

**OPTIMASI SELEKSI FITUR MENGGUNAKAN *WRAPPER* DAN
PARTICLE SWARM OPTIMIZATION PADA *RANDOM FOREST* UNTUK
PREDIKSI RISIKO DIABETES TAHAP AWAL**

SKRIPSI

Diajukan Untuk Memenuhi Sebagian Syarat Guna
Memperoleh Gelar Sarjana Komputer (S.Kom.)
Pada Program Studi Sistem Informasi



OLEH :

MEDIANA NURLAILLI

NPM: 2213030130

FAKULTAS TEKNIK DAN ILMU KOMPUTER (FTIK)
UNIVERSITAS NUSANTARA PERSATUAN GURU REPUBLIK INDONESIA
UN PGRI KEDIRI

2026

Skripsi oleh :

MEDIANA NURLAILI

NPM : 2213030130

Judul

**OPTIMASI SELEKSI FITUR MENGGUNAKAN *WRAPPER* DAN
PARTICLE SWARM OPTIMIZATION PADA *RANDOM FOREST* UNTUK
PREDIKSI RISIKO DIABETES TAHAP AWAL**

Telah disetujui untuk diajukan Kepada
Panitia Ujian Sidang Skripsi Program Studi Sistem Informasi
FTIK UN PGRI Kediri

Tanggal: 25 Juni 2026

Pembimbing I



Dr. Sucipto, M.Kom

NIDN. 0721029101

Pembimbing II



Anita Sari Wardani, M.Kom

NIDN. 0713018402

Skripsi oleh :

MEDIANA NURLAILLI

NPM : 2213030130

Judul

**OPTIMASI SELEKSI FITUR MENGGUNAKAN *WRAPPER* DAN
PARTICLE SWARM OPTIMIZATION PADA *RANDOM FOREST* UNTUK
PREDIKSI RISIKO DIABETES TAHAP AWAL**

Telah dipertahankan di depan Panitia Ujian/Sidang Skripsi

Program Studi Sistem Informasi FTIK UN PGRI Kediri

Tanggal: 26 Juni 2026

Dan Dinyatakan telah Memenuhi Persyaratan

Panitia Penguji :

1. Ketua : Dr. Sucipto, M.Kom

[.....]

2. Penguji 1 : Muhammad Najibulloh Muzaki, M.Cs

[.....]

3. Penguji 2 : Anita Sari Wardani, M.Kom

[.....]

Mengetahui,

Dekan Fakultas Teknik dan Ilmu Komputer



Dr. Sulistiono, M.Si

NIDN.0007076801

PERNYATAAN

Yang bertanda tangan di bawah ini saya,

Nama : Mediana Nurlaili
Jenis Kelamin : Perempuan
Tempat/tgl. lahir : Bangkalan, 16 Juli 2002
NPM : 2213030130
Fak/Prodi : FTIK/ S1-Sistem Informasi

menyatakan dengan sebenarnya, bahwa dalam Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu perguruan tinggi, dan sepanjang pengetahuan saya tidak terdapat karya tulis atau pendapat yang pernah diterbitkan oleh orang lain, kecuali yang secara sengaja dan tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Kediri, 23 Juni 2026

Yang Menyatakan



Mediana Nurlaili

NPM: 2213030130

MOTTO

"Kita selalu bisa memilih: bersabar atau marah, bersyukur atau ingkar."

— *Bahar Safar, terinspirasi dari Janji* —

ABSTRAK

Mediana Nurlailli: Optimasi Seleksi Fitur Menggunakan Wrapper Dan Particle Swarm Optimization Pada Random Forest Untuk Prediksi Risiko Diabetes Pada Dua Dataset Diabetes, Skripsi, Program Studi Sistem Informasi, Fakultas Teknik dan Ilmu Komputer, Universitas Nusantara PGRI Kediri, 2026.

Kata Kunci: *Random Forest, Seleksi Fitur, Sequential Forward Selection, Sequential Backward Selection, Particle Swarm Optimization, Diabetes.*

Tingginya prevalensi diabetes secara global mendorong kebutuhan akan sistem deteksi dini berbasis *machine learning* yang akurat dan efisien. Namun, keberadaan fitur yang tidak relevan dalam dataset medis dapat menurunkan kinerja model klasifikasi secara signifikan. Penelitian ini mengkaji pengaruh tiga metode seleksi fitur, yaitu *Sequential Forward Selection* (SFS), *Sequential Backward Selection* (SBS), dan *Particle Swarm Optimization* (PSO), terhadap kinerja algoritma *Random Forest* pada dua dataset diabetes, yaitu *Early Stage Diabetes Risk Prediction* Dataset dari UCI (520 data, 16 fitur) dan *Pima Indians Diabetes* Dataset (768 data, 8 fitur). Pengujian dilakukan menggunakan skema 10-fold cross validation.

Hasil pengujian pada Dataset 1 menunjukkan peningkatan akurasi dari 98,27% (tanpa seleksi fitur, 16 fitur) menjadi 98,65% dengan SFS, 99,23% dengan SBS, dan 98,85% dengan PSO, masing-masing menggunakan 11 fitur terpilih. Pada Dataset 2, ketiga metode seleksi fitur secara konsisten meningkatkan akurasi dari 76,30% menjadi 77,47% dengan mengurangi jumlah fitur dari 8 menjadi 6. Hasil penelitian membuktikan bahwa penerapan seleksi fitur mampu meningkatkan performa *Random Forest* sekaligus mengurangi dimensi data. Metode SBS menghasilkan akurasi tertinggi pada Dataset 1, sedangkan ketiga metode menghasilkan performa yang setara pada Dataset 2.

KATA PENGANTAR

Puji Syukur Kami panjatkan Kehadirat Allah Tuhan Yang Maha Kuasa, karena penyusunan skripsi ini dapat diselesaikan. Skripsi dengan judul “Optimasi Seleksi Fitur Menggunakan *Wrapper* Dan *Particle Swarm Optimization* Pada *Random Forest* Untuk Prediksi Risiko Diabetes” ini ditulis guna memenuhi sebagian syarat untuk memperoleh gelar Sarjana Komputer. Pada kesempatan ini diucapkan terima kasih dan penghargaan yang setulus-tulusnya kepada:

1. Dr. Zainal Afandi, M.Pd selaku Rektor UN PGRI.
2. Dr. Sulistiono, M.Si selaku Dekan Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri.
3. Dr. Sucipto, M.Kom selaku Ka Prodi Sistem Informasi Fakultas Teknik dan Ilmu Komputer UN PGRI Kediri.
4. Dr. Sucipto, M.Kom dan Anita Sari Wardani, M.Kom selaku Pembimbing skripsi yang dengan sabar membimbing dalam menyelesaikan skripsi ini.
5. Kedua orang tua, Saudara yang telah mendukung dalam do’a dan motivasi dalam penyelesaian skripsi ini.
6. Teman-teman yang selalu membantu memberi semangat dalam penyelesaian skripsi ini.

Disadari bahwa skripsi ini masih banyak kekurangan, maka diharapkan tegur sapa, kritik dan saran-saran, dari berbagai pihak sangat diharapkan.

Kediri, 25 Juni 2026



Mediana Nurlailli
NPM: 2213030130

DAFTAR ISI

Halaman Sampul	i
Halaman Persetujuan.....	i
Halaman Pengesahan	ii
Halaman Pernyataan.....	iv
Halaman Motto.....	iv
ABSTRAK	v
KATA PENGANTAR.....	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR.....	x
DAFTAR TABEL	x
DAFTAR LAMPIRAN	xii
BAB I PENDAHULUAN.....	1
A. Latar Belakang	1
B. Batasan Masalah.....	7
C. Rumusan Masalah.....	7
D. Tujuan Penelitian	7
E. Manfaat Penelitian	8
BAB II KAJIAN PUSTAKA	9
A. Kajian Teori.....	9
B. Kajian Hasil Penelitian Terdahulu	19
C. Kerangka Berpikir.....	23
BAB III METODE PENELITIAN	26
BAB IV HASIL DAN PEMBAHASAN	37
BAB V PENUTUP.....	74
A. Kesimpulan	74
B. Saran.....	74
DAFTAR PUSTAKA	76

LAMPIRAN.....	79
---------------	----

DAFTAR GAMBAR

Gambar 2.1. Kerangka Berfikir.....	Error! Bookmark not defined.
Gambar 3.1 Alur Penelitian CRISP-DM.....	26
Gambar 4.1 Distribusi Fitur dan Korelasi Early.....	44
Gambar 4.2 Distribusi Fitur dan Korelasi Early.....	45
Gambar 4.3 Distribusi Fitur dan Korelasi Pima.....	46
Gambar 4.4 Kurva Akurasi dan jumlah fitur SFS Early	49
Gambar 4.5 Kurva Akurasi dan Jumlah Fitur SFS Pima.....	51
Gambar 4.6 Kurva Akurasi dan Jumlah Fitus SBS Early.....	54
Gambar 4.7 Kurva Akurasi dan Jumlah Fitur SBS Pima.....	56
Gambar 4.7 Kurva Akurasi dan Jumlah Fitur SBS Pima.....	56
Gambar 4.8 Kurva Konvergensi PSO Early dan Pima.....	57
Gambar 4.9 Heatmap Fitur Terpilih Eraly.....	58
Gambar 4.10 Heatmap Fitur Terpilih Pima.....	59
Gambar 4.11 Perbandingan Confusion Matrix Early.....	61
Gambar 4.12 Perbandingan Confusion Matrix Pima.....	62
Gambar 4.13 Perbandingan Jumlah Fitur Early.....	66
Gambar 4.14 Perbandingan Jumlah Fitur Pima.....	67
Gambar 4.15 Perbandingan Akurasi Kedua Dataset.....	68
Gambar 4.16 Perbandingan Precision, Recall, dan f1-Score.....	69

DAFTAR TABEL

Tabel 2.1. Tabel Confuxion Matrix	12
Tabel 2.2. Tabel Hipotesis	24
Tabel 3.1. Tabel Deskripsi Fitur Dataset Early.....	28
Tabel 3.2. Tabel Deskripsi Fitur Dataset Pima	29
Tabel 3.3. Tabel Perbandingan Metode	35
Tabel 4.1. Tabel Data Awal Uci.....	36
Tabel 4.2. Tabel Tipe dan informasi	37
Tabel 4.3. Tabel Deskripsi Dataset	39
Tabel 4.4. Tabel Data Awal Pima	41
Tabel 4.5. Tabel Tipe dan Informasi	42
Tabel 4.6. Tabel Deskripsi Statistik Dataset2	43
Tabel 4.7. Tabel Proses Seleksi Fitur SFS Early.....	48
Tabel 4.8. Tabel Proses Seleksi Fitur SFS Pima.....	50
Tabel 4.9. Tabel Proses Seleksi Fitur SBS Early	53
Tabel 4.10. Tabel Proses Seleksi Fitur SBS Pima	55
Tabel 4.11. Tabel Matrix Evaluasi Metode Early	63
Tabel 4.12. Tabel Metrix Evaluasi Metode Pima	64
Tabel 4.13. Tabel Rekapitulasi Hasil Perbandingan	71

DAFTAR LAMPIRAN

Lampiran 1 Kartu Bimbingan Skripsi	77
Lampiran 2 Surat Keterangan Bebas Similarity.....	79
Lampiran 3 Bukti Halaman Awal Cek Similarity.....	81
Lampiran 4 Bukti Submit Artikel/Loa/Terbit	82
Lampiran 5 Lembar Berita Acara Ujian.....	83
Lampiran 6 Lembar Revisi Ujian.....	84

BAB I

PENDAHULUAN

A. Latar Belakang

Diabetes melitus merupakan salah satu penyakit metabolik kronis yang prevalensinya terus meningkat secara signifikan di tingkat global. Penyakit ini ditandai oleh gangguan regulasi kadar glukosa darah dalam jangka panjang yang disebabkan oleh gangguan produksi maupun fungsi insulin dalam tubuh. Kondisi tersebut berpotensi menimbulkan berbagai komplikasi serius seperti penyakit *kardiovaskular*, gagal ginjal, *neuropati*, hingga gangguan penglihatan yang dapat menurunkan kualitas hidup penderita secara signifikan. Selain itu, peningkatan jumlah penderita diabetes juga berdampak langsung pada meningkatnya beban biaya kesehatan serta tekanan terhadap sistem pelayanan kesehatan. Dalam jangka panjang, kondisi ini tidak hanya mempengaruhi individu, tetapi juga memberikan dampak luas terhadap produktivitas masyarakat dan stabilitas sistem kesehatan suatu negara. Oleh karena itu, diabetes menjadi salah satu permasalahan kesehatan global yang memerlukan perhatian serius, terutama dalam upaya pencegahan dan pengendalian sejak dini (World Health Organization, 2023).

Urgensi deteksi dini risiko diabetes menjadi semakin penting karena sebagian besar komplikasi yang ditimbulkan oleh penyakit ini berkembang secara bertahap dan sering kali baru terdeteksi pada tahap lanjut. Kondisi tersebut menyebabkan penanganan menjadi lebih kompleks serta membutuhkan biaya yang lebih besar. Deteksi dini memungkinkan dilakukannya *intervensi preventif* seperti perubahan pola hidup dan penanganan medis awal yang lebih efektif dalam menekan risiko komplikasi. Selain itu, deteksi dini juga memberikan peluang bagi tenaga medis untuk melakukan pemantauan kondisi pasien secara lebih optimal sebelum terjadi kondisi yang lebih serius. Namun demikian, metode *konvensional* dalam praktik klinis masih cenderung bersifat reaktif dan bergantung pada pemeriksaan laboratorium setelah gejala muncul.

Pendekatan ini dinilai kurang efektif dalam mengidentifikasi individu yang berisiko sejak tahap awal, sehingga diperlukan metode yang lebih proaktif dan berbasis data untuk mendukung proses deteksi dini secara lebih akurat (Sneha & Gangil, 2019).

Seiring dengan perkembangan teknologi informasi, sektor kesehatan mengalami transformasi melalui digitalisasi data medis dan rekam kesehatan pasien. Data kesehatan saat ini tersimpan dalam jumlah besar dengan karakteristik *multivariat* yang mencakup data *demografis*, gejala klinis, serta hasil pemeriksaan medis. Ketersediaan data dalam jumlah besar ini membuka peluang untuk dilakukan analisis lebih lanjut menggunakan pendekatan komputasi guna menemukan pola tersembunyi yang tidak mudah diidentifikasi secara manual. Dalam penelitian ini digunakan *Early Stage Diabetes Risk Prediction Dataset* yang diperoleh dari *UCI Machine Learning Repository*, sebagaimana digunakan dalam penelitian oleh (Nugroho et al. 2021). Dataset ini terdiri dari 520 data dengan 16 fitur tanpa nilai hilang, sehingga cukup *representatif* untuk pengembangan model prediksi risiko diabetes tahap awal. Dataset kedua adalah *Pima Indians Diabetes Dataset* yang terdiri dari 768 data dengan 8 fitur numerik dan telah lama menjadi tolok ukur dalam penelitian prediksi diabetes. Penggabungan dua dataset dengan karakteristik berbeda ini bertujuan untuk menguji konsistensi dan kemampuan generalisasi metode seleksi fitur yang diterapkan (Kaliappan et al., 2024).

Perkembangan metode *machine learning* telah memberikan kontribusi yang signifikan dalam bidang kesehatan, khususnya dalam membantu proses diagnosis dan prediksi berbagai penyakit secara lebih cepat dan akurat. Berbagai algoritma klasifikasi telah dikembangkan untuk mengolah data medis sehingga mampu menghasilkan model prediksi yang memiliki tingkat akurasi tinggi serta dapat digunakan sebagai sistem pendukung keputusan dalam diagnosis penyakit diabetes (Ansari et al., 2025). Selain itu, *machine learning* juga mampu mengolah data dalam jumlah besar dengan tingkat kompleksitas yang tinggi sehingga dapat

menghasilkan model yang lebih adaptif terhadap berbagai kondisi data. Beberapa penelitian menunjukkan bahwa algoritma *machine learning* mampu memberikan tingkat akurasi yang lebih tinggi dibandingkan metode tradisional dalam prediksi diabetes. Dengan demikian, penerapan *machine learning* menjadi salah satu solusi yang menjanjikan dalam meningkatkan kualitas layanan kesehatan berbasis data (Khan et al., 2021).

Meskipun demikian, dataset medis umumnya memiliki karakteristik yang kompleks dengan kombinasi fitur kategorikal dan numerik yang heterogen. Permasalahan seperti adanya noise, redundansi fitur, serta ketidakseimbangan kontribusi antar fitur terhadap variabel target sering kali muncul dalam data medis. Kondisi ini dapat menyebabkan model mengalami kesulitan dalam menangkap pola yang relevan sehingga berdampak pada penurunan performa model prediksi. Selain itu, adanya fitur yang tidak relevan juga dapat memperbesar ruang pencarian solusi sehingga proses pembelajaran menjadi kurang efisien. Oleh karena itu, diperlukan strategi yang tepat dalam mengelola data agar model yang dihasilkan mampu bekerja secara optimal dan memberikan hasil prediksi yang akurat (Patel, 2023).

Penggunaan seluruh fitur tanpa melalui proses seleksi cenderung meningkatkan kompleksitas komputasi serta risiko terjadinya *overfitting*. Model yang mengalami *overfitting* biasanya memiliki performa yang sangat baik pada data latih, tetapi tidak mampu melakukan generalisasi dengan baik pada data baru. Hal ini tentu menjadi kendala dalam penerapan model pada kondisi nyata karena hasil prediksi menjadi kurang dapat diandalkan. Selain itu, penggunaan fitur yang berlebihan juga dapat memperlambat proses komputasi dan meningkatkan kebutuhan sumber daya sistem. Penelitian menunjukkan bahwa pemilihan fitur yang tidak optimal dapat menurunkan kinerja model prediksi meskipun menggunakan algoritma klasifikasi yang kuat (Sneha & Gangil, 2019).

Seleksi fitur bertujuan untuk menghilangkan atribut yang kurang relevan sehingga model dapat bekerja lebih efisien tanpa mengurangi

bahkan dapat meningkatkan performa klasifikasi. Proses ini juga membantu mengurangi kompleksitas model dan mempercepat waktu komputasi (Idris, 2025). Hal ini sangat penting dalam konteks kesehatan karena hasil model perlu dipahami oleh tenaga medis. Metode seleksi fitur berbasis *wrapper* merupakan salah satu pendekatan yang efektif karena mengevaluasi subset fitur berdasarkan kinerja model klasifikasi. Pendekatan ini memungkinkan pemilihan fitur yang benar-benar berkontribusi terhadap peningkatan performa model. Penelitian sebelumnya menunjukkan bahwa metode *wrapper* mampu meningkatkan akurasi *Random Forest* dalam prediksi diabetes dibandingkan tanpa seleksi fitur, meskipun metode ini memiliki keterbatasan dalam pencarian solusi global (Nugroho et al., 2021).

Sequential Forward Selection (SFS) dan *Sequential Backward Selection* (SBS) merupakan metode *wrapper* yang bekerja secara bertahap dalam memilih fitur terbaik. SFS memulai dari himpunan kosong dan menambahkan fitur satu per satu, sedangkan SBS memulai dari seluruh fitur dan menghapus fitur yang kurang relevan secara bertahap. Kedua metode ini telah terbukti mampu meningkatkan akurasi model dengan jumlah fitur yang lebih efisien. Namun demikian, sifat *greedy* yang dimiliki menyebabkan proses seleksi fitur hanya berfokus pada solusi lokal sehingga berpotensi tidak menemukan kombinasi fitur terbaik secara global. Hal ini menjadi salah satu keterbatasan utama dari metode seleksi fitur *konvensional* yang perlu diperhatikan dalam pengembangan model prediksi (Kaliappan et al., 2024).

Particle Swarm Optimization (PSO) sebagai metode *metaheuristic* berbasis *swarm intelligence* menawarkan pendekatan yang berbeda dalam proses seleksi fitur. PSO bekerja dengan memanfaatkan populasi partikel yang bergerak dalam ruang solusi untuk mencari kombinasi fitur terbaik berdasarkan nilai *fitness* tertentu. Pendekatan ini memungkinkan eksplorasi ruang solusi secara lebih luas dibandingkan metode *konvensional* sehingga berpotensi menghasilkan subset fitur yang lebih optimal. Selain itu, PSO juga memiliki kemampuan dalam menyeimbangkan proses eksplorasi dan

eksploitasi sehingga solusi yang dihasilkan cenderung lebih stabil. Beberapa penelitian menunjukkan bahwa integrasi PSO mampu meningkatkan performa model prediksi diabetes secara signifikan (Rahman et al., 2025).

Random Forest merupakan salah satu algoritma *ensemble learning* yang banyak digunakan karena mampu menangani data berdimensi tinggi, mengurangi risiko *overfitting*, serta menghasilkan performa klasifikasi yang stabil pada berbagai jenis dataset medis. Keunggulan tersebut menjadikan *Random Forest* sebagai salah satu algoritma yang sering diterapkan dalam penelitian prediksi diabetes (Deberneh & Kim, 2021). Algoritma ini bekerja dengan menggabungkan beberapa pohon keputusan untuk meningkatkan akurasi dan stabilitas model. Kombinasi antara *Random Forest* dan metode seleksi fitur yang tepat dapat meningkatkan kinerja model secara signifikan. Namun demikian, penelitian yang membandingkan secara komprehensif metode SFS, SBS, dan PSO dalam konteks *Random Forest* untuk prediksi diabetes tahap awal masih terbatas, sehingga diperlukan penelitian lebih lanjut untuk mengisi kesenjangan tersebut (Khan et al., 2021).

Berdasarkan kajian tersebut, dapat diidentifikasi adanya kesenjangan penelitian pada aspek perbandingan metode seleksi fitur berbasis *wrapper konvensional* dan *metaheuristik* dalam meningkatkan kinerja *Random Forest* pada prediksi risiko diabetes tahap awal. Penelitian ini memiliki kebaruan dalam integrasi dan evaluasi sistematis metode SFS, SBS, dan PSO untuk memperoleh subset fitur yang optimal. Dengan pendekatan ini, diharapkan dapat dihasilkan model prediksi yang tidak hanya memiliki tingkat akurasi yang tinggi, tetapi juga efisien dalam penggunaan fitur. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan metode *data mining* di bidang kesehatan, khususnya dalam meningkatkan kualitas prediksi risiko diabetes tahap awal (Nugroho et al., 2021).

Dengan demikian, selain meningkatkan akurasi, proses optimasi melalui metode seleksi fitur juga dapat menghasilkan model yang lebih

sederhana karena hanya menggunakan atribut-atribut penting. Model yang lebih sederhana umumnya memiliki kemampuan generalisasi yang lebih baik ketika diterapkan pada data baru (Natarajan et al., 2025). Penelitian ini untuk menganalisis serta membandingkan efektivitas metode seleksi fitur *Sequential Forward Selection* (SFS), *Sequential Backward Selection* (SBS), dan *Particle Swarm Optimization* (PSO) dalam meningkatkan kinerja algoritma *Random Forest* pada prediksi risiko diabetes tahap awal. Selain itu, penelitian ini juga bertujuan untuk mengetahui kombinasi fitur yang paling optimal dalam menghasilkan model klasifikasi yang akurat dan efisien. Dari segi akademis, penelitian ini diharapkan dapat memperkaya kajian mengenai penerapan metode seleksi fitur berbasis *wrapper* dan *metaheuristik* dalam bidang kesehatan, khususnya pada pengolahan data medis berbasis *machine learning*. Lebih lanjut, hasil penelitian ini diharapkan dapat menjadi dasar bagi pengembangan penelitian lanjutan pada kasus penyakit lain yang juga memerlukan deteksi dini berbasis data, seperti penyakit jantung, kanker, maupun gangguan metabolik lainnya. Dengan semakin berkembangnya penelitian di bidang ini, diharapkan dapat mendorong terciptanya sistem prediksi medis yang lebih adaptif, akurat, dan mampu mendukung pengambilan keputusan secara lebih cerdas di masa mendatang.

B. Identifikasi Masalah

1. Model prediksi risiko diabetes tahap awal berbasis *machine learning* belum sepenuhnya optimal karena masih banyak menggunakan seluruh fitur dataset tanpa proses seleksi fitur yang tepat.
2. Penggunaan seluruh fitur pada dataset medis menimbulkan permasalahan *redundansi*, kompleksitas komputasi, serta risiko *overfitting* yang berdampak pada penurunan kinerja model klasifikasi.
3. Minimnya penelitian yang membandingkan SFS, SBS, dan PSO secara bersamaan pada dua dataset diabetes dengan karakteristik yang berbeda menggunakan algoritma *Random Forest*.

C. Batasan Masalah

1. Penelitian Penelitian menggunakan dua dataset: (1) *Early Stage Diabetes Risk Prediction* dari UCI (520 data, 16 fitur) dan (2) *Pima Indians Diabetes Dataset* (768 data, 8 fitur numerik).
2. Metode seleksi fitur yang dibandingkan adalah SFS, SBS, dan PSO, dengan Random Forest sebagai algoritma klasifikasi dan *10-fold cross validation* sebagai skema validasi.
3. Analisis difokuskan pada kinerja model klasifikasi berdasarkan metrik akurasi, *precision*, *recall*, dan *F1-score*, tanpa membahas implementasi klinis.

D. Rumusan Masalah

1. Berapa nilai *accuracy* algoritma *Random Forest* tanpa seleksi fitur pada masing-masing dataset diabetes?
2. Berapa perubahan nilai akurasi dan jumlah fitur yang dihasilkan oleh metode SFS, SBS, dan PSO pada masing-masing dataset?
3. Metode seleksi fitur manakah yang menghasilkan akurasi tertinggi dengan jumlah fitur paling efisien pada masing-masing dataset?

E. Tujuan Penelitian

1. Menganalisis Mengukur dan membandingkan kinerja algoritma Random Forest sebelum dan sesudah penerapan metode seleksi fitur SFS, SBS, dan PSO pada dua dataset diabetes.
2. Membandingkan Mengidentifikasi metode seleksi fitur yang memberikan peningkatan performa terbaik berdasarkan metrik evaluasi klasifikasi pada masing-masing dataset.
3. Menghasilkan model klasifikasi prediksi risiko diabetes tahap awal yang lebih akurat dan efisien melalui optimasi seleksi fitur pada dua dataset dengan karakteristik berbeda.

F. Manfaat Penelitian

1. Hasil penelitian ini diharapkan dapat memberikan kontribusi ilmiah dalam pengembangan ilmu data mining dan *machine learning* di bidang kesehatan, khususnya pada kajian seleksi fitur dan klasifikasi prediktif, dengan memperkaya literatur terkait penerapan dan perbandingan metode seleksi fitur *Sequential Forward Selection* (SFS), *Sequential Backward Selection* (SBS), dan *Particle Swarm Optimization* (PSO) pada algoritma Random Forest untuk prediksi risiko diabetes tahap awal.
2. Penelitian ini diharapkan dapat menjadi referensi akademik bagi peneliti, mahasiswa, dan akademisi dalam mengembangkan penelitian sejenis, terutama yang berkaitan dengan optimasi seleksi fitur dan evaluasi kinerja model klasifikasi berbasis data kesehatan, serta menjadi dasar bagi penelitian lanjutan dengan pendekatan atau dataset yang berbeda.
3. Hasil penelitian ini diharapkan dapat memberikan rekomendasi metode seleksi fitur yang paling optimal untuk meningkatkan kinerja model prediksi risiko diabetes tahap awal, sehingga dapat mendukung pengembangan sistem pendukung keputusan berbasis data di bidang kesehatan yang lebih akurat, efisien, dan berbasis bukti.

DAFTAR PUSTAKA

- Abd El-Latif, E. I., & Moneim, I. A. (2024). Exploring Feature Selection and Machine Learning Algorithms for Predicting Diabetes Disease. *International Journal of Intelligent Systems and Applications*, 16(1), 1–10. <https://doi.org/10.5815/ijisa.2024.01.01>
- Abnoosian, K., Farnoosh, R., & Behzadi, M. H. (2023). Prediction of diabetes disease using an ensemble of machine learning multi-classifier models. *BMC Bioinformatics*, 24(1), 1–24. <https://doi.org/10.1186/s12859-023-05465-z>
- Dahlan, D., & Pradipta, A. (2025). *Penerapan Data Mining Untuk Prediksi Resiko Diabetes Tahap Awal Menggunakan Metode Naive Bayes*. 1(2), 54–61.
- Devian, D., Nurul Sabrina, P., & Komarudin, A. (2024). Prediksi Penyakit Diabetes Dengan Metode K-Nearest Neighbor (Knn) Dan Seleksi Fitur Information Gain. *JATI (Jurnal Mahasiswa Teknik Informatika)*, 8(6), 11320–11326. <https://doi.org/10.36040/jati.v8i6.11364>
- Dutta, A., Hasan, M. K., Ahmad, M., Awal, M. A., Islam, M. A., Masud, M., & Meshref, H. (2022). Early Prediction of Diabetes Using an Ensemble of Machine Learning Models. *International Journal of Environmental Research and Public Health*, 19(19), 1–25. <https://doi.org/10.3390/ijerph191912378>
- Kaliappan, J., Saravana Kumar, I. J., Sundaravelan, S., Anesh, T., Rithik, R. R., Singh, Y., Vera-Garcia, D. V., Himeur, Y., Mansoor, W., Atalla, S., & Srinivasan, K. (2024). Analyzing classification and feature selection strategies for diabetes prediction across diverse diabetes datasets. *Frontiers in Artificial Intelligence*, 7. <https://doi.org/10.3389/frai.2024.1421751>
- Kangra, K., & Singh, J. (2024). A genetic algorithm-based feature selection approach for diabetes prediction. *IAES International Journal of Artificial Intelligence (IJ-AI)*, 13(2), 1489–1498. <https://doi.org/10.11591/ijai.v13.i2.pp1489-1498>
- Khan, F. A., Zeb, K., Al-Rakhami, M., Derhab, A., & Bukhari, S. A. C. (2021). Detection and Prediction of Diabetes Using Data Mining: A Comprehensive Review. *IEEE Access*, 9, 43711–43735. <https://doi.org/10.1109/ACCESS.2021.3059343>
- Musa, J., & Abdulazeez, A. M. (2023). A Review on Diabetes Classification Based on Machine Learning Algorithms. *Indonesian Journal of Computer Science*, 12(2), 284–301. <http://ijcs.stmikindonesia.ac.id/ijcs/index.php/ijcs/article/view/3135>
- Nugroho, A., Fanani, A. Z., & Shidik, G. F. (2021). *Proceedings, 2021 International Seminar on Application for Technology of Information and Communication (iSemantic): IT opportunities and creativities for digital innovation and communication within global pandemic : September 18th - 19th, 2021, Universit*. 179–183.

- Rahman, F., Hossain, S., Tiang, J. J., & Nahid, A. Al. (2025). Diabetes Prediction Using Feature Selection Algorithms and Boosting-Based Machine Learning Classifiers. *Diagnostics*, 15(20), 1–24. <https://doi.org/10.3390/diagnostics15202622>
- Sheta, A., Elashmawi, W. H., Al-Qerem, A., & Othman, E. S. (2024). Utilizing Various Machine Learning Techniques for Diabetes Mellitus Feature Selection and Classification. *International Journal of Advanced Computer Science and Applications*, 15(3), 1372–1384. <https://doi.org/10.14569/IJACSA.2024.01503134>
- Sneha, N., & Gangil, T. (2019). Analysis of diabetes mellitus for early prediction using optimal features selection. *Journal of Big Data*, 6(1). <https://doi.org/10.1186/s40537-019-0175-6>
- Swati D. Patel. (2023). A survey on Diabetes Prediction Models Using Data Mining Techniques: issues and challenges. *International Journal of Scientific Research in Science, Engineering and Technology*, 263–267. <https://doi.org/10.32628/ijrsrset23103208>
- Zhou, H., Xin, Y., & Li, S. (2023). A diabetes prediction model based on Boruta feature selection and ensemble learning. *BMC Bioinformatics*, 24(1), 1–34. <https://doi.org/10.1186/s12859-023-05300-5>
- Abd El-Latif, A. A., & Moneim, A. (2024). Machine Learning-Based Diabetes Classification Using Ensemble Models. *Healthcare Analytics*.
- Anggreini, N. L., Yuliana, A., Ramdan, D. S., & Al-Dayyeni, W. (2025). Improving Diabetes Prediction Performance Using Random Forest Classifier with Hyperparameter Tuning. *Jurnal Teknik Informatika (JUTIF)*, 6(4), 1847–1860. <https://doi.org/10.52436/1.jutif.2025.6.4.4755>
- Ansari, G. A., Shafi, S., Ansari, M. D., & Shadab, A. (2025). Advanced Supervised Machine Learning Methods for Precise Diabetes Mellitus Prediction Using Feature Selection. *Frontiers in Medicine*, 12. <https://doi.org/10.3389/fmed.2025.1620268>
- Deberneh, H. M., & Kim, I. (2021). Prediction of Type 2 Diabetes Based on Machine Learning Algorithm. *International Journal of Environmental Research and Public Health*, 18(6), 3317. <https://doi.org/10.3390/ijerph18063317>
- Hairani, H., Anggrawan, A., & Priyanto, D. (2023). Improvement Performance of the Random Forest Method on Unbalanced Diabetes Data Classification Using SMOTE-Tomek Link. *International Journal of Informatics and Visualization*, 7(1), 258–264.
- Idris, N. F. (2025). A Review of Feature Selection Methods on Diabetes Mellitus Classification. *International Journal on Advanced Science, Engineering and Information Technology*, 15(3), 686–692.
- Iskandar, R. F. N., Gutama, D. H., Wijaya, D. P., & Danianti, D. (2024). Klasifikasi Menggunakan Metode Random Forest untuk Awal Deteksi Diabetes Melitus Tipe 2. *Jurnal Teknik Industri Terintegrasi*, 7(3), 1620–1626. <https://doi.org/10.31004/jutin.v7i3.26916>

- Jawza, D. N., Mazdadi, M. I., Farmadi, A., Saragih, T. H., Kartini, D., & Abdullayev, V. (2025). Enhancing Diabetes Prediction Accuracy Using Random Forest and XGBoost with PSO and GA-Based Feature Selection. *Journal of Electronics, Electromedical Engineering, and Medical Informatics*, 7(2), 295–306.
- Natarajan, K., Baskaran, D., & Kamalanathan, S. (2025). An Adaptive Ensemble Feature Selection Technique for Model-Agnostic Diabetes Prediction. *Scientific Reports*, 15, 6907.
- Ogaden, G. B. A., & Dwidasmara, I. B. G. (2025). Perbandingan RFE dan SelectKBest untuk Klasifikasi Penyakit Diabetes dengan Random Forest. *Jurnal Nasional Teknologi Informasi dan Aplikasinya*, 3(3), 641–650. <https://doi.org/10.24843/JNATIA.2025.v03.i03.p19>
- Oladimeji, O. O., Oladimeji, A., & Oladimeji, O. (2021). Classification Models for Likelihood Prediction of Diabetes at Early Stage Using Feature Selection. *Applied Computing and Informatics*.
- Prajogo, V. J., & Narulita, S. (2025). Deteksi Penderita Diabetes dengan Algoritma Random Forest dan Backward Elimination. *Jurnal Rekam Medis dan Manajemen Informasi Kesehatan*, 5(1).
- Rahman, F., Hossain, S., Tiang, J. J., & Nahid, A. A. (2025). Diabetes Prediction Using Feature Selection Algorithms and Boosting-Based Machine Learning Classifiers. *Diagnostics*, 15(20), 2622. <https://doi.org/10.3390/diagnostics15202622>
- Wibowo, A., & Sutrisno. (2025). Integration of Random Forest, ADASYN, and SHAP for Diabetes Prediction and Interpretation. *Scientific Journal of Informatics*, 12(2).
- Zhou, X., et al. (2023). Comparative Analysis of Machine Learning Models for Diabetes Prediction. *Healthcare*.

6/23/26, 1:46 PM

ppi.unpkediri.ac.id/ppi/modul/admin/export/cetak_pdf.php?id=2982

SURAT KETERANGAN BEBAS PLAGIARISME



SURAT KETERANGAN BEBAS SIMILARITY

Ketua UPT Perpustakaan, Publikasi dan Inovasi Universitas Nusantara PGRI Kediri
menerangkan bahwa mahasiswa dengan identitas berikut:

Nama Mahasiswa : Mediana Nurlailli
NPM : 2213030130
Program Studi : S1-Sistem Informasi

Judul Karya Ilmiah:

“Optimasi Seleksi Fitur Menggunakan Wrapper dan Particle Swarm Optimization pada
Random Forest untuk Prediksi Risiko Diabetes Tahap Awal”

Dinyatakan sudah memenuhi syarat batas maksimal 30% *similarity* sesuai dengan
ketentuan yang berlaku pada setiap subbab naskah Laporan **Tugas Akhir/Skripsi/Tesis**
yang disusun.

Demikian Surat Keterangan ini kami berikan untuk dapat dipergunakan seperlunya.

Kediri, 23 Juni 2026
Ka UPT PPI,



Dr. Abdul Aziz Hunaifi, M.A



25% Overall Similarity

The combined total of all matches, including overlapping sources, for each database.

Filtered from the Report

- Bibliography

Top Sources

- 14%  Internet sources
- 9%  Publications
- 21%  Submitted works (Student Papers)

Integrity Flags

0 Integrity Flags for Review

Our system's algorithms look deeply at a document for any inconsistencies that would set it apart from a normal submission. If we notice something strange, we flag it for you to review.

A Flag is not necessarily an indicator of a problem. However, we'd recommend you focus your attention there for further review.

KARTU BIMBINGAN SEMINAR PROPOSAL



Nama : Mediana Nurlaili
NPM : 2213030130
Prodi : Sistem Informasi
Semester : GJL
Thn Akademik : 25/26

No.	Sem/TA	Tanggal	Dosen Pembimbing	Materi	Catatan
1	GJL 25/26	05 Oktober 2025	ARIE NUGROHO	Topik Penelitian	Bimbingan Awal Penentuan Topik.
2	GJL 25/26	13 Oktober 2025	ARIE NUGROHO	Judul Penelitian	Mencari Jurnal Utama.
3	GJL 25/26	30 Desember 2025	ARIE NUGROHO	BAB I	Bimbingan Awal Bab 1.
4	GJL 25/26	03 November 2025	ARIE NUGROHO	BAB I	.
5	GJL 25/26	04 November 2025	ARIE NUGROHO	BAB I	.
6	GJL 25/26	10 November 2025	ARIE NUGROHO	BAB II	Bimbingan Bab 2.
7	GJL 25/26	11 November 2025	ARIE NUGROHO	BAB II	Revisi Bab 2.
8	GJL 25/26	17 November 2025	ARIE NUGROHO	BAB II	Membuat Kerangka Berpikir.
9	GJL 25/26	18 November 2025	ARIE NUGROHO	BAB II	Acc bab 2.
10	GJL 25/26	24 November 2025	ARIE NUGROHO	BAB III	Bimbingan Bab 3.
11	GJL 25/26	25 November 2025	ARIE NUGROHO	BAB III	Revisi awal.
12	GJL 25/26	01 Desember 2025	ARIE NUGROHO	BAB III	.
13	GJL 25/26	02 Desember 2025	ARIE NUGROHO	BAB III	Acc Bab 3.
14	GJL 25/26	02 Desember 2025	ARIE NUGROHO	Daftar Pustaka	.
15	GJL 25/26	08 Desember 2025	ARIE NUGROHO	Daftar Pustaka	.
16	GJL 25/26	09 Desember 2025	ARIE NUGROHO	Daftar Pustaka	.
17	GNP 25/26	01 April 2026	SUCIPTO	Topik Penelitian	Topik penelitian yang diambil.
18	GNP 25/26	10 Juni 2026	ANITA SARI WARDANI	Topik Penelitian	Topik Penelitian yang diambil.
19	GNP 25/26	16 Juni 2026	SUCIPTO	Judul Penelitian	Pengecekan kembali judul skripsi.
20	GNP 25/26	26 Juni 2026	ANITA SARI WARDANI	Judul Penelitian	pengecekan judul skripsi.
21	GNP 25/26	30 Juni 2026	SUCIPTO	BAB I	pengecekan bab satu, daftar isi dan lain-lain
22	GNP 25/26	15 April 2026	ANITA SARI WARDANI	BAB I	pengecekan bab 1.
23	GNP 25/26	24 April 2026	SUCIPTO	BAB II	pengecekan bab 2.
24	GNP 25/26	06 Mei 2026	ANITA SARI WARDANI	BAB II	pengecekan bab 2.
25	GNP 25/26	11 Mei 2026	SUCIPTO	BAB III	pengecekan bab 3.
26	GNP 25/26	18 Mei 2026	ANITA SARI WARDANI	BAB III	pengecekan bab 3.
27	GNP 25/26	21 Mei 2026	SUCIPTO	BAB IV	mengelola data.
28	GNP 25/26	29 Mei 2026	ANITA SARI WARDANI	BAB IV	mengelola data.
29	GNP 25/26	25 Mei 2026	SUCIPTO	Publikasi	publikasi jurnal.
30	GNP 25/26	30 Juni 2026	ANITA SARI WARDANI	Publikasi	publikasi jurnal.

31	GNP 25/26	02 Juni 2026	SUCIPTO	BAB V	pengecekan bab 5.
32	GNP 25/26	20 Juni 2026	ANITA SARI WARDANI	BAB V	pengecekan bab 5.
33	GNP 25/26	17 Juni 2026	ANITA SARI WARDANI	BAB V	pengecekan bab 5.
34	GNP 25/26	19 Juni 2026	SUCIPTO	Daftar Pustaka	.
35	GNP 25/26	09 Juni 2026	SUCIPTO	Topik Penelitian	.
36	GNP 25/26	25 Juni 2026	ANITA SARI WARDANI	Publikasi	.
37	GNP 25/26	23 Mei 2026	ANITA SARI WARDANI	Publikasi	.

Kaprodi Sistem Informasi
Utama

Dosen Pembimbing Akademik

Dosen Pembimbing

Valid

Valid

Valid

(Dr. SUCIPTO, S.Kom, M.Kom)

(ARIE NUGROHO, S.Kom, M.M, M.Kom
(RINA FIRLIANA, M.Kom)

)



BINER

Jurnal Ilmiah Informatika dan Komputer

ISSN (print): 2828-0210 | ISSN (online): 2828-0229

<https://ojs.unsiq.ac.id/index.php/biner>

LETTER OF ACCEPTANCE
NO. 027/LOA/BINER/VII/2026

Dengan ini pengelola jurnal **Biner: Jurnal Ilmiah Informatika dan Komputer** dengan p-ISSN : 2828-0210, e-ISSN : 2828-0229, terindeks [SINTA](#), [Garuda](#), [Google Scholar](#) menerangkan bahwa:

Judul Artikel : **OPTIMASI SELEKSI FITUR PADA RANDOM FOREST UNTUK
PREDIKSI RISIKO DIABETES TAHAP AWAL MENGGUNAKAN SFS,
SBS, DAN PSO**

Penulis : Mediana Nurlaili, Sucipto, Anita Sari Wardani

Afiliasi : Universitas Nusantara PGRI Kediri

Telah memenuhi kriteria publikasi di jurnal **Biner: Jurnal Ilmiah Informatika dan Komputer** dan dinyatakan **DITERIMA** untuk dipublikasikan pada edisi Vol. 5, No. 2, Juli 2026. Artikel tersedia pada

Indexed by:



akhir bulan penerbitan secara online di laman jurnal

<https://ojs.unsiq.ac.id/index.php/biner>.

Demikian surat ini kami buat untuk dipergunakan sebagaimana mestinya.

Wonosobo, 13 Juli 2026

Editor in Chief,



HIDAYATUS SIBYAN, S.Kom .,
M.Kom.



Universitas Nusantara PGRI Kediri

Status Terakreditasi Baik Sekali

SK BAN-PT No.671/SK/BAN-PT/Akred/PT/VI/2021 Tanggal 21 Juli 2021

Lembaga Penelitian Dan Pengabdian Pada Masyarakat (LPPM)

Kampus I Jl. K.H. Achmad Dahlan 76 Kediri 64112 Telp.(0354)771576,771503 Kediri

Website: <https://lp2m.unpkediri.ac.id> email: lemlit@unpkediri.ac.id

BERITA ACARA UJIAN SKRIPSI

Pada hari ini Jumat tanggal 26 Juni 2026 bertempat di Ruang UJIAN M4 Universitas Nusantara PGRI Kediri, telah dilaksanakan Ujian Skripsi Mahasiswa Universitas Nusantara PGRI Kediri :



Nama : Mediana Nurlailli
NPM : 2213030130
Prodi : S1-Sistem Informasi
Semester/Tahun Akademik : Genap, 25/26

Judul : Optimasi Seleksi Fitur Menggunakan Wrapper dan Particle Swarm
Optimization pada Random Forest untuk Prediksi Risiko Diabetes Tahap Awal

Dengan Hasil : Revisi Naskah
Nilai : 88 (A)

Ketua Penguji

☑ Valid

(Dr. SUCIPTO, S.Kom, M.Kom)

Penguji 1

☑ Valid

(MUHAMMAD NAJIBULLOH MUZAKI,
S.Kom, M.Cs)

Penguji 2

☑ Valid

(ANITA SARI WARDANI, S.Kom, M.Kom)

nb:

- Input Nilai dilakukan oleh Ketua Penguji, Penguji 1, dan Penguji ke-n melalui siakad
- Nilai Final berasal dari gabungan seluruh dosen penguji
- Daftar Dosen yang telah mengisi nilai: MUHAMMAD NAJIBULLOH MUZAKI, ANITA SARI WARDANI, SUCIPTO
- Pastikan semua penguji telah melakukan input nilai



Universitas Nusantara PGRI Kediri

Status Terakreditasi Baik Sekali

SK BAN-PT No.671/SK/BAN-PT/Akred/PT/VII/2021 Tanggal 21 Juli 2021

Lembaga Penelitian Dan Pengabdian Pada Masyarakat (LPPM)

Kampus I Jl. K.H. Achmad Dahlan 76 Kediri 64112 Telp.(0354)771576,771503 Kediri

Website: <https://lp2m.unpkediri.ac.id> email: lemlit@unpkediri.ac.id

LEMBAR REVISI UJIAN SKRIPSI



Nama : Mediana Nurlailli
 NPM : 2213030130
 Prodi : S1-Sistem Informasi
 Semester/Tahun : Genap, 25/26
 Akademik
 Hari/Tanggal : Jumat / 26 Juni 2026
 Ujian
 Judul : Optimasi Seleksi Fitur Menggunakan Wrapper dan Particle Swarm
 Optimization pada Random Forest untuk Prediksi Risiko Diabetes
 Tahap Awal
 Dengan Hasil : Revisi Naskah
 Nilai : 88 (A)

Catatan Revisi:

lihat pada naskah skripsi , revisi pada naskah, Perbaikan di landasan teori, penyajian hasil penelitian dan penulisan

Ketua Penguji

(Dr. SUCIPTO, S.Kom, M.Kom)

Penguji 1

(MUHAMMAD NAJIBULLOH MUZAKI,
S.Kom, M.Cs)

Penguji 2

(ANITA SARI WARDANI, S.Kom, M.Kom)

nb:

- Tanda Tangan revisi (Ketua Penguji, Penguji 1, dan Penguji ke-n) dilakukan secara manual
- Nilai berasal dari gabungan dari seluruh dosen penguji
- Daftar Dosen yang telah mengisi nilai: MUHAMMAD NAJIBULLOH MUZAKI, ANITA SARI WARDANI, SUCIPTO

