Jurnal Majalah Ilmiah Informasi dan Teknologi Ilmiah (INTI)
ISSN 2301-9425 (Media Cetak), ISSN 2798-0340 (Media Online)
Volume 12, No 2, Februari 2025
Hal: 162-170

Prediksi Risiko Diabetes Menggunakan Random Forest dengan Evaluasi Klasifikasi dan Analisis Fitur

Berto Nadeak

Teknik Informatika,Fakultas Ilmu Komputer & Teknologi Informasi, Universitas Budi Darma, Medan, Indonesia
Email: Nadeak.Baru@Gmail.Com
Email penulis korespondensi: Nadeak.Baru@Gmail.Com

Abstrak- Prediksi risiko diabetes berbasis data merupakan salah satu penerapan pembelajaran mesin yang memerlukan evaluasi kinerja dan keterbatasan penggunaan secara jelas. Penelitian ini menyusun implementasi Random Forest untuk klasifikasi biner risiko diabetes sekaligus menganalisis kontribusi relatif atribut. Data yang digunakan berupa 768 record simulasi dengan delapan variabel numerik dan label dua kelas. Prosedur meliputi persiapan data, pembagian stratifikasi 80:20, pembentukan model dengan 180 pohon, evaluasi pada 154 data uji, serta repeated stratified five-fold cross-validation sebanyak dua pengulangan. Hasil pengujian menunjukkan accuracy 83,77%, precision 76,67%, recall 80,70%, F1-score 78,63%, dan ROC-AUC 87,39%. Pengujian silang menghasilkan rerata accuracy 80,33% dengan simpangan baku 3,17%, F1-score 73,06% dengan simpangan baku 3,82%, serta ROC-AUC 85,95% dengan simpangan baku 2,95%. Atribut Glucose memiliki nilai feature importance tertinggi sebesar 0,4495, diikuti BMI sebesar 0,1357 dan DiabetesPedigreeFunction sebesar 0,0876. Temuan ini menunjukkan bahwa alur pelatihan, evaluasi, dan penjelasan model dapat diverifikasi secara numerik. Namun, data berasal dari simulasi dan tidak membuktikan ketepatan klinis, generalisasi populasi, ataupun kelayakan penggunaan sebagai alat diagnosis. Pengujian pada data pasien terverifikasi, validasi eksternal, pemeriksaan kalibrasi, dan evaluasi risiko kesalahan diperlukan sebelum pemanfaatan di layanan kesehatan.

Kata Kunci: diabetes; Random Forest; klasifikasi risiko; pembelajaran mesin; ROC-AUC; feature importance

Abstract- Data-driven diabetes risk prediction is a potential application of machine learning, but its performance and limitations must be reported transparently. This study presents a Random Forest implementation for binary diabetes risk classification and an analysis of the relative contributions of predictive attributes. The verification dataset consists of 768 simulated records, eight numerical variables, and a binary outcome. The workflow includes data preparation, an 80:20 stratified split, training a 180-tree model, evaluation on 154 test records, and repeated stratified five-fold cross-validation with two repetitions. Test performance reached an accuracy of 83.77%, precision of 76.67%, recall of 80.70%, F1-score of 78.63%, and ROC-AUC of 87.39%. Cross-validation yielded mean accuracy of 80.33% (standard deviation 3.17%), F1-score of 73.06% (3.82%), and ROC-AUC of 85.95% (2.95%). Glucose showed the highest model-based feature importance (0.4495), followed by BMI (0.1357) and DiabetesPedigreeFunction (0.0876). These findings verify the computational workflow, evaluation protocol, and descriptive model interpretation under the simulated conditions. They do not establish clinical effectiveness, population generalizability, or diagnostic suitability because no authentic patient records were evaluated. The next stage requires ethically obtained, representative clinical data, independent validation, assessment of probability calibration, and prospective evaluation of the consequences of misclassification. Predictions must not be interpreted as medical diagnoses or as substitutes for qualified clinical assessment.

Keywords: diabetes; Random Forest; risk classification; machine learning; ROC-AUC; feature importance

1. PENDAHULUAN

Diabetes melitus merupakan kondisi kronis yang berkaitan dengan gangguan regulasi glukosa darah dan dapat menimbulkan komplikasi pada berbagai organ. Identifikasi faktor risiko dan pemeriksaan lebih awal memiliki nilai penting bagi layanan kesehatan. Namun, penilaian risiko memerlukan pemahaman terhadap konteks populasi, kualitas data, dan konsekuensi kesalahan klasifikasi. Penggunaan istilah prediksi risiko pada penelitian informatika tidak boleh disamakan dengan diagnosis diabetes yang memerlukan standar pemeriksaan dan interpretasi klinis.

Pendekatan pembelajaran mesin telah dimanfaatkan untuk membangun model prediktif dari berbagai karakteristik kesehatan. Penelitian Lai dkk. [1], Xie dkk. [2], Dinh dkk. [3], dan Mujumdar serta Vaidehi [4] menunjukkan bahwa data kesehatan dapat diolah menjadi model klasifikasi, dengan pilihan algoritma serta sumber data yang beragam. Kajian sistematis oleh Kaul dan Kumar [5] dan studi Muhammad dkk. [6] menguatkan bahwa metode pembelajaran terawasi perlu dievaluasi secara terukur sebelum digunakan pada situasi baru.

Random Forest merupakan teknik ensemble yang menggabungkan banyak decision tree melalui bootstrap sampling, pemilihan fitur secara acak, dan agregasi prediksi [16]. Dalam klasifikasi risiko, kemampuan

mempelajari hubungan nonlinear menjadi salah satu alasan penerapannya. Butt dkk. [7], Olisah dkk. [8], dan Chang dkk. [9] mendiskusikan beragam strategi pembelajaran mesin untuk prediksi diabetes. Namun, hasil performa antarpenelitian tidak dapat dibandingkan langsung apabila dataset, definisi label, penanganan nilai hilang, pembagian data, serta metrik penilaian berbeda.

Kodama dkk. [10] dan Olusanya dkk. [11] menyoroti variasi efektivitas algoritma pada studi diabetes. Penelitian Ginting dkk. [12], Sriyanto dan Supriyatna [13], Andi dkk. [14], serta Noviyanti dan Alamsyah [15] juga memberi konteks penerapan Random Forest dalam deteksi atau klasifikasi terkait diabetes. Literature tersebut mendukung relevansi topik, tetapi tidak otomatis membuktikan bahwa model baru akurat di setiap populasi. Oleh karena itu, keterlacakan konfigurasi, evaluasi data uji terpisah, dan analisis batas generalisasi perlu diperhatikan.

Pedoman TRIPOD+AI [17] menekankan transparansi pelaporan model prediksi yang dibangun menggunakan machine learning. Selain diskriminasi, kalibrasi probabilitas merupakan aspek penting untuk penggunaan klinis [18]. Pada data dengan proporsi kelas tidak seimbang, accuracy perlu dibaca bersama recall, F1-score, dan metrik lain yang merepresentasikan pola kesalahan [19]. Penilaian feature importance juga hanya menunjukkan kontribusi relatif menurut mekanisme model dan tidak langsung membuktikan hubungan kausal; teknik penjelasan yang lebih rinci dapat dipertimbangkan [20].

Kesenjangan yang dijawab pada artikel ini bersifat prosedural: mendokumentasikan konfigurasi Random Forest, menyajikan hasil pembagian data dan pengujian silang, serta membaca bobot kepentingan atribut secara konsisten dalam contoh verifikasi. Dataset yang tersedia dalam laporan sumber berupa data simulasi 768 record, bukan rekam medis asli. Dengan demikian, artikel ini tidak mengklaim inovasi prediksi klinis, tidak menilai kesiapan implementasi layanan kesehatan, dan tidak menyatakan model lebih baik daripada seluruh metode pembanding. Fokusnya adalah verifikasi alur eksperimen yang dapat ditelusuri.

Tujuan penelitian ialah (1) mengimplementasikan model Random Forest berbasis delapan atribut numerik; (2) melaporkan kinerja pada data uji terpisah menggunakan confusion matrix, accuracy, precision, recall, F1-score, dan ROC-AUC; (3) menilai kestabilan hasil melalui repeated stratified cross-validation; serta (4) mendeskripsikan urutan feature importance. Hasil dapat menjadi titik awal penelitian lanjutan menggunakan dataset klinis yang layak, tetapi hasil simulasi tidak dapat dipakai sebagai keputusan medis.

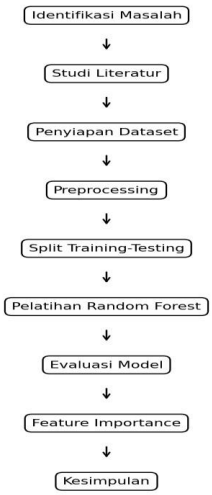
2. METODOLOGI PENELITIAN

2.1 Rancangan dan Sumber Data

Penelitian menggunakan rancangan eksperimen klasifikasi biner berbasis data simulasi. Total terdapat 768 record yang terdiri dari 485 kelas negatif/risiko rendah dan 283 kelas positif/risiko tinggi. Data ini dibuat untuk verifikasi algoritma dan tidak berasal dari pasien nyata atau penyedia layanan kesehatan. Delapan atribut yang digunakan adalah Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin, BMI, DiabetesPedigreeFunction, dan Age. Definisi kelas serta karakteristik data tidak boleh dianggap mewakili prevalensi atau pola penyakit pada populasi tertentu.

Pembagian data dilakukan secara stratified menjadi 614 data latih dan 154 data uji (sekitar 80:20). Pembagian stratifikasi menjaga proporsi kelas sedekat mungkin dengan proporsi awal. Data uji dipisahkan dari pelatihan sehingga metrik hold-out dapat dibaca sebagai pemeriksaan pada data yang tidak digunakan untuk membentuk pohon. Akan tetapi, naskah sumber tidak menyertakan berkas dataset mentah maupun skrip lengkap, sehingga reproduksi angka secara mandiri masih membutuhkan lampiran tambahan.

2.2 Tahapan Penelitian



Gambar 1. Tahapan Penelitian Random Forest

Tahap penelitian pada Gambar 1 mencakup identifikasi masalah, studi pustaka, persiapan dataset, preprocessing, pembagian training-testing, pelatihan Random Forest, evaluasi, analisis feature importance, dan penyusunan kesimpulan. Validasi kualitas data seharusnya mencakup pemeriksaan duplikasi, nilai kosong, nilai tidak wajar, dan satuan pengukuran. Setiap transformasi yang menggunakan statistik data harus dipelajari hanya dari training set agar tidak terjadi data leakage. Dalam naskah sumber, detail distribusi dan pembersihan data tidak cukup untuk diverifikasi; aspek tersebut merupakan batas reproduksibilitas yang perlu diperbaiki.

2.3 Konfigurasi Random Forest

Random Forest membangun himpunan pohon berdasarkan variasi sampel dan subset fitur. Prediksi akhir diperoleh melalui voting antar pohon, sedangkan estimasi probabilitas biasanya berasal dari agregasi keluaran pohon. Konfigurasi pada laporan dipertahankan tanpa optimasi ulang, sehingga nilai hasil tidak diperlakukan sebagai estimasi dari pencarian hyperparameter yang luas.

Parameter	Nilai
n_estimators	180
max_depth	9
min_samples_leaf	2
max_features	sqrt
class_weight	balanced_subsample
random_state	42

Jumlah 180 pohon dipakai untuk membentuk ensemble, max_depth membatasi kompleksitas, min_samples_leaf mengurangi pecahan cabang yang sangat kecil, dan max_features=sqrt menerapkan pemilihan fitur acak di tiap percabangan. class_weight=balanced_subsample menyesuaikan bobot kelas pada sampel bootstrap. random_state=42 mendokumentasikan sumber ketetapan bilangan acak. Konfigurasi tersebut berasal dari laporan sumber dan belum menunjukkan hasil pencarian parameter otomatis.

2.4 Evaluasi Klasifikasi

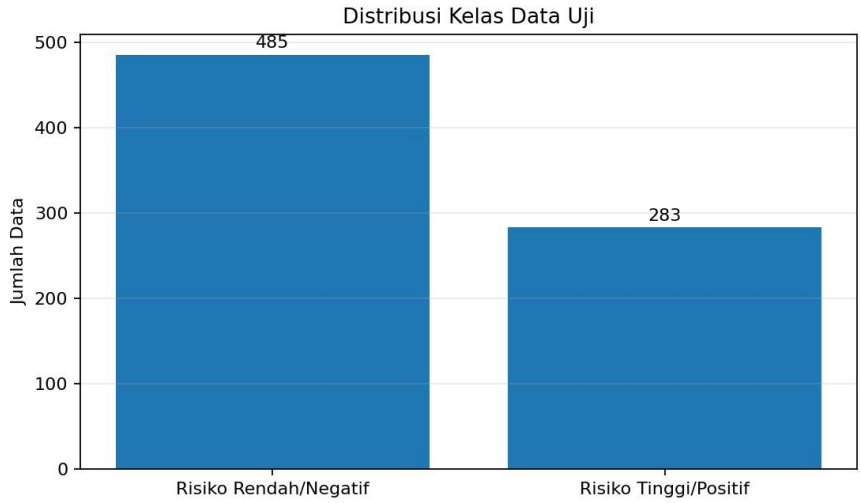
Confusion matrix membedakan true positive (TP), true negative (TN), false positive (FP), serta false negative (FN). Kinerja dievaluasi menggunakan accuracy=(TP+TN)/(TP+TN+FP+FN), precision=TP/(TP+FP), recall=TP/(TP+FN), dan F1-score=2×precision×recall/(precision+recall). ROC-AUC digunakan sebagai ringkasan kemampuan diskriminasi pada berbagai ambang klasifikasi. Nilai recall perlu ditafsirkan bersama false negative karena individu yang diprediksi negatif secara keliru dapat luput dari pemeriksaan lanjutan apabila sistem disalahgunakan sebagai alat diagnosis.

Sebagai pemeriksaan kestabilan, dilakukan repeated stratified five-fold cross-validation dengan dua pengulangan. Rata-rata dan simpangan baku antarfold dilaporkan untuk accuracy, F1-score, dan ROC-AUC. Laporan sumber tidak memberikan skor per fold dan rincian apakah validasi silang hanya menggunakan data training atau keseluruhan dataset; oleh karena itu belum dapat dipastikan kemandirian estimasi cross-validation terhadap data uji akhir. Desain penelitian lanjutan sebaiknya menggunakan nested cross-validation bila hyperparameter diseleksi dari data.

3. HASIL DAN PEMBAHASAN

3.1 Karakteristik Data Simulasi

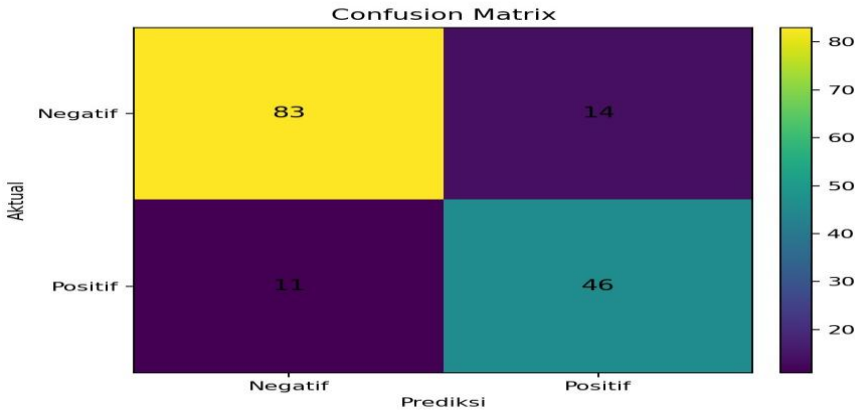
Distribusi data simulasi terdiri dari 485 kategori negatif (63,2%) dan 283 kategori positif (36,8%). Proporsi ini menunjukkan ketidakseimbangan kelas moderat. Stratifikasi bertujuan agar pemisahan training dan testing mempertahankan proporsi tersebut. Distribusi label penting karena model yang cenderung memprediksi kelas mayoritas bisa memperoleh accuracy yang tampak baik, tetapi kurang peka dalam mengenali kelas positif.



Gambar 2. Distribusi Kelas pada 768 Data Simulasi

3.2 Hasil Pengujian pada Data Testing

Pada 154 record data testing diperoleh 83 true negative, 14 false positive, 11 false negative, dan 46 true positive. Jumlah keempat sel confusion matrix sama dengan ukuran testing set, yaitu 154. Dengan demikian, jumlah prediksi benar sebanyak 129 dan prediksi salah sebanyak 25. Pembacaan sel dilakukan dengan kelas positif sebagai kategori risiko tinggi sesuai label simulasi.

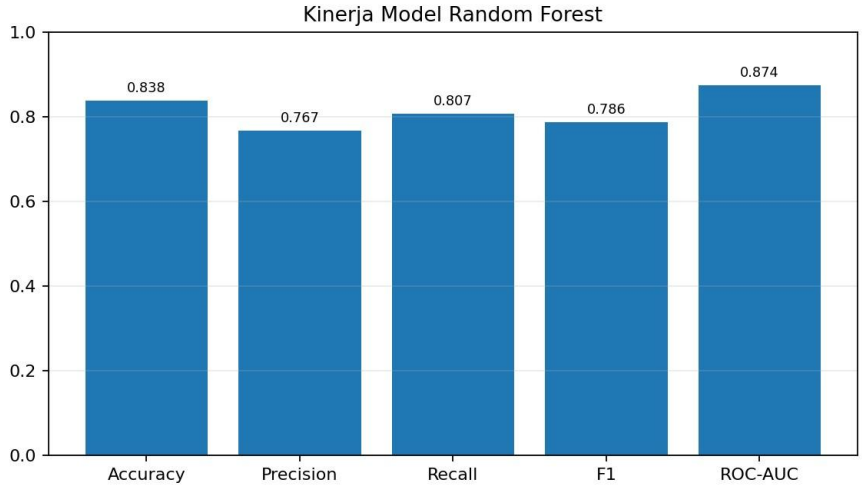


Gambar 3. Confusion Matrix pada Data Testing

Metrik	Nilai	Interpretasi pada Data Uji
Accuracy	0,8377	129/154 prediksi benar

Precision	0,7667	46/60 prediksi positif benar
Recall	0,8070	46/57 kelas positif terdeteksi
F1-score	0,7863	Harmonik precision dan recall
ROC-AUC	0,8739	Diskriminasi skor model

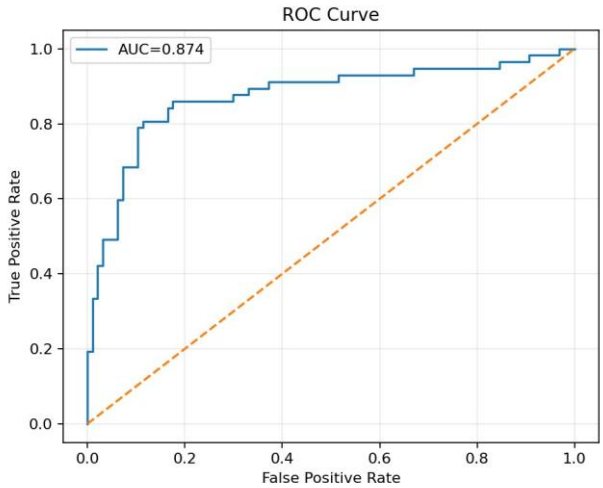
Accuracy 0,8377 menggambarkan proporsi klasifikasi benar pada data testing. Precision 0,7667 berarti sebagian prediksi risiko tinggi merupakan positif palsu, sedangkan recall 0,8070 menunjukkan masih terdapat 11 kasus positif pada data uji yang diprediksi negatif. F1-score 0,7863 merangkum keseimbangan precision dan recall, sementara ROC-AUC 0,8739 menunjukkan kemampuan membedakan kelas pada variasi threshold. Karena data simulasi bukan pengamatan klinis, angka tersebut tidak dapat disamakan dengan sensitivitas atau spesifisitas perangkat skrining di lapangan.



Gambar 4. Perbandingan Metrik Pengujian Random Forest

Dari confusion matrix dapat pula dihitung specificity=83/(83+14)=0,8557 atau 85,57%. Nilai tersebut melengkapi recall untuk menunjukkan proporsi negatif yang dikenali dengan benar. Pemeriksaan turunan ini menggunakan nilai sel yang sudah tersedia, bukan eksperimen baru. Model tidak memiliki ambang keputusan klinis yang telah disepakati, sehingga trade-off antara false positive dan false negative masih membutuhkan pembahasan bersama pakar kesehatan.

3.3 Kurva ROC dan Interpretasi Diskriminasi



Gambar 5. Kurva ROC Model Random Forest

Kurva ROC pada Gambar 5 menggambarkan perubahan true positive rate dan false positive rate pada berbagai ambang. Luas area 0,8739 merupakan ringkasan kemampuan pemeringkatan prediksi dalam data

simulasi. Nilai AUC tidak menyatakan probabilitas prediksi sudah terkalibrasi. Model dengan AUC baik dapat tetap menghasilkan probabilitas yang terlalu tinggi atau rendah, sehingga pemeriksaan calibration plot dan Brier score sebaiknya dilakukan pada data relevan sebelum penggunaan praktis [18].

3.4 Repeated Stratified Cross-Validation

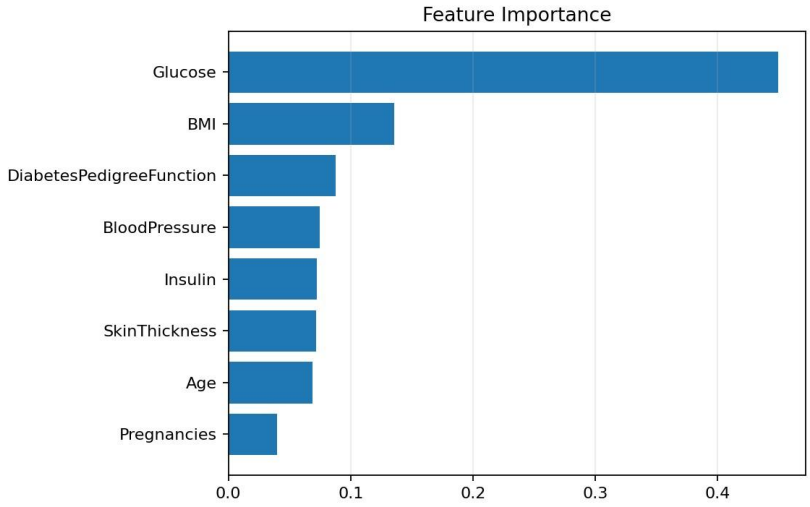
Metrik	Rata-rata	Simpangan baku
Accuracy	0,8033	0,0317
F1-score	0,7306	0,0382
ROC-AUC	0,8595	0,0295

Pengujian silang menghasilkan rata-rata accuracy 0,8033, F1-score 0,7306, dan ROC-AUC 0,8595. Hasil ini lebih rendah daripada sebagian metrik pada satu pembagian data testing. Perbedaan tersebut dapat muncul karena variasi komposisi sampel antar fold. Simpangan baku accuracy 0,0317 dan F1-score 0,0382 menunjukkan fluktuasi yang patut dilaporkan. Namun, pengujian silang masih berbasis dataset simulasi yang sama dan tidak setara dengan validasi pada rumah sakit, wilayah, atau periode berbeda.

Perbandingan antara satu hold-out dan repeated cross-validation perlu dilakukan secara hati-hati. Jika CV menggunakan keseluruhan dataset yang juga mencakup testing set, interpretasinya menjadi pemeriksaan internal tambahan dan bukan validasi independen atas model akhir. Untuk pengembangan empiris, data eksternal perlu disimpan terpisah sejak awal dan proses pencarian hyperparameter harus berada di dalam fold pelatihan. Pelaporan sebaiknya memenuhi prinsip TRIPOD+AI [17].

3.5 Analisis Kepentingan Fitur

Peringkat	Fitur	Importance
1	Glucose	0,4495
2	BMI	0,1357
3	DiabetesPedigreeFunction	0,0876
4	BloodPressure	0,0746
5	Insulin	0,0721
6	SkinThickness	0,0719
7	Age	0,0688
8	Pregnancies	0,0397



Gambar 6. Feature Importance pada Model Simulasi

Fitur Glucose memiliki kontribusi relatif terbesar pada pembentukan prediksi model sebesar 0,4495. BMI menempati urutan kedua dengan 0,1357 dan DiabetesPedigreeFunction ketiga dengan 0,0876. Nilai ini merupakan feature importance yang dilaporkan oleh model, bukan besaran risiko klinis dan bukan pengukuran hubungan kausal. Bila dua variabel saling berkorelasi, pembagian importance dapat dipengaruhi cara pohon

memilih split. Analisis lanjutan dapat menggunakan permutation importance atau penjelasan berbasis SHAP, disertai kajian stabilitas penjelasan pada beberapa sampel [20].

3.6 Diskusi Dibandingkan Penelitian Terdahulu

Hasil pada laporan menunjukkan bahwa Random Forest dapat menjalankan tugas klasifikasi biner dengan sejumlah atribut terstruktur. Penelitian terdahulu pada prediksi diabetes [1]–[15] menggunakan variasi model, dataset dan strategi evaluasi. Kesamaan penggunaan Random Forest dengan Ginting dkk. [12], Sriyanto dan Supriyatna [13], Andi dkk. [14], serta Noviyanti dan Alamsyah [15] memberikan relevansi metode, tetapi belum cukup untuk menyimpulkan adanya peningkatan performa. Perbandingan angka dari sumber lain hanya sah apabila label, populasi, preprocessing, pembagian dataset, dan definisi metrik dapat disejajarkan.

Kontribusi nyata dari naskah yang tersedia adalah dokumentasi verifikasi pipeline, tabel hyperparameter, evaluasi hold-out serta repeated cross-validation, dan pelaporan feature importance. Klaim novelty metodologis perlu dibatasi karena Random Forest dan metriknya telah banyak digunakan. Jika hendak menargetkan jurnal penelitian dengan tuntutan kontribusi empiris kuat, peneliti perlu menyertakan data asli yang dapat digunakan secara sah, baseline seperti Logistic Regression atau XGBoost, uji perbandingan statistik, dan evaluasi validasi eksternal. Peningkatan panjang pembahasan atau jumlah referensi tidak bisa menggantikan kebutuhan bukti penelitian.

3.7 Keterbatasan dan Implikasi

Keterbatasan utama adalah 768 record bersifat simulasi, sehingga pola distribusi fitur, prevalensi, hubungan antarvariabel, dan tingkat kesulitan klasifikasi dapat berbeda secara mendasar dari data pasien. Selain itu, file sumber belum menyertakan dataset mentah, kode implementasi, protokol preprocessing rinci, atau provenance label. Tanpa unsur ini, reviewer tidak dapat memastikan apakah model benar-benar menggeneralisasi pada data lain maupun apakah seluruh hasil dapat direplikasi. Karena konteks domain merupakan kesehatan, penggunaan hasil sebagai alat medis tidak dibenarkan.

Sebelum aplikasi operasional dikembangkan, perlu validasi perbedaan antara target klasifikasi diagnosis saat ini dan prediksi risiko kejadian di masa depan. Penggunaan istilah “risiko” secara ilmiah mensyaratkan definisi target, horizon waktu, dan outcome yang jelas; pada data simulasi hanya tersedia label biner. Pengembangan lanjutan dapat menilai fairness antarkelompok, kualitas pengukuran laboratorium, penanganan missingness, pembaruan model ketika populasi berubah, perlindungan data pribadi, serta mekanisme oversight oleh tenaga kesehatan.

3.8 Rekomendasi Rancangan Validasi Lanjutan

Untuk menaikkan kualitas bukti dari verifikasi menjadi penelitian empiris, pengumpulan data lanjutan perlu menetapkan tujuan model sebelum pelatihan. Bila sasaran adalah prediksi kejadian diabetes baru, data harus memiliki waktu pengamatan awal, periode tindak lanjut, dan definisi kejadian yang sesuai. Sebaliknya, bila sasaran adalah klasifikasi status diabetes pada saat pemeriksaan, penelitian perlu menjelaskan pemeriksaan rujukan yang membentuk label. Kedua tujuan memiliki konsekuensi validasi berbeda. Selain itu, karakteristik usia dan jenis kelamin populasi sumber, kriteria inklusi-eksklusi, serta alasan penghapusan record harus dilaporkan secara konsisten. Ketidakhadiran informasi ini membatasi interpretasi hasil meski metrik numeriknya tinggi.

Rancangan validasi yang lebih kuat dimulai dengan pemisahan data pasien secara unik untuk mencegah catatan individu yang sama muncul pada bagian training dan testing. Imputasi, encoding, seleksi fitur, dan penyetelan parameter harus dilakukan hanya berdasarkan bagian training. Evaluasi akhir sebaiknya mempertahankan dataset yang benar-benar tidak disentuh selama pengembangan model. Pelaporan hasil dapat mencakup interval kepercayaan, metrik per kelompok, calibration curve, dan ukuran potensi manfaat klinis. Dalam penelitian dengan beberapa institusi, pemisahan menurut lokasi atau waktu dapat memberi gambaran ketahanan model menghadapi variasi distribusi data. Tidak satu pun pemeriksaan tambahan tersebut telah dilakukan pada laporan sumber; uraian ini merupakan rancangan pengembangan, bukan temuan baru.

Aspek interpretabilitas juga perlu dibuktikan pada tingkat individu, bukan hanya dengan satu tabel feature importance global. Peneliti dapat membandingkan interpretasi berbasis fitur dengan pengetahuan klinis melalui konsultasi ahli dan pemeriksaan kasus salah klasifikasi. Penting untuk mengevaluasi stabilitas fitur ketika data diambil dari periode yang berbeda. Variabel Glucose yang dominan dalam simulasi mungkin berasal dari cara

data dibentuk dan belum menjamin dominasinya pada data lapangan. Dokumentasi kode, versi perangkat lunak, konfigurasi random seed, serta ketersediaan dataset yang dapat dibagikan secara aman akan mempermudah replikasi dan mengurangi risiko klaim yang tidak teruji.

4. KESIMPULAN

Penelitian ini menyajikan verifikasi alur klasifikasi berbasis Random Forest pada 768 record simulasi yang memiliki delapan atribut numerik dan label biner. Model dikonfigurasi dengan 180 pohon, kedalaman maksimum sembilan, min_samples_leaf dua, dan pembobotan kelas balanced_subsample. Pada 154 data uji, confusion matrix memperlihatkan 83 true negative, 14 false positive, 11 false negative, dan 46 true positive. Evaluasi menghasilkan accuracy 0,8377, precision 0,7667, recall 0,8070, F1-score 0,7863, dan ROC-AUC 0,8739. Repeated stratified five-fold cross-validation dua pengulangan menghasilkan rata-rata accuracy 0,8033, F1-score 0,7306, dan ROC-AUC 0,8595, sehingga variasi hasil antar pembagian data tetap perlu diperhatikan. Analisis atribut menempatkan Glucose sebagai feature importance tertinggi sebesar 0,4495, diikuti BMI dan DiabetesPedigreeFunction. Hasil tersebut membuktikan keberhasilan verifikasi prosedural dan konsistensi ringkasan perhitungan pada data simulasi, bukan efektivitas diagnosis maupun prediksi risiko klinis. Penelitian ini belum mendemonstrasikan generalisasi pada pasien nyata, tidak melakukan validasi eksternal atau kalibrasi, dan tidak membandingkan model terhadap baseline dalam desain eksperimen setara. Untuk memenuhi tuntutan penelitian empiris, pengembangan berikutnya memerlukan dataset klinis terverifikasi, dokumentasi preprocessing, lampiran kode, evaluasi risiko kesalahan, pelaporan sesuai TRIPOD+AI, serta keterlibatan tenaga kesehatan. Oleh karena itu, rekomendasi model harus dibatasi sebagai contoh komputasional dan bukan dasar penetapan diagnosis atau keputusan medis individual.

REFERENCES

[1] H. Lai, H. Huang, K. Keshavjee, A. Guergachi, and X. Gao, “Predictive models for diabetes mellitus using machine learning techniques,” BMC Endocrine Disorders, vol. 19, art. 101, 2019, doi: 10.1186/s12902-019-0436-6.

[2] Z. Xie, O. Nikolayeva, J. Luo, and D. Li, “Building Risk Prediction Models for Type 2 Diabetes Using Machine Learning Techniques,” Preventing Chronic Disease, vol. 16, art. E130, 2019, doi: 10.5888/pcd16.190109.

[3] A. Dinh, S. Miertschin, A. Young, and S. D. Mohanty, “A data-driven approach to predicting diabetes and cardiovascular disease with machine learning,” BMC Medical Informatics and Decision Making, vol. 19, art. 211, 2019, doi: 10.1186/s12911-019-0918-5.

[4] A. Mujumdar and V. Vaidehi, “Diabetes Prediction using Machine Learning Algorithms,” Procedia Computer Science, vol. 165, pp. 292–299, 2019, doi: 10.1016/j.procs.2020.01.047.

[5] S. Kaul and Y. Kumar, “Artificial Intelligence-based Learning Techniques for Diabetes Prediction: Challenges and Systematic Review,” SN Computer Science, vol. 1, art. 322, 2020, doi: 10.1007/s42979-020-00337-2.

[6] L. J. Muhammad, E. A. Algehyne, and S. S. Usman, “Predictive Supervised Machine Learning Models for Diabetes Mellitus,” SN Computer Science, vol. 1, art. 240, 2020, doi: 10.1007/s42979-020-00250-8.

[7] U. M. Butt, S. Letchmunan, M. Ali, F. H. Hassan, A. Baqir, and H. H. R. Sherazi, “Machine Learning Based Diabetes Classification and Prediction for Healthcare Applications,” Journal of Healthcare Engineering, vol. 2021, art. 9930985, 2021, doi: 10.1155/2021/9930985.

[8] C. Olisah, O. Adeleye, L. Smith, and M. Smith, “A Robust Machine Learning Framework for Diabetes Prediction,” Future Technologies Conference 2021, vol. 359, pp. 775–792, 2021, doi: 10.1007/978-3-030-89880-9_58.

[9] V. Chang, M. A. Ganatra, K. Hall, L. Golightly, and Q. A. Xu, “An assessment of machine learning models and algorithms for early prediction and diagnosis of diabetes using health indicators,” Healthcare Analytics, vol. 2, art. 100118, 2022, doi: 10.1016/j.health.2022.100118.

[10] S. Kodama et al., “Predictive ability of current machine learning algorithms for type 2 diabetes mellitus: A meta-analysis,” Journal of Diabetes Investigation, vol. 13, no. 5, pp. 900–908, 2022, doi: 10.1111/jdi.13736.

[11] M. O. Olusanya, R. E. Ogunsakin, M. Ghai, and M. A. Adeleke, “Accuracy of Machine Learning Classification Models for the Prediction of Type 2 Diabetes Mellitus: A Systematic Survey and Meta-Analysis Approach,” IJERPH, vol. 19, no. 21, art. 14280, 2022, doi: 10.3390/ijerph192114280.

[12] J. B. Ginting, T. Suci, C. N. Ginting, and D. D. Hartono, “Early detection system of risk factors for diabetes mellitus type 2 utilization of machine learning-random forest,” Journal of Family and Community Medicine, vol. 30, no. 3, pp. 171–179, 2023, doi: 10.4103/jfcm.jfcm_33_23.

[13] Sriyanto and A. R. Supriyatna, "Prediksi Penyakit Diabetes Menggunakan Algoritma Random Forest," TEKNIKA, vol. 17, no. 1, pp. 163–172, 2023, doi: 10.5281/zenodo.8051410.

[14] Andi, Thamrin, A. Susanto, E. Wijaya, and D. Djohan, "Analysis of the random forest and grid search algorithms in early detection of diabetes mellitus disease," Jurnal Mantik, vol. 7, no. 2, 2023, doi: 10.35335/mantik.v7i2.3981.

[15] C. N. Noviyanti and A. Alamsyah, "Early Detection of Diabetes Using Random Forest Algorithm," Journal of Information System Exploration and Research, vol. 2, no. 1, 2024, doi: 10.52465/joiser.v2i1.245.

[16] L. Breiman, "Random Forests," Machine Learning, vol. 45, pp. 5–32, 2001, doi: 10.1023/A:1010933404324.

[17] G. S. Collins et al., "TRIPOD+AI statement: updated guidance for reporting clinical prediction models that use regression or machine learning methods," BMJ, vol. 385, art. e078378, 2024, doi: 10.1136/bmj-2023-078378.

[18] B. Van Calster, D. J. McLernon, M. van Smeden, L. Wynants, and E. W. Steyerberg, "Calibration: the Achilles heel of predictive analytics," BMC Medicine, vol. 17, art. 230, 2019, doi: 10.1186/s12916-019-1466-7.

[19] D. Chicco and G. Jurman, "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation," BMC Genomics, vol. 21, art. 6, 2020, doi: 10.1186/s12864-019-6413-7.

[20] S. M. Lundberg et al., "From local explanations to global understanding with explainable AI for trees," Nature Machine Intelligence, vol. 2, pp. 56–67, 2020, doi: 10.1038/s42256-019-0138-9.