

Klasifikasi Multikelas Tingkat Diabetes Berdasarkan Indikator Kesehatan Pasien Menggunakan Strategi One-vs-Rest

Tabitha Martha Agustine*, Robet, Octara Pribadi

Program Studi Teknik Informatika, STMIK TIME, Medan, Indonesia

Email: ^{1,*}tabithamartha885@gmail.com, ²robertdetime@gmail.com, ³octarapribadi@gmail.com

Email Penulis Korespondensi: tabithamartha885@gmail.com*

Submitted: 06/08/2025; Accepted: 01/11/2025; Published: 31/12/2025

Abstrak - Diabetes merupakan penyakit tidak menular yang prevalensinya terus meningkat setiap tahun, dan sering kali tidak terdiagnosis pada tahap awal, terutama pada kondisi *prediabetes* yang umumnya tidak menunjukkan gejala mencolok. Penelitian ini bertujuan melakukan komparasi model klasifikasi multikelas untuk memprediksi tingkat keparahan diabetes, yaitu non-diabetes, *prediabetes*, dan diabetes, berdasarkan indikator kesehatan pasien. Metode yang digunakan adalah pendekatan *One-vs-Rest (OvR)*, yang melatih masing-masing kelas terhadap gabungan dua kelas lainnya. Dataset penelitian berasal dari survei kesehatan Nasional tahun 2015, mencakup lebih dari 250.000 data pasien dengan indikator kesehatan seperti tekanan darah, indeks massa tubuh, kadar kolesterol, riwayat penyakit jantung, dan aktivitas fisik. Untuk mengatasi ketidakseimbangan kelas, digunakan metode *Synthetic Minority Oversampling Technique (SMOTE)*. Algoritma *Random Forest* dipilih sebagai metode utama karena terbukti memberikan kinerja yang stabil dalam penerapan strategi *OvR*, sedangkan *Logistic Regression* digunakan sebagai pembandingan untuk menilai konsistensi hasil. Evaluasi dilakukan menggunakan metrik klasifikasi dan validasi silang (*K-Fold Cross Validation*). Hasil penelitian menunjukkan bahwa penerapan strategi *One-vs-Rest (OvR)* dengan *Random Forest* efektif dalam meningkatkan kinerja klasifikasi multikelas tingkat diabetes, dengan akurasi dan *F1-score* masing-masing sebesar 93%, sedangkan *Logistic Regression* hanya mencapai akurasi 52%. Temuan ini menegaskan potensi penerapan pembelajaran mesin dalam mendukung deteksi dini diabetes serta menjadi dasar bagi penelitian lanjutan di bidang kesehatan berbasis data.

Kata Kunci: Diabetes; Klasifikasi Multikelas; One-vs-Rest; Indikator Kesehatan; Random Forest

Abstract - Diabetes is a non-communicable disease whose prevalence continues to increase every year. It often goes undiagnosed in its early stages, especially in *prediabetes*, which generally does not show any noticeable symptoms. This study aims to compare a multi-class classification model to predict the severity level of diabetes, namely non-diabetes, *prediabetes*, and diabetes, based on patient health indicators. The method employed is the *One-vs-Rest (OvR)* approach, which trains each class against the combination of the remaining classes. The dataset was obtained from the 2015 National Health Survey, comprising over 250,000 patient records with health indicators such as blood pressure, body mass index, cholesterol levels, a history of heart disease, and physical activity. To address class imbalance, the *Synthetic Minority Oversampling Technique (SMOTE)* was applied. The *Random Forest* algorithm was chosen as the primary method due to its stable performance in implementing the *OvR* strategy. At the same time, *Logistic Regression* was used as a comparison to assess the consistency of results. Evaluation was carried out using classification metrics and *K-fold cross-validation*. The results show that applying the *OvR* strategy with *Random Forest* is efficacious in improving multiclass diabetes classification performance, achieving 93% accuracy and *F1-score*, whereas *Logistic Regression* achieved about 52%. These findings emphasise the potential of *machine learning* to support early diabetes detection and provide a foundation for further data-driven health research.

Keywords: Diabetes; Multiclass Classification; One-vs-Rest; Health Indicators; Random Forest

1. PENDAHULUAN

Diabetes Mellitus (DM) merupakan penyakit metabolik kronis yang berdampak luas terhadap berbagai sistem organ tubuh, sehingga sering disebut sebagai “induk dari berbagai penyakit” karena potensinya memicu berbagai komplikasi serius seperti nefropati, neuropati, dan retinopati [1]. Selain itu, DM juga menjadi salah satu penyebab utama kematian akibat komplikasi berat seperti gagal ginjal, kebutaan, stroke, dan penyakit jantung [2]. Deteksi dini menjadi hal yang krusial karena pada tahap *prediabetes* yakni kondisi sebelum seseorang benar-benar terkena DM tipe 2 biasanya tidak muncul gejala khas [3]. Untuk itu, edukasi dan dukungan pengelolaan mandiri diperlukan guna menurunkan risiko komplikasi jangka pendek maupun panjang [4]. Risiko transisi dari *prediabetes* ke diabetes tipe 2 bahkan diperkirakan mencapai lebih dari 70% dalam kurun waktu lima tahun apabila tidak disertai intervensi gaya hidup yang tepat [5].

Berdasarkan laporan Riskesdas tahun 2018, prevalensi diabetes di Indonesia mengalami peningkatan yang signifikan dalam lima tahun terakhir, yaitu dari 6,9% menjadi 8,5%. Selama ini, diabetes lebih dikenal sebagai penyakit yang menyerang kelompok usia lanjut. Namun, pola konsumsi dan gaya hidup masyarakat yang semakin tidak sehat turut mendorong peningkatan kasus pada kelompok usia yang lebih muda, termasuk anak-anak dan remaja. Di Indonesia, data registri nasional menunjukkan lebih dari 1.200 anak dan remaja penderita diabetes

tersebar di kota-kota besar seperti Jakarta, Bandung, Yogyakarta, dan Denpasar, dengan sebagian besar kasus berada pada kelompok usia 11–15 tahun [6].

Prediabetes merupakan gangguan metabolik yang ditandai oleh kadar glukosa darah yang melebihi ambang normal, namun belum mencapai batas diagnosis diabetes secara klinis. Selain itu, *prediabetes* atau gangguan toleransi glukosa juga menjadi isu kesehatan yang signifikan, di mana prevalensinya pada remaja Indonesia tercatat mencapai lebih dari 36% [7]. Individu dengan *prediabetes* memiliki risiko yang jauh lebih tinggi untuk berkembang menjadi diabetes tipe 2 dalam kurun waktu lima tahun. Sayangnya, sebagian besar studi terdahulu masih mengelompokkan pasien hanya dalam dua kategori, yaitu diabetes dan non-diabetes, sehingga fase kritis *prediabetes* kerap terabaikan dalam proses identifikasi dini. Menurut Menke et al., *prediabetes* sering kali tidak terdeteksi akibat ketiadaan gejala klinis yang mencolok [8].

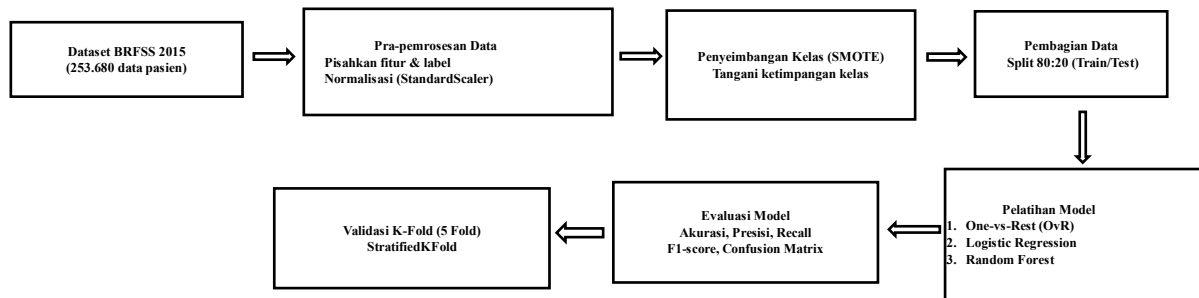
Kemajuan dalam bidang *machine learning* (ML) telah membuka peluang besar dalam klasifikasi berbagai jenis penyakit, termasuk diabetes. Teknologi ini mampu membangun model dengan kompleksitas tinggi dan akurasi yang baik, serta dapat diadaptasi untuk klasifikasi multikelas berdasarkan data medis dan perilaku gaya hidup [9]. Beragam algoritma telah digunakan dalam upaya deteksi dini diabetes, salah satunya adalah *Random Forest* yang dipadukan dengan teknik *SMOTE* [10], *Support Vector Machine* (SVM) yang dipadukan dengan teknik *SMOTE* untuk mengatasi ketidakseimbangan data [11], hingga pendekatan *hibrid* antara *Random Forest* dan *SVM + SMOTE* yang dilaporkan mampu mencapai akurasi sebesar 89% [12]. Selain itu, pendekatan berbasis *deep learning* seperti *Long Short-Term Memory* (LSTM) dan *Convolutional Neural Network* (CNN) juga mulai diterapkan dalam deteksi diabetes, karena kemampuannya mengenali pola sekuensial pada data medis secara lebih akurat [13] [14].

Meskipun demikian, sebagian besar penelitian masih berfokus pada pemisahan dua kelas (*biner*), yakni antara diabetes dan non-diabetes. Oleh karena itu, dibutuhkan pendekatan klasifikasi *multikelas* yang mampu mengelompokkan pasien ke dalam tiga kategori, yakni non-diabetes, *prediabetes*, dan diabetes. Salah satu metode yang relevan untuk tujuan ini adalah strategi *One-vs-Rest* (*OvR*), yaitu pendekatan di mana sebuah model dikembangkan untuk membedakan satu kelas tertentu dari keseluruhan kelas lainnya. Strategi ini dinilai efektif untuk menangani masalah ketidakseimbangan data seperti yang umum ditemukan dalam populasi penderita *Diabetes Mellitus*, karena setiap kelas diperlakukan secara terpisah [9]. Sebuah studi dalam *Scientific Reports* [15] menunjukkan keberhasilan penggunaan pendekatan *OvR* dalam mendeteksi DM dengan tingkat akurasi mencapai 89%. Di tingkat Nasional, penelitian oleh Rahmawati et al. [16] membandingkan performa beberapa algoritma seperti *SVM*, *Logistic Regression*, dan *Naïve Bayes* yang dikombinasikan dengan *SMOTE*, menggunakan data dari rumah sakit regional. Hasilnya, model *SVM* dengan *SMOTE* menghasilkan akurasi 95,8% dan *AUC* sebesar 99,1%, lebih tinggi dibanding model-model lain tanpa teknik penyeimbangan data. Penelitian terbaru juga mengusulkan pendekatan *multikelas* berbasis *LightGBM* dan *Gradient Boosting* yang terbukti menghasilkan klasifikasi yang seimbang dan efisien dalam mendeteksi stadium *diabetes* [17] [18].

Merujuk pada uraian sebelumnya, penelitian ini bertujuan membangun model klasifikasi multikelas tingkat diabetes (non-diabetes, *prediabetes*, dan diabetes) menggunakan strategi *One-vs-Rest* (*OvR*). Algoritma *Random Forest* dipilih sebagai metode utama karena konsistensi kinerjanya, sedangkan *Logistic Regression* hanya digunakan sebagai pembanding untuk memastikan konsistensi hasil. Dataset yang digunakan adalah *BRFSS 2015* yang mencakup lebih dari 250.000 responden dengan berbagai indikator kesehatan. Penelitian ini diharapkan dapat mendukung pengembangan sistem deteksi dini diabetes berbasis data medis yang akurat serta adaptif di berbagai konteks, termasuk Indonesia.

2. METODOLOGI PENELITIAN

Metodologi penelitian ini dirancang untuk membangun model klasifikasi multikelas tingkat diabetes yang akurat dan andal. Tahapannya meliputi akuisisi dan eksplorasi dataset *BRFSS 2015*, pra-pemrosesan data, penyeimbangan kelas dengan *SMOTE*, pembagian data untuk pelatihan dan pengujian, serta pelatihan dan evaluasi model. Validitas hasil dijamin melalui stratified K-fold cross-validation. Seluruh proses penelitian dirangkum dalam alur tahapan berikut.



Gambar 1. Tahapan Penelitian

2.1 Akuisisi dan Eksplorasi Data

Dataset yang digunakan dalam penelitian ini bersumber dari *Behavioral Risk Factor Surveillance System (BRFSS)* tahun 2015, sebuah survei Nasional mengenai kesehatan masyarakat yang dikembangkan oleh *Centers for Disease Control and Prevention (CDC)*, Amerika Serikat. *BRFSS* dapat diakses secara publik dan telah banyak digunakan dalam berbagai penelitian yang berkaitan dengan epidemiologi dan penyakit kronis. Dataset dapat diunduh melalui dua sumber berikut:

- Kaggle: *BRFSS 2015 Dataset* on Kaggle
- Situs resmi CDC: *CDC BRFSS 2015*

Dataset yang digunakan memiliki 253.680 entri dan 22 atribut yang merepresentasikan berbagai indikator kesehatan, seperti tekanan darah tinggi, indeks massa tubuh, kebiasaan merokok, hingga konsumsi buah dan sayur. Variabel target dalam dataset ini adalah ‘Diabetes_012’, yang mengelompokkan kondisi responden menjadi tiga kategori: non-diabetes (0), *prediabetes* (1), dan diabetes (2). Beberapa atribut penting dalam dataset ini ditampilkan pada tabel berikut:

Tabel 1. Atribut dalam Dataset *BRFSS 2015*

NO	Nama Atribut	Deskripsi
1	HighBP	Riwayat tekanan darah tinggi
2	HighChol	Kolesterol tinggi
3	BMI	Indeks massa tubuh (Body Mass Index)
4	Smoker	Riwayat merokok
5	Stroke	Riwayat stroke
6	HeartDiseaseorAttack	Riwayat serangan jantung
7	PhysActivity	Aktivitas fisik secara rutin
8	Fruits	Konsumsi buah-buahan secara rutin
9	Veggies	Konsumsi sayur-sayuran secara rutin
10	GenHlth	Penilaian kondisi kesehatan secara umum (skala 1–5)
... (total 22 fitur hingga label Diabetes_012)

Eksplorasi awal dilakukan menggunakan *Python* untuk melihat struktur data dan distribusi kelas target. Hasil eksplorasi menunjukkan bahwa distribusi kelas tidak merata, dengan mayoritas data berada pada kelas 0 (non-diabetes), diikuti oleh kelas 2 (diabetes), dan hanya sebagian kecil pada kelas 1 (*prediabetes*). Ketidakseimbangan ini perlu diatasi agar model tidak bias terhadap kelas mayoritas.

2.2 Pra-pemrosesan Data

Tahapan pra-pemrosesan bertujuan untuk mempersiapkan data agar layak digunakan dalam proses pelatihan model *machine learning*. Langkah awal adalah memisahkan variabel fitur (X) dan label (y). Setelah itu, seluruh fitur dinormalisasi menggunakan metode *StandardScaler* agar berada pada skala yang seragam. Normalisasi ini penting untuk menghindari dominasi fitur dengan nilai besar yang dapat mempengaruhi proses pembelajaran algoritma klasifikasi. Setelah dilakukan normalisasi, data siap untuk tahapan selanjutnya, yaitu penyeimbangan kelas (*balancing data*) yang akan dilakukan dengan teknik *SMOTE* pada subbab berikutnya.

2.3 Penyeimbangan Data Menggunakan SMOTE

Pada dataset *BRFSS* 2015, ditemukan adanya ketidakseimbangan distribusi kelas pada variabel target *Diabetes_012*, yaitu:

- Kelas 0 (*non-diabetes*) memiliki jumlah sampel terbanyak.
- Kelas 2 (*diabetes*) jumlahnya sedang.
- Kelas 1 (*prediabetes*) merupakan kelas minoritas yang jumlahnya paling sedikit.

Ketidakseimbangan distribusi antar kelas dalam dataset dapat mengakibatkan model cenderung bias terhadap kelas yang dominan. Untuk mengatasi hal ini, digunakan pendekatan *Synthetic Minority Over-sampling Technique* (*SMOTE*), yang bekerja dengan menghasilkan sampel sintetis pada kelas minoritas guna menyetarakan proporsinya dengan kelas mayoritas, tanpa melakukan duplikasi data asli. Hasil dari proses ini adalah dataset yang seimbang secara kuantitas antar kelas, memungkinkan model untuk belajar dari ketiga kelas secara adil.

2.4 Pembagian Data Pelatihan dan Pengujian

Usai dilakukan normalisasi serta penyeimbangan kelas dengan metode *SMOTE*, dataset kemudian dipisahkan menjadi dua kelompok utama, yakni 80% untuk pelatihan model dan 20% sisanya untuk pengujian performa model. Pembagian ini merupakan praktik umum dalam pengembangan model *machine learning*, yang bertujuan untuk menghindari *overfitting*, yaitu kondisi ketika model terlalu menyesuaikan diri terhadap data pelatihan sehingga kehilangan kemampuan untuk mengenali pola baru.

Dataset pelatihan berperan dalam membentuk model dan mengoptimalkan parameter yang digunakan, sementara dataset pengujian dimanfaatkan untuk mengukur kemampuan generalisasi model terhadap data baru yang belum dikenali sebelumnya. Proporsi 80:20 dipilih agar model memiliki cukup data untuk belajar sekaligus tetap menyediakan data yang representatif untuk pengujian performa secara objektif.

2.5 Pelatihan Model dengan Strategi One-vs-Rest (OvR)

Untuk klasifikasi multikelas, penelitian ini menggunakan strategi *One-vs-Rest* (*OvR*), yang bekerja dengan cara melatih satu model untuk masing-masing kelas terhadap gabungan dua kelas lainnya. Pendekatan ini dinilai efektif dalam menangani masalah klasifikasi yang tidak seimbang dan banyak digunakan dalam domain medis [19]. Dua algoritma utama yang digunakan adalah:

- Logistic Regression*: dipilih karena sifatnya sederhana dan efisien, serta sudah menjadi metode andalan dalam prediksi klinis [20]. Perannya dalam penelitian ini adalah sebagai pembanding untuk memastikan konsistensi dan validitas hasil evaluasi, bukan sebagai fokus utama.
- Random Forest*: algoritma ini digunakan karena dinilai efektif dalam mengelola data yang memiliki kompleksitas tinggi serta hubungan antar fitur yang tidak linier. Hal ini sejalan dengan hasil penelitian sebelumnya yang menunjukkan bahwa integrasi *Random Forest* dengan algoritma *evolutioner* seperti *Particle Swarm Optimization* (*PSO*) dapat meningkatkan akurasi dan kinerja prediksi [21].

2.5.1 Konsep Matematis Strategi OvR

Strategi *OvR* mengubah klasifikasi multikelas menjadi beberapa klasifikasi biner. Untuk setiap kelas $i \in \{1, 2, \dots, k\}$, dibentuk satu model klasifikasi biner:

$$h_i(x) = \{ 1, \text{jika } y = i; 0, \text{jika } y \neq i \} \quad (1)$$

Kemudian untuk prediksi, digunakan fungsi *argmax* untuk memilih kelas dengan skor tertinggi dari semua model:

$$\hat{y} = \arg \max h_i(x), i \in \{1, 2, \dots, k\} \quad (2)$$

2.5.2 Pseudocode Strategi One-vs-Rest (OvR)

Input:

- Dataset $D = (X, y)$, dengan $y \in \{1, 2, \dots, k\}$
- Algoritma dasar (*Logistic Regression* atau *Random Forest*)

Langkah:

- Untuk setiap kelas i dari 1 sampai k :
 - Buat label biner y_i :
$$y_i = \{ 1, \text{jika } y = i; 0, \text{lainnya} \} \quad (3)$$
 - Latih model $_i$ menggunakan (X, y_i)
- Simpan semua model: $[\text{model}_1, \text{model}_2, \dots, \text{model}_k]$
- Untuk prediksi sampel baru x :

1. Hitung skor dari setiap model i
2. Pilih kelas dengan skor tertinggi:

$$\hat{y} = \arg \max(model1(x), \dots, modelk(x)) \quad (4)$$

2.6 Evaluasi Model

Evaluasi dilakukan untuk menilai performa model dalam mengklasifikasikan tiga tingkat diabetes. Metrik yang digunakan antara lain:

- a. Akurasi: indikator yang menunjukkan seberapa besar persentase prediksi model yang sesuai dengan label sebenarnya dari keseluruhan data.

$$Accuracy = \frac{(TP + TN)}{(TP + TN + FP + FN)} \quad (5)$$

- b. Presisi: menunjukkan seberapa tepat model dalam mengidentifikasi data yang benar-benar termasuk ke dalam kelas positif dari seluruh prediksi positif yang dihasilkan.

$$Precision = \frac{TP}{(TP + FP)} \quad (6)$$

- c. *Recall*: merepresentasikan sejauh mana model mampu mengidentifikasi seluruh data yang benar-benar termasuk dalam kelas positif.

$$Recall = \frac{TP}{(TP + FN)} \quad (7)$$

- d. *F1-score*: merupakan nilai rata-rata harmonik dari metrik presisi dan recall, yang memberikan gambaran keseimbangan antara keduanya.

$$F1 - Score = \frac{2 (Precision \cdot Recall)}{(Precision + Recall)} \quad (8)$$

- e. *Confusion Matrix*: menunjukkan rincian prediksi benar dan salah untuk tiap kelas.

Keterangan:

1. TP = True Positive
2. TN = True Negative
3. FP = False Positive
4. FN = False Negative

Evaluasi dilakukan dengan pendekatan *macro average*, yaitu menghitung metrik untuk setiap kelas secara individual, lalu dirata-rata tanpa mempertimbangkan proporsi kelas. Hal ini penting untuk memastikan bahwa kelas minoritas seperti *prediabetes* tetap mendapatkan perhatian yang setara.

2.6.1. Visualisasi Hasil Evaluasi

- a. Visualisasi *Confusion Matrix*

Confusion Matrix divisualisasikan untuk setiap model guna memberikan representasi detail mengenai jumlah prediksi yang tepat dan keliru pada tiap kelas. Visualisasi ini berperan penting dalam menganalisis pola kesalahan klasifikasi yang terjadi. Model *Logistic Regression* dan *Random Forest* masing-masing divisualisasikan dalam bentuk *heatmap* menggunakan pustaka *seaborn*. *Confusion Matrix* memudahkan analisis terhadap kelas-kelas yang cenderung sering diklasifikasikan secara keliru serta membantu mendeteksi potensi bias model terhadap kategori tertentu.

- b. Grafik *F1-Score* per kelas

Untuk melengkapi analisis numerik, dilakukan visualisasi dalam bentuk grafik *F1-score* per kelas. Visualisasi ini menunjukkan performa masing-masing model dalam membedakan kategori non-diabetes, *prediabetes*, dan diabetes secara lebih jelas. Visualisasi ini sangat penting karena dapat menunjukkan kelemahan model terhadap kelas tertentu, khususnya *prediabetes* yang seringkali memiliki karakteristik mirip dengan kelas lainnya. Grafik ini juga mempermudah pembaca dalam membandingkan performa *Logistic Regression* dan *Random Forest* secara intuitif.

2.6.2. Validasi Model dengan *Stratified K-Fold Cross Validation*

Sebagai pelengkap dari evaluasi menggunakan pembagian data tunggal (80% data latih dan 20% data uji), penelitian ini juga melakukan validasi tambahan menggunakan teknik *stratified K-fold cross validation* sebanyak

5 *fold*. Validasi silang ini dilakukan untuk memastikan bahwa performa model benar-benar konsisten dan tidak bergantung pada subset data tertentu saja.

Teknik *StratifiedKfold* dipilih karena mempertahankan proporsi kelas target (non-diabetes, *prediabetes*, diabetes) pada setiap *fold*. Ini sangat penting dalam konteks klasifikasi multikelas dengan distribusi kelas yang tidak seimbang, karena memastikan bahwa setiap model diuji pada data yang mencerminkan distribusi nyata dari seluruh populasi.

Proses ini sangat selaras dengan pendekatan *One-vs-Rest (OvR)* yang digunakan dalam penelitian ini, di mana setiap model dilatih untuk mengenali satu kelas terhadap gabungan dua kelas lainnya. Dengan menerapkan validasi *K-Fold* pada strategi *OvR*, maka setiap kombinasi pelatihan dan pengujian akan mencerminkan distribusi kelas yang seimbang dan adil. Hal ini meningkatkan keandalan hasil, serta memberikan gambaran menyeluruh terhadap konsistensi kinerja masing-masing model *OvR* pada data yang berbeda-beda.

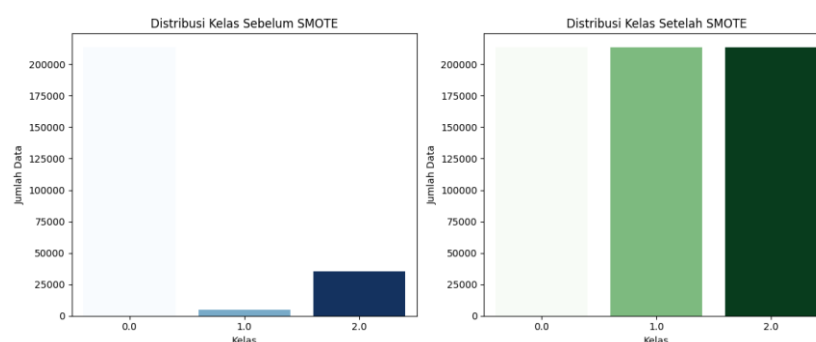
2.6.3 Parameter Eksperimen

Pada proses pelatihan model, digunakan beberapa parameter, antara lain *max_iter* sebesar 1000 dan *random_state* bernilai 42 untuk algoritma *Logistic Regression*. Sementara itu, *Random Forest* menggunakan *n_estimators* sebanyak 100 dengan nilai *random_state* yang sama, yaitu 42. Semua eksperimen dilakukan dalam *pipeline* yang mencakup *StandardScaler* untuk proses normalisasi, serta pendekatan *One-vs-Rest* sebagai strategi klasifikasi multikelas. Validasi model dilakukan menggunakan metode *StratifiedKfold*, yang dibagi ke dalam 5 lipatan (*n_splits*=5), dengan data diacak terlebih dahulu (*shuffle*=*True*) dan pengacakan dikendalikan menggunakan *random_state*=42. Penilaian performa model dilakukan dengan pendekatan *macro average* pada metrik akurasi, *presisi*, *recall*, dan *F1-score*, yang dihitung menggunakan pustaka *scikit-learn* versi terkini.

3. HASIL DAN PEMBAHASAN

3.1 Hasil Penerapan Metode Klasifikasi

Alur eksperimen mencakup pra-pemrosesan, normalisasi fitur, penyeimbangan kelas dengan *SMOTE*, strategi multikelas *One-vs-Rest (OvR)*, serta *Stratified K-Fold (5 folds)* pada dua algoritma utama: *Logistic Regression* dan *Random Forest*. Sebelum penyeimbangan, distribusi kelas sangat timpang, seperti terlihat pada sisi kiri Gambar 2, kelas non-diabetes (0) jauh lebih dominan daripada *prediabetes* (1) dan diabetes (2). Untuk mengurangi ketimpangan kelas, digunakan metode *SMOTE* yang mensintesis data baru pada kelas minoritas hingga distribusinya seimbang dengan kelas mayoritas. Visualisasi hasil distribusi setelah *SMOTE* dapat dilihat pada sisi kanan Gambar 2.



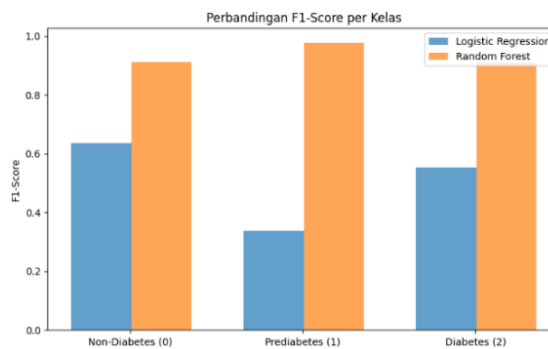
Gambar 2. Perbandingan Distribusi Kelas Sebelum dan Sesudah *SMOTE*

Dataset berisi >250.000 entri dibagi stratified menjadi 80% data latih (± 200.000) dan 20% data uji (± 50.000) dengan *random_state*=42. Kedua subset memuat indikator kesehatan (mis. tekanan darah, IMT, kolesterol, riwayat jantung, aktivitas fisik). Seluruh eksperimen dikemas dalam *pipeline scikit-learn* yang menerapkan *StandardScaler* (dipelajari hanya dari data latih), skema *OvR*, dan model klasifikasi. Dua algoritma dievaluasi: *Logistic Regression* (*max_iter*=1000, *random_state*=42) dan *Random Forest* (*n_estimators*=100, *random_state*=42). Pelatihan pada 80% data latih, menggunakan *StratifiedKfold 5 folds* (*n_splits*=5, *shuffle*=*True*, *random_state*=42). Kinerja akhir dilaporkan pada 20% data uji dengan hasil evaluasi pada akurasi, *presisi*, *recall*, dan *F1-score*, sehingga setiap kelas memperoleh bobot penilaian yang setara.

Tabel 2. Hasil Evaluasi Model Terhadap Data Uji *BRFSS 2015*

Metrik	Logistic Regression	Random Forest (utama)
<i>Accuracy</i>	52%	93%
<i>Precision</i>	51%	93%
<i>Recall</i>	52%	93%
<i>F1-score</i>	51%	93%

Tabel 2 menunjukkan hasil evaluasi model menggunakan *Logistic Regression* dan *Random Forest* dengan strategi *One-vs-Rest*. *Logistic Regression* hanya memperoleh akurasi rata-rata 52%. Sementara itu, *Random Forest* sebagai metode utama memberikan hasil yang jauh lebih tinggi dan stabil dengan akurasi 93%, *precision* 93%, *recall* 93%, dan *F1-score* 93%. Temuan ini menguatkan alasan pemilihan *Random Forest* sebagai fokus utama dalam penelitian ini.



Gambar 3. Perbandingan *F1-score* per kelas antara *Logistic Regression* dan *Random Forest*

Gambar 3 memperlihatkan bahwa *Random Forest* menghasilkan nilai *F1-score* yang tinggi dan konsisten di ketiga kelas target. Peningkatan paling signifikan terlihat pada kelas 1 (*prediabetes*), di mana *F1-score* *Random Forest* mencapai 98%, sedangkan *Logistic Regression* hanya mencapai 34%. Hal ini menandakan bahwa *Random Forest* lebih andal dalam mengenali kondisi *prediabetes*, yang sangat penting untuk upaya pencegahan dini terhadap diabetes tipe 2.

3.2 Evaluasi Hasil

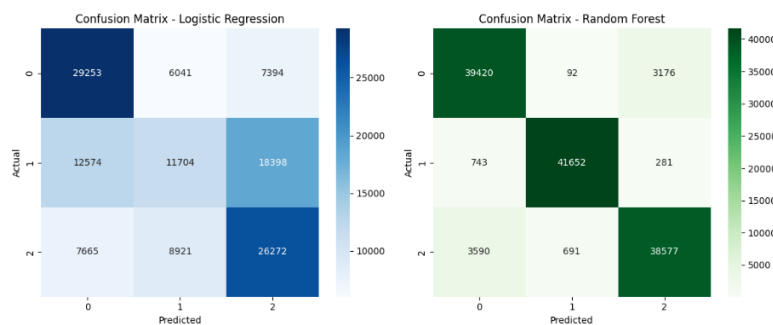
3.2.1 Evaluasi Menggunakan Confusion Matrix

Evaluasi lanjutan dilakukan dengan memanfaatkan *confusion matrix*, yang memberikan gambaran lebih rinci mengenai ketepatan klasifikasi tiap kelas target.

Tabel 3. Ringkasan Evaluasi Confusion Matrix Model terhadap Data Uji *BRFSS 2015*

Model	Confusion Matrix
<i>Logistic Regression</i>	Banyak kesalahan klasifikasi pada kelas 1 dan 2
<i>Random Forest</i>	Akurasi tinggi pada semua kelas, dominan prediksi benar

Tabel 3 menyajikan ringkasan hasil evaluasi *confusion matrix* dari dua model klasifikasi yang digunakan, yaitu *Logistic Regression* dan *Random Forest*. Secara umum, terlihat bahwa *Logistic Regression* sering mengklasifikasikan *prediabetes* (1) sebagai diabetes (2) atau bahkan non-diabetes (0). Sebaliknya, algoritma *Random Forest* mampu menghasilkan klasifikasi yang tinggi tingkat akurasinya dan konsisten pada ketiga kategori kelas.



Gambar 4. Confusion Matrix: Visualisasi Akurasi Klasifikasi

Gambar 4 memperlihatkan perbandingan *confusion matrix* antara model *Logistic Regression* dan *Random Forest* dalam klasifikasi tiga tingkat diabetes. Hasil menunjukkan bahwa *Logistic Regression* memiliki performa yang kurang optimal dalam mengenali kelas *prediabetes* (kelas 1), dengan proporsi besar data yang justru diklasifikasikan sebagai kelas lain, terutama kelas diabetes (kelas 2) sebanyak 18.398 kasus. Sebaliknya, *Random Forest* menunjukkan tingkat akurasi yang sangat tinggi, khususnya pada kelas *prediabetes*, dengan 41.652 data berhasil diklasifikasikan secara tepat dan tingkat kesalahan yang rendah. Temuan ini mengindikasikan bahwa *Random Forest* lebih efektif dalam mengenali pola karakteristik pasien pada fase *prediabetes*, menjadikannya kandidat yang lebih unggul untuk digunakan dalam sistem pendeteksian dini berbasis data medis.

3.2.2 Validasi Model dengan *Stratified K-Fold*

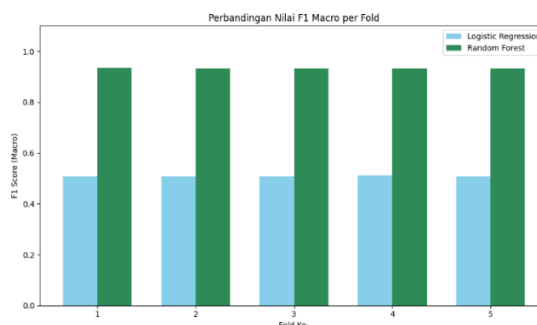
Sebagai pelengkap dari evaluasi berbasis data uji, dilakukan validasi tambahan menggunakan teknik *Stratified K-Fold Cross Validation* sebanyak 5 *fold*. Metode ini menjaga distribusi kelas target secara proporsional dalam setiap *fold*, sehingga meningkatkan stabilitas dan mengurangi potensi bias dalam evaluasi model.

Proses validasi ini sangat relevan dengan pendekatan *One-vs-Rest (OvR)* yang digunakan dalam penelitian, di mana model dilatih untuk mengenali satu kelas terhadap dua kelas lainnya. Dengan penerapan *K-Fold* pada strategi *OvR*, evaluasi menjadi lebih adil karena distribusi kelas tetap seimbang di seluruh kombinasi pelatihan dan pengujian.

Tabel 4. Hasil Rata-rata dan Standar Deviasi Validasi K-Fold

Model	F1 Macro Mean	F1 Std Dev
Logistic Regression	0.5087	0.0017
Random Forest	0.9339	0.0005

Tabel 4 menunjukkan bahwa *Random Forest* tidak hanya memiliki performa rata-rata lebih tinggi, tetapi juga lebih stabil dibandingkan *Logistic Regression*, terlihat dari nilai standar deviasi yang sangat kecil (0.0005).



Gambar 5. Grafik F1-Macro per Fold – *Logistic Regression* vs *Random Forest*

Gambar 5 menampilkan nilai *F1 Macro* pada masing-masing *fold* untuk kedua algoritma. Model *Random Forest* secara konsisten menghasilkan skor di atas 93% pada setiap *fold*. Sebaliknya, *Logistic Regression* memiliki *F1-score* yang stagnan dan jauh lebih rendah di setiap *fold*, berkisar di angka 51%. Pola ini memperkuat bukti bahwa *Random Forest* dengan pendekatan *One-vs-Rest* lebih andal dan konsisten pada klasifikasi multikelas berbasis data medis.

Sebagai bentuk verifikasi, dilakukan perhitungan manual metrik evaluasi berdasarkan *confusion matrix* algoritma *Random Forest* yang memuat nilai TP, FP, FN, TN beserta *precision*, *recall*, dan *F1-score* untuk masing-masing kelas.

Tabel 5. Verifikasi Manual Metrik Evaluasi Random Forest Berdasarkan Confusion Matrix

Kelas	TP	FP	FN	TN	Precision	Recall	F1-score
Non-Diabetes (0)	39.493	4.201	3.247	81.281	90.39%	92.40%	91.38%
<i>Prediabetes</i> (1)	41.725	740	1.016	84.741	98.26%	97.62%	97.94%
Diabetes (2)	38.613	3.450	4.128	82.031	91.80%	90.34%	91.06%
Macro Avg	–	–	–	–	93.48%	93.46%	93.46%

Berdasarkan Tabel 5, kelas *Prediabetes* (1) menunjukkan performa terbaik dengan *precision* 98.26%, *recall* 97.62%, dan *F1-score* 97.94%, dan kelas Non-Diabetes (0) dan Diabetes (2) juga tinggi dengan nilai di atas 90%. Rata-rata (*macro average*) *precision*, *recall*, dan *F1-score* sebesar 93.5% dengan akurasi global 93.46%. Hasil ini konsisten dengan Tabel 2 yang melaporkan akurasi, *precision*, *recall*, dan *F1-score* 93% untuk *Random Forest*, sehingga memvalidasi hasil evaluasi otomatis *scikit-learn*.

3.3 Pembahasan.

Hasil eksperimen menunjukkan bahwa algoritma *Random Forest* dengan strategi *One-vs-Rest* secara signifikan mengungguli *Logistic Regression* dalam klasifikasi multikelas tingkat diabetes. Keunggulan ini tercermin dari seluruh metrik evaluasi, khususnya nilai *F1-score* yang tinggi dan konsisten pada ketiga kelas target, dengan performa terbaik pada kelas 1 (*prediabetes*). Kemampuan *Random Forest* dalam menghasilkan *F1-score* sebesar 98% pada kelas *prediabetes* memiliki implikasi penting dalam konteks medis.

Fase *prediabetes* merupakan tahap kritis yang memungkinkan dilakukannya intervensi dini guna mencegah progresi menuju diabetes tipe 2. Oleh karena itu, model klasifikasi yang mampu mengenali kondisi ini secara akurat akan sangat berguna dalam mendukung program skrining dan pengambilan keputusan klinis berbasis data (*data-driven decision-making*).

Sebaliknya, *Logistic Regression* menunjukkan keterbatasan signifikan, terutama karena merupakan model linier yang cenderung gagal dalam menangkap hubungan kompleks antar fitur kesehatan seperti tekanan darah, indeks massa tubuh, aktivitas fisik, dan kebiasaan merokok. Model ini hanya menghasilkan akurasi 52% dan memiliki performa rendah pada kelas *prediabetes*, dengan banyak kasus yang salah diklasifikasikan sebagai kelas lain. Hasil penelitian ini konsisten dengan studi sebelumnya yang juga menekankan keunggulan algoritma berbasis pohon keputusan dalam domain medis. Mandal et al. [12] menunjukkan bahwa kombinasi pendekatan *multi-class classification* menggunakan strategi *One-vs-Rest* dan metode *SMOTE* bersama *Random Forest* menghasilkan performa akurat dan stabil untuk klasifikasi penyakit kronis.

Dalam penelitian tersebut, *Random Forest* mampu mengatasi ketidakseimbangan kelas dan menghasilkan *F1-score* yang kompetitif. Studi lainnya oleh Liu et al. [22] memperkuat temuan bahwa algoritma *Random Forest* unggul pada prediksi rawat ulang pasien diabetes. Penelitian ini menunjukkan *Random Forest* tampil paling optimal dengan *F1-score* sekitar 83% dan akurasi 88%, mengungguli model lain seperti *XGBoost* dan *SVM* dalam data lebih dari 100.000 rekam medis dari 130 rumah sakit AS, diolah dengan strategi pemilihan fitur *GWO*.

Selain itu, validasi tambahan menggunakan teknik *Stratified K-Fold Cross Validation* sebanyak 5 *fold* dalam penelitian ini menunjukkan bahwa *Random Forest* tidak hanya unggul dari segi performa rata-rata, tetapi juga dari sisi konsistensi, dengan standar deviasi yang sangat rendah (0.0005). Hal ini membuktikan bahwa model tidak hanya baik pada subset tertentu, namun juga *robust* terhadap variasi data, sehingga layak diterapkan dalam lingkungan nyata (*real-world applications*). Dengan mempertimbangkan semua aspek di atas, dapat disimpulkan bahwa penerapan strategi *One-vs-Rest* bersama algoritma *Random Forest* dan penyeimbangan data menggunakan *SMOTE* merupakan pendekatan yang sangat efektif dan andal dalam klasifikasi multikelas berbasis data medis, khususnya untuk prediksi tingkat diabetes.

4. KESIMPULAN

Penelitian ini berhasil melakukan komparasi model klasifikasi multikelas tingkat diabetes berdasarkan indikator kesehatan pasien menggunakan pendekatan *One-vs-Rest (OvR)* dengan algoritma *Random Forest* sebagai metode utama. Dengan memanfaatkan dataset *Behavioral Risk Factor Surveillance System (BRFSS) 2015*, model ini mampu mengklasifikasikan pasien ke dalam tiga kategori yaitu non-diabetes, *prediabetes*, dan diabetes dengan mempertimbangkan indikator kesehatan seperti indeks massa tubuh, tekanan darah tinggi, kadar kolesterol, aktivitas fisik, dan status merokok. Hasil pengujian menunjukkan bahwa kombinasi *SMOTE*, *OvR*, dan *Random Forest* menghasilkan akurasi, presisi, *recall*, dan *F1-score* konsisten di angka 93%, dengan kemampuan mendeteksi kasus *prediabetes* mencapai 98% yang sangat penting dalam pencegahan dini diabetes. *Logistic Regression* digunakan sebagai pembanding, dan hanya mencapai akurasi 52% serta *F1-score* 51%, sehingga menegaskan keunggulan *Random Forest*. Evaluasi tambahan dengan *Stratified K-Fold Cross Validation* memperlihatkan performa *Random Forest* yang stabil, sementara *SMOTE* terbukti efektif mengatasi ketidakseimbangan kelas. Dengan demikian, dapat disimpulkan bahwa kombinasi *SMOTE + OvR + Random Forest* merupakan pendekatan yang efektif dan andal untuk klasifikasi multikelas diabetes, meskipun penelitian ini memiliki keterbatasan karena dataset yang digunakan berasal dari populasi Amerika Serikat, sehingga penerapannya di konteks Indonesia perlu dikaji lebih lanjut dengan data lokal, serta model yang dikembangkan belum diimplementasikan dalam sistem nyata seperti aplikasi mobile atau platform berbasis web. Penelitian lanjutan direkomendasikan untuk mengintegrasikan model ke dalam sistem aplikasi yang dapat dimanfaatkan oleh tenaga kesehatan maupun masyarakat secara langsung.

REFERENCES

- [1] A. B. Pulungan, G. Fadiana, and D. Annisa, "Type 1 diabetes mellitus in children: Experience in Indonesia," *Clin. Pediatr. Endocrinol.*, vol. 30, no. 1, pp. 11–18, 2021, doi: 10.1297/cpe.30.11.
- [2] C. Boutari and C. S. Mantzoros, "A 2022 update on the epidemiology of obesity and a call to action: as its twin COVID-19 pandemic appears to be receding, the obesity and dysmetabolism pandemic continues to rage on," *Metabolism.*, vol. 133, 2022, doi: 10.1016/j.metabol.2022.155217.
- [3] C. E. Franks and A. M. Gronowski, "The Faces of Women in AACC - A Historical Reflection on AACC Leadership and Awards," *Clin. Chem.*, vol. 67, no. 2, pp. 449–453, 2021, doi: 10.1093/clinchem/hvaa253.
- [4] D. Care and S. S. Suppl, "Introduction : Standards of Medical Care in Diabetes — 2022," vol. 45, no. December 2021, pp. 2021–2022, 2022.
- [5] R. M. M. Khan, Z. J. Y. Chua, J. C. Tan, Y. Yang, Z. Liao, and Y. Zhao, "From pre-diabetes to diabetes: Diagnosis, treatments and translational research," *Med.*, vol. 55, no. 9, pp. 1–30, 2019, doi: 10.3390/medicina55090546.
- [6] M. Faizi, G. Fadiana, D. Nadira, A. Angela, H. A. Puteri, and A. Pulungan, "Pediatric Type 1 Diabetes Care in Indonesia: A Review of Current Challenges and Practice," *J. Clin. Res. Pediatr. Endocrinol.*, pp. 1–6, 2024, doi: 10.4274/jcrpe.galenos.2024.2024-9-4.
- [7] F. Dany *et al.*, "Urban-rural distinction of potential determinants for *prediabetes* in Indonesian population aged ≥ 15 years: A cross-sectional analysis of Indonesian Basic Health Research 2018 among normoglycemic and prediabetic individuals," *BMC Public Health*, vol. 20, no. 1, pp. 1–9, 2020, doi: 10.1186/s12889-020-09592-7.
- [8] M. R. Rooney *et al.*, "Risk of Progression to Diabetes among Older Adults with *Prediabetes*," *JAMA Intern. Med.*, vol. 181, no. 4, pp. 511–519, 2021, doi: 10.1001/jamainternmed.2020.8774.
- [9] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 104–116, 2017, doi: 10.1016/j.csbj.2016.12.005.
- [10] M. Faisal and I. B. Santoso, "Algoritma Random Forest dan Synthetic Minority Oversampling Technique (*SMOTE*) untuk Deteksi Diabetes," vol. 10, no. 2, pp. 223–234, 2025.
- [11] A. Wibowo, A. F. N. Masruriyah, and S. Rahmawati, "Refining Diabetes Diagnosis Models: The Impact of *SMOTE* on SVM, Logistic Regression, and Naïve Bayes for Imbalanced Datasets," *J. Electron. Electromed. Eng. Med. Informatics*, vol. 7, no. 1, pp. 197–207, 2025, doi: 10.35882/jeeemi.v7i1.596.
- [12] X. Wang *et al.*, "Exploratory study on classification of diabetes mellitus through a combined Random Forest Classifier," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, pp. 1–14, 2021, doi: 10.1186/s12911-021-01471-4.
- [13] M. F. Rabby, Y. Tu, M. I. Hossen, I. Lee, A. S. Maida, and X. Hei, "Stacked LSTM based deep recurrent neural network with kalman smoothing for blood glucose prediction," *BMC Med. Inform. Decis. Mak.*, vol. 21, no. 1, pp. 1–15, 2021, doi: 10.1186/s12911-021-01462-5.
- [14] H. Naz and S. Ahuja, "Deep learning approach for diabetes prediction using PIMA Indian dataset," *J. Diabetes Metab. Disord.*, vol. 19, no. 1, pp. 391–403, 2020, doi: 10.1007/s40200-020-00520-5.
- [15] B. Zgardzińska *et al.*, "Studies on healthy and neoplastic tissues using positron annihilation lifetime spectroscopy and focused histopathological imaging," *Sci. Rep.*, vol. 10, no. 1, pp. 1–10, 2020, doi: 10.1038/s41598-020-68727-3.
- [16] S. Rahmawati, A. Wibowo, and A. F. N. Masruriyah, "Improving Diabetes Prediction Accuracy in Indonesia: A Comparative Analysis of SVM, Logistic Regression, and Naive Bayes with *SMOTE* and ADASYN," *J. RESTI*, vol. 8, no. 5, pp. 607–614, 2024, doi: 10.29207/resti.v8i5.5980.

- [17] K. M. Kuo, P. Talley, Y. H. Kao, and C. H. Huang, “A multi-class classification model for supporting the diagnosis of type II diabetes mellitus,” *PeerJ*, vol. 8, pp. 1–24, 2020, doi: 10.7717/peerj.9920.
- [18] D. D. Rufo, T. G. Debelee, A. Ibenthal, and W. G. Negera, “Diagnosis of diabetes mellitus using gradient boosting machine (Lightgbm),” *Diagnostics*, vol. 11, no. 9, pp. 1–14, 2021, doi: 10.3390/diagnostics11091714.
- [19] M. E. Shipe, S. A. Deppen, F. Farjah, and E. L. Grogan, “Developing prediction models for clinical use using logistic regression: An overview,” *J. Thorac. Dis.*, vol. 11, no. Suppl 4, pp. S574–S584, 2019, doi: 10.21037/jtd.2019.01.25.
- [20] J. K. Harris, “Primer on binary logistic regression,” *Fam. Med. Community Heal.*, vol. 9, pp. 1–7, 2021, doi: 10.1136/fmch-2021-001290.
- [21] P. F. Pratama, D. Rahmadani, and R. S. Nahampun, “Random Forest Optimization Using Particle Swarm Optimization for Diabetes Classification,” vol. 1, no. July, pp. 41–46, 2023.
- [22] V. B. Liu, L. Y. Sue, and Y. Wu, “Comparison of machine learning models for predicting 30-day readmission rates for patients with diabetes,” pp. 0–1, 2024, doi: 10.21037/jmai-24-70.