai glukosa darah berada di atas normal tetapi masih di bawah nilai diagnostik diabetes (WHO, 2023). Kondisi ini tidak disebabkan oleh kekurangan insulin, melainkan terjadi ketika hormon yang diproduksi oleh plasenta mencegah tubuh menggunakan insulin dengan efektif sehingga mengakibatkan penumpukan glukosa dalam darah dan tidak diserap oleh sel. Keadaan ini dikenal sebagai resistensi insulin (Rodriguez, 2022).

Pada kebanyakan kasus, resistensi insulin selama kehamilan akan hilang setelah melahirkan, tetapi bagi wanita yang sudah mengalami masalah dalam penggunaan insulin oleh tubuh bahkan sebelum kehamilan terjadi, memiliki tingkat risiko lebih tinggi menderita GDM. Bertambahnya kasus obesitas diseluruh dunia, berdampak pula pada peningkatan jumlah wanita hamil yang didiagnosis menderita GDM (Wenrui, 2022). Mereka memiliki tingkat risiko lebih tinggi untuk mengalami komplikasi baik selama kehamilan maupun saat melahirkan. Preeklampsia, persalinan premature, dan makrosomia pada bayi merupakan beberapa risiko komplikasi yang dapat terjadi. Selain itu, para wanita tersebut dan mungkin anak-anak mereka juga berisiko lebih tinggi terkena diabetes tipe 2 di masa depan (WHO, 2023).

Menurut *National Center for Health Statistics* (NCHS), dari tahun 2011 hingga 2019 di US, tercatat bahwa angka GDM meningkat secara signifikan dari 47,6 menjadi 63,5 per 1000 kelahiran hidup dengan rata-rata perubahan tahunan sebesar 3,7% per tahun (Shah, NS, dkk, 2022). Sedangkan berdasarkan *International Diabetes Federation*, melaporkan pada tahun 2021 prevalensi GDM di dunia tercatat sebesar 16,7% (IDF, 2021).

*Diabetes mellitus gestasional* adalah diabetes mellitus yang timbul selama kehamilan (Catalano PM, 2014). Pada kehamilan normal, wanita mengalami penurunan sensitivitas insulin sebesar 60%. Pada wanita dengan fungsi pankreas normal, dapat meningkatkan produksi insulin yang cukup untuk mengimbangi resistensi insulin

fisiologis yang terjadi selama kehamilan. Diabetes gestasional berkembang ketika pankreas tidak mampu mengimbangi resistensi insulin karena kekurangan sel beta fungsional. Patofisiologi resistensi insulin dan gangguan sekresi insulin pada pasien diabetes gestasional mirip dengan yang terjadi pada pasien diabetes tipe 2 (Johns EC, 2018). Wanita yang memiliki kadar glukosa yang normal sebelum kehamilan tetapi mengalami diabetes gestasional pada akhir kehamilan cenderung mengalami disfungsi metabolik subklinis yang belum terdeteksi sebelum pembuahan, dibandingkan dengan wanita yang memiliki toleransi glukosa yang normal selama kehamilan (Catalano PM, 2014).

Saat ini penyebab GDM masih belum diketahui dengan pasti, namun terdapat beberapa teori yang menjelaskan mengapa kondisi ini terjadi. Selama kehamilan terjadi, plasenta berperan dalam menyuplai nutrisi dan air ke janin serta memproduksi hormon-hormon tertentu untuk menjaga kehamilan. Beberapa hormon tersebut seperti estrogen, kortisol, dan laktogen dapat menyebabkan efek kontra insulin dimana kinerja insulin jadi terhambat. Pada umumnya, kondisi ini terjadi sekitar 20-24 minggu usia kehamilan. Seiring dengan perkembangan plasenta, hormon ini semakin banyak diproduksi dan meningkatkan risiko resistensi insulin. Pada umumnya, pankreas dapat memproduksi lebih banyak insulin untuk mengatasi hal tersebut, tetapi jika produksi insulin tidak mencukupi untuk mengatasi efek hormon plasenta, maka akan terjadi diabetes gestasional.

Dengan adanya dampak yang diakibatkan oleh diabetes gestasional baik pada ibu maupun bayi, perlu dilakukan upaya untuk mengidentifikasi GDM semenjak dini. Deteksi sejak dini diharapkan dapat menurunkan resiko komplikasi baik pada ibu maupun pada bayi yang dilahirkan dimasa mendatang.

Dalam bidang kesehatan, pemanfaatan *machine learning* telah banyak digunakan untuk mendiagnosis suatu penyakit. *Machine learning* mempelajari data dan memungkinkan komputer untuk melakukan klasifikasi dan prediksi berdasarkan data tersebut. Penggunaan *machine learning* dalam bidang medis dapat menekan biaya dan mempercepat proses diagnosis. Deteksi dini penyakit umumnya menggunakan teknik klasifikasi, yaitu dengan

mengenali pola atau model dari dataset untuk memprediksi apakah seseorang menderita penyakit atau tidak. Model ini dibangun berdasarkan analisis data *training* dan dapat digunakan untuk mengklasifikasikan dan memprediksi tren data di masa depan.

Akurasi dan tingkat performa model menjadi faktor penentu kualitas klasifikasi dan prediksi yang dihasilkan. Contoh metode klasifikasi yang banyak digunakan yaitu *Random Forest*. Chae Zahrah (2022) telah melakukan penelitian untuk membandingkan metode *Support Vector Machine (SVM)* dengan *Random Forest* untuk deteksi awal risiko diabetes melitus. Hasilnya diketahui bahwa metode *Random Forest* lebih unggul dengan tingkat akurasi 98.08%.

Dalam penelitian ini, akan dilakukan perbandingan performa *random forest* dengan metode lainnya yaitu metode *K-Nearest Neighbors (KNN)* dalam mendiagnosis GDM pada wanita hamil. Penelitian ini bertujuan untuk menentukan model terbaik dengan tingkat akurasi yang optimal sehingga dapat digunakan untuk deteksi dini GDM pada wanita hamil secara lebih akurat.

*Random Forest* adalah salah satu algoritma machine learning yang umum digunakan dimana pertama kali diperkenalkan oleh Leo Breiman dan Adele Cutler. Algoritma ini merupakan pengembangan dari metode bagging Breiman, hanya saja pada *random forest* digunakan keacakan fitur yang bertujuan untuk membuat hutan pohon keputusan yang tidak berkorelasi. Keacakan fitur juga dapat disebut sebagai *random sub sampling* atau atau pemilihan  $m$  variable. *Random forest* menggabungkan output dari beberapa *decision tree* untuk mencapai hasil tunggal dimana hal ini dapat digunakan baik untuk klasifikasi maupun untuk regresi.

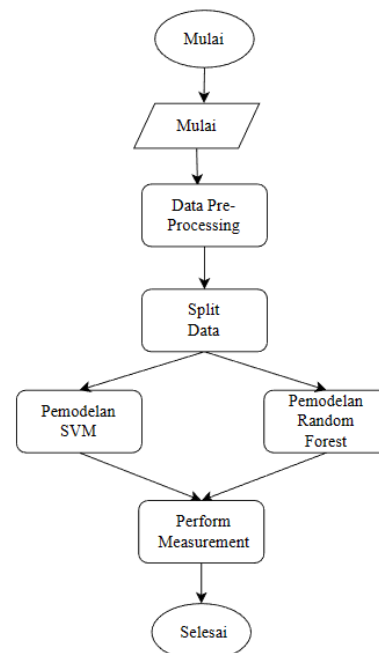
Algoritma *K-Nearest Neighbors (KNN)* merupakan salah satu metode *supervised learning* yang dapat digunakan untuk melakukan klasifikasi dan prediksi berdasarkan kedekatan jarak. Tujuan dari algoritma ini adalah untuk menentukan kelompok data baru berdasarkan kumpulan atribut dari data *training*. Data yang belum diketahui kelompoknya diolah dengan memilih tetangga dari sejumlah  $k$  (tetangga terdekat) yang berjarak paling dekat dari data tersebut. Dalam

algoritma klasifikasi ini, perhitungan jarak antara data satu dengan yang lain dibutuhkan.

## METODE PENELITIAN

Data yang digunakan merupakan data sekunder yang didapat dari hasil *study cohort* yang dilakukan oleh Rocha, A.d.S., dkk. Data terdiri atas 133 data pengamatan terhadap wanita hamil dengan 14 variabel yang terdiri atas 1 variabel *dependent (Y)* dan 13 variabel *independent (X)*.

Metode yang dipakai dalam penelitian ini yaitu algoritma KNN dan *Random Forest*. Analisis data dilakukan dengan menggunakan bahasa pemrograman *python* memanfaatkan *Google Colaboratory* sebagai alat bantu. Analisis dilakukan dengan prosedur penelitian sebagaimana pada Gambar 1.



Gambar 1. Metode penelitian

1. **Data Preprocessing**  
Data *preprocessing* merupakan langkah awal dalam data analisis dimana bertujuan untuk mempersiapkan data yang akan digunakan. Proses ini meliputi *profiling*, pembersihan, validasi, dan transformasi.
2. **Splitting Data Training dan Test**  
Dalam langkah ini, data dibagi menjadi data *training* dan data *testing*. Data *training* digunakan untuk membentuk model pada *machine learning* sedangkan data *testing*

digunakan untuk menguji performa model yang dihasilkan.

3. Proses Klasifikasi  
Tahap ini dilakukan dengan menerapkan algoritma *K-Nearest Neighbors* (KNN) dan *Random Forest* untuk mengklasifikasikan dan memprediksi diagnosis *gestational diabetes mellitus* (GDM) pada Wanita hamil.
4. Perbandingan *Performance Measure*

Untuk melihat hasil prediksi, digunakan *confusion matrix*

**Tabel 1.** *confusion matrix*

		PREDICTION	
		TP	FN
ACTUAL	FP		
	TN		

Perbandingan performa didasarkan atas nilai *accuracy*, *recall*, *precision*, dan *f1-score*.

- a. Accuracy, untuk mengetahui berapa kali model machine learning memprediksi dengan benar secara keseluruhan. Nilai ini dilihat dari presentase berasal dari jumlah prediksi benar (positif dan negatif) dari total jumlah prediksi yang dihasilkan.

$$accuracy = \frac{TP+TN}{TP+FP+TN+FN} \times 100\% \quad (1)$$

- b. Recall, untuk mengetahui berapa kali model mampu mengidentifikasi kategori true positive yaitu berdasarkan presentase dari prediksi benar positif (true positive/TP), dibandingkan dengan keseluruhan data positif.

$$recall = \frac{TP}{TP+FN} \times 100\% \quad (2)$$

- c. Precision, untuk mengetahui adalah seberapa baik model dalam memprediksi kategori tertentu yaitu berdasarkan presentase dari prediksi benar positif (true positive) dibandingkan seluruh hasil prediksi yang positif.

$$precision = \frac{TP}{TP+FP} \times 100\% \quad (3)$$

- d. F1 Score, skor ini memperhitungkan presisi dan recall dan didasarkan pada keseimbangan keduanya. Skor F1 berada pada rentang 0 hingga 1, dimana skor 1 berarti memiliki presisi dan recall yang sempurna, sedangkan skor F1 terendah adalah 0 yang berarti nilai recall atau presisi adalah nol.

$$F_1 = 2 \frac{precision \times recall}{precision+recall} = \frac{2TP}{TP+2(FP+FN)} \quad (4)$$

## PEMBAHASAN DAN HASIL

### Dataset

Pada penelitian ini, data yang digunakan berasal dari studi *cohort* yang dilakukan oleh Rocha, A.d.S., dkk dalam penelitian berjudul *Visceral Adipose Tissue Measurements During Pregnancy*. Data diperoleh dengan melakukan pengamatan pada wanita hamil sejak usia kehamilan 20 minggu hingga persalinan.

Sampel penelitian terdiri dari 154 wanita yang diamati dari bulan Oktober 2016 hingga Desember 2017 di Departemen Ultrasonografi di Pusat Kesehatan Pendidikan Murialdo, sebuah klinik yang menyediakan layanan pengobatan janin bagi pengguna Sistem Kesehatan Terpadu di kota Porto Alegre, Rio Grande do Sul, Brasil. Dari 154 wanita yang dipilih pada awalnya, sebanyak 21 (13%) wanita tidak dapat diamati hingga akhir, sehingga menghasilkan sampel akhir sebanyak 133 wanita.

Pengukuran jaringan *adiposea visceral* (VAT) dilakukan selama USG kebidanan rutin. Evaluasi biometrik juga dilakukan dimana informasi diperoleh dari perawatan prenatal. Hasil kehamilan termasuk *gestational diabetes mellitus* (GDM) diperoleh melalui evaluasi rekam medis. Data tersebut dikumpulkan sebagai bagian dari penelitian yang bertujuan untuk mengevaluasi apakah VAT dapat memprediksi GDM pada saat persalinan. Dari 133 data diketahui bahwa sebanyak 18 pasien menderita GDM sedangkan sisanya sebanyak 115 pasien tidak menderita GDM.

Variabel dalam dataset ini terdiri dari 14 variabel dimana variabel *gestational dm* menjadi variabel target atau *variable dependent*, untuk mengetahui lebih jelas

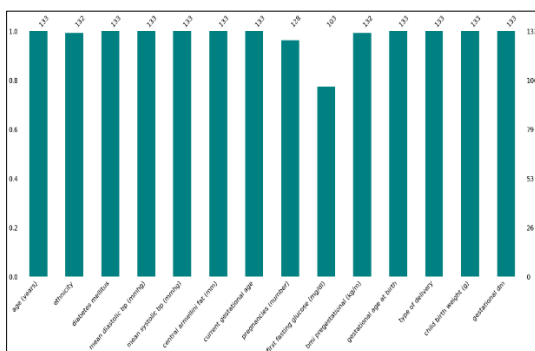
mengenai variable yang digunakan dapat dilihat pada tabel berikut:

**Tabel 2.** Variabel dalam dataset

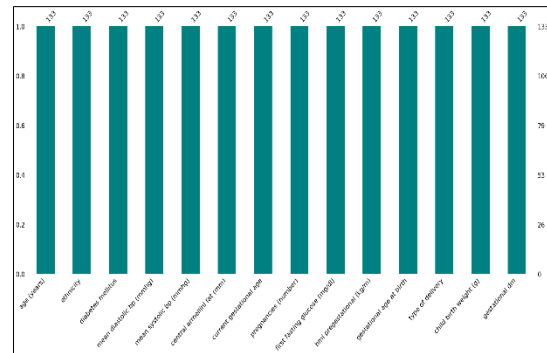
Variabel	Keterangan
Age (years)	Usia pasien
Ethnicity	Etnik pasien (0 = berkulit putih; 1 = bukan berkulit putih)
Diabetes mellitus	Riwayat diabetes mellitus
Mean diastolic BP	Tekanan darah diastolic
Mean systolic BP	Tekanan darah systolic
Central Armellini fat (mm)	<i>visceral adipose tissue</i>
Current Gestational age	Umur kehamilan
Pregnancies (number)	Jumlah kehamilan yang pernah dialami
First fasting glucose (mg/dl)	Kadar glukosa puasa pertama
BMI pregestational (kg/m)	Body Mass Indeks Ketika hamil
Gestational age at birth	Umur kehamilan Ketika melahirkan
Type of delivery	Jenis Persalinan (0=normal, 1 = operasi caesar)
Child birth weight (g)	Berat bayi Ketika lahir
Gestational DM	Gestational Diabetes Mellitus (0 = Tidak; 1 = Ya)

### Data Preprocessing

Tahap data *pre-processing* yang dilakukan berupa *data cleansing* dengan tujuan untuk membersihkan data dari *missing value* dan *outlier* yang terdapat dalam dataset. Hasilnya didapat sebagai berikut:



**Gambar 2.** Missing value sebelum pre-processing



**Gambar 3.** Missing value setelah pre-processing

Berdasarkan gambar 2, diketahui bahwa terdapat 4 variabel yang memiliki *missing value*, diantaranya yaitu variable *ethnicity* (1 data), *pregnancies (number)* (5 data), *first fasting glucose (mg/dl)* (30 data), dan *bmi pregestational (kg/m)* (1 data). Data yang hilang tersebut ditangani dengan mengisinya menggunakan nilai modus (untuk data kategorik) serta mean dan median (untuk data numerik). Hasil dari penanganan missing value ini, dapat dilihat pada gambar 3, terlihat bahwa sudah tidak terdapat missing value pada data.

Selain missing value, dilakukan pula *outlier checking*, dimana hasilnya terdapat outlier pada beberapa variable. Penanganan pada outlier tidak dilakukan karena bukan merupakan kesalahan input data. Outlier tersebut juga dapat menjadi *insight* dan menandakan bahwa terdapat perbedaan pada data pasien yang dapat mengindikasikan penyebab terjadinya *gestational diabetes mellitus (GDM)*.

### Split Data

Jumlah data yang digunakan dalam penelitian ini sebanyak 133 data. Sebelum dilakukan pemodelan, data perlu dibagi menjadi data *training* dan *testing*. Dalam penelitian ini, digunakan proporsi 80:20 dalam pembagian data, dimana 80% atau sebanyak 106 data digunakan sebagai data training dan 20% atau sebanyak 27 data digunakan sebagai data testing

### Perancangan Model

Pada penerapan algoritma *random forest*, langkah pertama yang dilakukan adalah menentukan parameter. *GridSearchCV* digunakan untuk menentukan parameter optimal yang akan digunakan dalam membangun model. tersebut dengan mencoba kombinasi parameter yang telah ditentukan. Hasilnya, diperoleh parameter optimal yang dapat digunakan yaitu nilai  $n\_estimator=50$ ,  $max\_depth = 3$  dan  $random\ state\ 42$ . Hasil klasifikasi dapat dilihat melalui *confusion matrix* seperti dibawah ini.

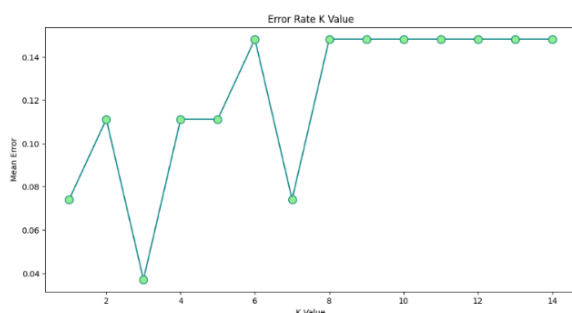
**Tabel 3.** Confusion Matrix Pada Data Training Model Random Forest

TRAINING			
Actual	Prediction		
	Negative	Positive	All
Negative	92	0	92
Positive	10	4	14
All	102	4	106

**Tabel 4.** Confusion Matrix Pada Data Testing Model Random Forest

TESTING			
Actual	Prediction		
	Negative	Positive	All
Negative	23	0	23
Positive	2	2	4
All	25	2	27

Selanjutnya dilakukan pemodelan dengan menggunakan algoritma K-Nearest Neighbors (KNN) dimana langkah pertama yang dilakukan yaitu mencari nilai  $k$  yang dapat memberikan performa optimal pada model. Patokan utama dalam menentukan nilai  $k$  adalah berdasarkan nilai mean error. Penentuan nilai  $k$  optimal dapat dilihat berdasarkan grafik pada gambar 5.



**Gambar 4.** Error Rate K-Value

Nilai  $k$  ditentukan berdasarkan nilai mean error terkecil seperti yang terlihat pada gambar 4. Namun untuk lebih tepat dilakukan iterasi guna mencari nilai  $k$  yang menghasilkan nilai skor maksimal baik pada data training maupun testing. Berdasarkan hasil iterasi, diperoleh nilai  $k$  optimal untuk data training adalah 1 dan untuk data testing adalah 3. Berdasarkan gambar juga dapat diketahui bahwa nilai  $k$  tersebut memiliki nilai mean error yang kecil.

Untuk melihat hasil klasifikasi menggunakan algoritma KNN dengan nilai  $k_{training}=1$  dan  $k_{testing}=3$  digunakan *confusion matrix* sebagai berikut:

**Tabel 5.** Confusion Matrix Pada Data Training Model KNN

TRAINING			
Actual	Prediction		
	Negative	Positive	All
Negative	92	0	92
Positive	0	14	14
All	92	14	106

**Tabel 6.** Confusion Matrix Pada Data Testing Model KNN

TESTING			
Actual	Prediction		
	Negative	Positive	All
Negative	23	0	23
Positive	1	3	4
All	24	3	27

### Performance Measure

Pengukuran performa digunakan berdasarkan nilai *precision*, *recall*, *f1-score*, dan *accuracy*. Tujuan dari perbandingan performa ini adalah untuk mengetahui metode manakah yang memiliki performa lebih baik dalam permasalahan klasifikasi penyakit *gestational diabetes melitus*. Adapun hasil perbandingan metode *K-Nearest Neighbors* dan *Random Forest* tertera pada tabel 7.

**Tabel 7.** Perbandingan Performa Metode KNN dan Random Forest

Metode	Precision	Recall	F1-Score	Training Accuracy	Test Accuracy
Random Forest	0.96	0.96	0.96	1.00	0.96
K-Nearest Neighbors (KNN)	0.93	0.93	0.92	0.91	0.93

Berdasarkan table diatas, diketahui bahwa metode *Random Forest* memiliki nilai *precision*, *recall*, *F1-Score* dan *Accuracy* lebih tinggi dibandingkan model *K-Nearest Neighbors* (KNN). Meskipun selisih dari kedua metode tersebut tidak berbeda secara signifikan, namun dapat disimpulkan bahwa metode *Random Forest* lebih baik dalam mendiagnosis *gestational diabetes mellitus* (GDM) dibanding metode *K-Nearest Neighbors* (KNN).

## SIMPULAN

Berdasarkan hasil penelitian perbandingan performa antara algoritma *Random Forest* dan *K-Nearest Neighbors* (KNN) terhadap 133 data pengamatan wanita hamil, didapat hasil bahwa metode *Random Forest* lebih unggul pada keseluruhan kategori penilaian performa (*accuracy*, *precision*, *recall*, dan *f1-score*) meskipun dengan selisih yang cukup kecil. Model *Random Forest* memiliki tingkat akurasi pada data test hingga 96% sedangkan model KNN dengan tingkat akurasi 93%. Sehingga pada klasifikasi dan prediksi *gestational diabetes mellitus*, metode *Random Forest* lebih baik dibanding *K-Nearest Neighbors* (KNN).

Penelitian lebih lanjut mengenai penerapan *machine learning* dalam diagnosis GDM pada wanita hamil masih penting untuk dilakukan. Faktor eksternal seperti genetik, gaya hidup, kebiasaan makan dan lingkungan social ekonomi dapat dipertimbangkan sebagai variabel yang dapat meningkatkan risiko GDM. Penambahan jumlah sampel penelitian juga dapat dilakukan agar dapat merepresentasikan populasi dengan lebih baik.

## REFERENSI

Arisandi, R. (2023) 'Perbandingan Model Klasifikasi Random Forest dengan Resampling dan Tanpa Resampling pada

Pasien Penderita Gagal Jantung', GAUSSIAN, 12(1), pp. 136-145, DOI: 10.14710/j.gauss.12.1.136-145

Becker, R., & Thrän, D. (2017). 'Completion of wind turbine data sets for wind integration studies applying Random Forest and K-Nearest Neighbors', Applied Energy, 208, 252–262. <https://doi.org/10.1016/j.apenergy.2017.10.044>

Catalano PM. (2014) 'Trying to understand gestational diabetes', Diabet Med, 31(3), pp 273-281.

Irfannandhy, R., Handoko, L. B., & Ariyanto, N. (2024). 'Analisis Performa Model Random Forest dan CatBoost dengan Teknik SMOTE dalam Prediksi Risiko Diabetes. Edumatic', Jurnal Pendidikan Informatika, 8(2), pp.714–723. <https://doi.org/10.29408/edumatic.v8i2.27990>

Junus, C. Z., Tarno, & Kartikasari, P. (2023) 'Klasifikasi Menggunakan Metode Support Vector Machine dan Random Forest Untuk Deteksi Awal Risiko Diabetes Melitus', GAUSSIAN, 11(3), pp. 386-396, DOI: 10.14710/j.gauss.11.3.386-396

N. M. Putry, (2022). "Komparasi Algoritma Knn Dan Naïve Bayes Untuk Klasifikasi Diagnosis Penyakit Diabetes Mellitus," EVOLUSI J. Sains dan Manaj., vol. 10, no. 1.

Prasetyo, R, P., Wahyuni, E, D., Kusumantara, P, M. (2024) 'Komparasi Performa Model Berbasis Algoritma Random Forest Dan Lightgbm Dalam Melakukan Klasifikasi Diabetes Melitus Gestasional', 12(3), pp.2582-2591

Ryfai, D. A., Hidayat, N., & Santoso, E. (2022) 'Klasifikasi Tingkat Resiko Serangan Penyakit Jantung menggunakan Metode K-Nearest Neighbor', J-PTIIK, 6(10), pp. 4701-4707.

- Sisodia, D., & Sisodia, D. S. (2018) 'Prediction of Diabetes using Classification Algorithms' *Procedia Computer Science*, 132, 1578–1585. Elsevier B.V. <https://doi.org/10.1016/j.procs.2018.05.122>
- Yolanda, V., Cholissodin, I., & Adikara, P. P. (2021). 'Klasifikasi Diagnosis Penyakit Diabetes Gestasional pada Ibu Hamil menggunakan Algoritme Neighbor Weighted K-Nearest Neighbor (NWKNN)', *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, 5(4), 1310–1321