



**PERBANDINGAN AKURASI
ALGORITMA *SUPPORT VECTOR MACHINE* (SVM) DAN
RANDOM FOREST (RF) DENGAN TEKNIK SMOTE UNTUK
PREDIKSI PENYAKIT DIABETES**

SKRIPSI

Diajukan sebagai salah satu syarat untuk mendapatkan gelar
sarjana Informatika Medis

ADELINA DWI WARDANI

2109001

**PROGRAM STUDI INFORMATIKA MEDIS PROGRAM
SARJANA UNIVERSITAS WIDYA HUSADA SEMARANG**

Agustus, 2025

PESETUJUAN SIAP UJIAN SKRIPSI

Judul : Perbandingan Akurasi Algoritma *Support Vector Machine*
(SVM) dan *Random Forest* (RF) dengan Teknik SMOTE
untuk Penyakit Diabetes

Nama mahasiswa : Adelina Dwi Wardani

NIM : 2109001

Siap dipertahankan di depan Tim Penguji

Pada 15 Agustus 2025

Menyetujui

Pembimbing Utama



Harsono S. Kom, M. Kes

NIP: 197709021997111015

Pembimbing Pendamping



Dr. Ambar Dwi Erawati, S. Si. T., M. Il. Kes

NUPTK: 7852759660231172

LEMBAR PENGESAHAN

Judul : Perbandingan Akurasi Algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dengan Teknik SMOTE untuk Penyakit Diabetes

Nama mahasiswa : Adelina Dwi Wardani

NIM : 2109001

Telah dipertahankan di depan Tim Penguji
Pada 15 Agustus 2025

Menyetujui,

Ketua Penguji	: Sigit Sugiharto, S.Kom, M.Kom	()
Anggota Penguji	: Okti Trihastuti Dyah R, S.K.M, M.K.M	()
Sekretaris Penguji	: Harsono S.Kom, M.Kes	()

Mengetahui

Rektor



(Prof. Dr. Chandrasa Soekardi, DEA)
NUPTK : 7836735636130062

Ketua Program Studi

(Sigit Sugiharto, S.Kom, M.Kom)
NUPTK : 0451756657130093

PERNYATAAN KEASLIAN PENELITIAN

Saya yang bertanda tangan dibawah ini :

Nama : Adelina Dwi Wardani

Tempat tanggal lahir : Semarang, 17 Mei 2002

NIM : 2109001

Program Studi : Informatika Medis

Dengan ini menyatakan dengan sebenar-benarnya bahwa:

1. Skripsi dengan judul “Perbandingan Akurasi Algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dengan Teknik SMOTE untuk Prediksi Penyakit Diabetes” adalah hasil karya saya, dan dalam naskah ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar Sarjana Informatika Medis di suatu Perguruan Tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain baik sebagianatau keseluruhan, kecuali secara tertulis dikutip dalam naskah ini dan diterbitkan dalam sumber kutipan atau daftar pustaka.
2. Apabila ternyata dalam naskah laporan tugas kahir studi ini dapat dibuktikan terdapat unsur- unsur plagiat, saya bersedia Skripsi ini digugurkan dan gelar akademik yang telah saya peroleh dibatalkan, serta diproses dengan ketentuan hukum yang berlaku.
3. Skripsi ini dapat dijadikan sumber Pustaka yang merupakan hak bebas *royalty* non eksklusif.

Dengan demikian pernyataan ini saya buat dengan sebenar-benarnya untuk dipergunkana sebagaimana mestinya.

Semarang, 15 Agustus 2025

Adelina Dwi Wardani

ABSTRAK

Adelina Dwi Wardani

Perbandingan Akurasi Algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) dengan Teknik SMOTE untuk Prediksi Penyakit Diabetes

11 lampiran depan + 61 halaman + 11 tabel + 19 gambar + 3 lampiran akhir

Latar Belakang: Diabetes melitus merupakan salah satu penyakit kronis yang banyak diderita masyarakat dan berpotensi menimbulkan komplikasi serius apabila tidak terdeteksi secara dini. Seiring berkembangnya teknologi, machine learning menjadi pendekatan yang efektif dalam memprediksi risiko diabetes secara cepat dan akurat.

Metode: Penelitian ini menggunakan algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) untuk memprediksi risiko diabetes berdasarkan dataset Pima Indian Diabetes. Data dibagi menggunakan metode split 70:30 dan diseimbangkan menggunakan teknik SMOTE untuk mengatasi ketidakseimbangan kelas. Evaluasi performa model dilakukan dengan metrik akurasi, *precision*, *recall*, *F1-score*, dan *ROC AUC*. Proses tuning hyperparameter dilakukan menggunakan *GridSearchCV* dengan *8-fold cross-validation*.

Hasil Penelitian: Model *Random Forest* dengan penerapan SMOTE menunjukkan performa terbaik dengan akurasi sebesar 78,35%, *precision* 84%, *recall* 72%, dan *F1-score* sebesar 70%. Hasil ini menunjukkan bahwa penggunaan teknik *balancing data* dan optimasi parameter dapat meningkatkan kinerja model secara signifikan dalam mendeteksi risiko diabetes.

Kesimpulan: Penerapan algoritma *machine learning* seperti *Random Forest* dan SVM mampu membantu prediksi risiko diabetes secara efektif. *Random Forest* dengan teknik SMOTE memberikan hasil yang lebih optimal dibandingkan metode lainnya. Penelitian ini diharapkan dapat menjadi dasar pengembangan sistem deteksi dini diabetes untuk mendukung pengambilan keputusan dalam bidang kesehatan.

Kata Kunci : Diabetes, *Random Forest*, *Support Vector Machine*, SMOTE, Prediksi, *Machine Learning*

Referensi : 21 (2017-2025)

ABSTRACT

Adelina Dwi Wardani

Comparison of the Accuracy of Support Vector Machine (SVM) and Random Forest (RF) Algorithms with the SMOTE Technique for Diabetes Prediction

11 Front Attachments + 61 Pages + 11 Tables + 19 Images + 3 Back Attachments

Background: Diabetes mellitus is a chronic disease that affect many people and has the potential to cause serious complications if not detected early. With the advancement of technology, machine learning has become an effective approach for predicting diabetes risk quickly and accurately.

Methods: This study used the Support Vector Machine (SVM) and Random Forest (RF) algorithms to predict diabetes risk based on the Pima Indian Diabetes dataset. The data was divided using a 70 : 30 split method and balanced using the SMOTE technique to address class imbalance. Model performance was evaluated using accuracy, precision, recall, F1-score, and ROC AUC metics. Hyperparameter tuning was performed using GridSearchCV with 8-fold cross-validation.

Research Result: The Random Forest model with SMOTE implementation demonstrated the best performance with an accuracy of 78.35%, a precision of 84%, a recall of 72%, and an F1-score of 70%. These result indicate that the use of data balancing techniques and parameter optimization can significantly improve model performance in detecting diabetes risk.

Conclusion: The application of machine learning algorithms such as Random Forest and SVM can help effectively predict diabetes risk. Random Forest with SMOTE techniques provided more optimal result than other methods. This research is expected to form the basis for developing an early diabetes detection system to support decision-making in the healthcare sector.

Keywords: Diabetes, Random Forest, Support Vector Machine, SMOTE, Prediction, Machine Learning.

References: 21 (2017-2025)

KATA PENGANTAR

Puji Syukur penulis panjatkan ke hadirat Tuhan Yang Maha Esa karena atas segala Rahmat, karunia, dan bimbingan-Nya sehingga penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Perbandingan Akurasi Algoritma *Support Vector Machine* (SVM) dan *Random Forest* (RF) untuk Prediksi Penyakit Diabetes”.

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar sarjana Informatika Medis. Penulis menyadari bahwa penyusunan skripsi ini tidak lepas dari dukungan berbagai pihak. Oleh karena itu, penulis menyampaikan ucapan terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Chandrasa Soekardi, DEA Rektor Universitas Widya Husada Semarang,
2. Sigit Sugiharto, S.Kom, M.Kom Ketua Prodi Program Studi Informatika Medis Program Sarjana
3. Harsono, S.Kom., M.Kes Dosen Pembimbing Utama
4. Dr. Ambar Dwi Erawati, S.Si.T., M.H.Kes Dosen Pembimbing Pendamping
5. Kedua orang tua serta keluarga yang telah mendoakan dan memberi semangat kepada penulis,
6. Bapak dan Ibu Dosen Program Studi Informatika Medis yang telah banyak memberikan ilmunya,
7. Teman-teman seperjuangan Angkatan 2021 Program Studi Informatika Medis yang sudah memberikan semangat, kebersamaan dan saling membantu selama proses penyusunan skripsi ini,
8. Penulis juga ingin menyampaikan terima kasih secara khusus kepada Kim Namjoon, Kim Seokjin, Min Yoongi, Jung Hoseok, Park Jimin, Kim Taehyung, Jeon Jungkook (BTS), atas karya, pemikiran, dan inspirasinya yang telah menjadi sumber semangat dan motivasi tersendiri bagi penulis dalam menghadapi proses penulisan skripsi ini.

Penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, kritik dan saran yang membangun sangat penulis harapkan demi perbaikan di masa mendatang. Kiranya dengan adanya skripsi ini, dapat bermanfaat bagi kita semua.

Semarang, Agustus 2025
Penyusun

(Adelina Dwi Wardani)



DAFTAR ISI

SKRIPSI.....	i
PESETUJUAN SIAP UJIAN SKRIPSI.....	ii
LEMBAR PENGESAHAN	iii
PERNYATAAN KEASLIAN PENELITIAN	iv
ABSTRAK	v
<i>ABSTRACT</i>	vi
KATA PENGANTAR.....	vii
DAFTAR ISI	ix
DAFTAR TABEL.....	xi
DAFTAR GAMBAR	xii
DAFTAR SINGKATAN	xiii
BAB I PENDAHULUAN.....	Error! Bookmark not defined.
A. Latar Belakang	Error! Bookmark not defined.
B. Perumusan Masalah	Error! Bookmark not defined.
C. Tujuan.....	Error! Bookmark not defined.
D. Manfaat	Error! Bookmark not defined.
E. Keaslian Penelitian.....	Error! Bookmark not defined.
BAB II TINJAUAN PUSTAKA.....	Error! Bookmark not defined.
A. Landasan Teori	Error! Bookmark not defined.
1. Diabetes.....	Error! Bookmark not defined.
2. <i>Data Mining</i>	Error! Bookmark not defined.
3. Algoritma <i>Support Vector Machine</i> (SVM)	Error! Bookmark not defined.
4. Algoritma <i>Random Forest</i>	Error! Bookmark not defined.
5. Evaluasi Model Prediksi	Error! Bookmark not defined.
6. <i>Pima Indian Diabetes Database</i>	Error! Bookmark not defined.
7. <i>K-Fold Cross-Validation</i>	Error! Bookmark not defined.

B. Kerangka Teori	Error! Bookmark not defined.
BAB III METODE PENELITIAN.....	Error! Bookmark not defined.
A. Kerangka Konsep	Error! Bookmark not defined.
B. Variabel Penelitian	Error! Bookmark not defined.
C. Ruang Lingkup Penelitian.....	Error! Bookmark not defined.
D. Rancangan Penelitian	Error! Bookmark not defined.
E. Instrument Penelitian	Error! Bookmark not defined.
F. Analisis Data	Error! Bookmark not defined.
G. Etika Penelitian	Error! Bookmark not defined.
BAB IV HASIL PENELITIAN DAN BAHASAN.....	Error! Bookmark not defined.
A. Gambaran Umum Tempat Penelitian	Error! Bookmark not defined.
B. Hasil Penelitian	Error! Bookmark not defined.
C. Bahasan	Error! Bookmark not defined.
D. Keterbatasan	Error! Bookmark not defined.
BAB V PENUTUP.....	Error! Bookmark not defined.
A. Kesimpulan	Error! Bookmark not defined.
B. Saran.....	Error! Bookmark not defined.
DAFTAR PUSTAKA	Error! Bookmark not defined.
LAMPIRAN.....	Error! Bookmark not defined.

DAFTAR TABEL

Tabel 1. 1 Keaslian Penelitian	Error! Bookmark not defined.
Tabel 2. 1 Rentang Nilai ROC AUC	Error! Bookmark not defined.
Tabel 3. 1 Variabel Independen	Error! Bookmark not defined.
Tabel 3. 2 Variabel Dependen	Error! Bookmark not defined.
Tabel 4. 1 Distribusi Data Sebelum dan Setelah SMOTE ...	Error! Bookmark not defined.
Tabel 4. 2 Evaluasi Model SVM Tanpa SMOTE ...	Error! Bookmark not defined.
Tabel 4. 3 Evaluasi Model SVM Dengan SMOTE	Error! Bookmark not defined.
Tabel 4. 4 Evaluasi Model RF Tanpa SMOTE.....	Error! Bookmark not defined.
Tabel 4. 5 Evaluasi Model RF Dengan SMOTE....	Error! Bookmark not defined.
Tabel 4. 6 Hasil Evaluasi Model	Error! Bookmark not defined.
Tabel 4. 7 Contoh <i>Input</i> Data	Error! Bookmark not defined.

DAFTAR GAMBAR

Gambar 2. 1 Fungsi Kernel	Error! Bookmark not defined.
Gambar 2. 2 <i>Hyperlane Super Vector</i>	Error! Bookmark not defined.
Gambar 2. 3 <i>Hyperparameter C</i>	Error! Bookmark not defined.
Gambar 2. 4 <i>Random Forest</i>	Error! Bookmark not defined.
Gambar 2. 5 Kerangka Teori	Error! Bookmark not defined.
Gambar 3. 1 Kerangka konsep	Error! Bookmark not defined.
Gambar 3. 2 Jadwal Penelitian	Error! Bookmark not defined.
Gambar 4. 1 Distribusi Data Kelas Diabetes vs Non Diabetes ..	Error! Bookmark not defined.
Gambar 4. 2 Visualisasi Kelas Data pada SMOTE	Error! Bookmark not defined.
Gambar 4. 3 <i>Confusion Matrix</i> SVM Tanpa SMOTE.....	Error! Bookmark not defined.
Gambar 4. 4 ROC AUC SVM Tanpa SMOTE.....	Error! Bookmark not defined.
Gambar 4. 5 <i>Confusion Matrix</i> SVM Dengan SMOTE	Error! Bookmark not defined.
Gambar 4. 6 ROC AUC SVM Dengan SMOTE....	Error! Bookmark not defined.
Gambar 4. 7 <i>Confusion Matrix</i> RF Tanpa SMOTE	Error! Bookmark not defined.
Gambar 4. 8 ROC AUC RF Tanpa SMOTE	Error! Bookmark not defined.
Gambar 4. 9 <i>Confusion Matrix</i> RF Dengan SMOTE.....	Error! Bookmark not defined.
Gambar 4. 10 ROC AUC RF Dengan SMOTE.....	Error! Bookmark not defined.
Gambar 4. 11 Perbandingan Evaluasi Model	Error! Bookmark not defined.
Gambar 4. 12 Visualisasi Tampilan Implementasi Sistem Prediksi.....	Error! Bookmark not defined.

DAFTAR SINGKATAN

AUC	: <i>Area Under the Curve</i>
BMI	: <i>Body Mass Index</i>
CV	: <i>Cross Validation</i>
DHS	: <i>Daftar Hasil Studi</i>
FN	: <i>False Negatif</i>
FP	: <i>False Positif</i>
FPR	: <i>False Positive Rate</i>
IDF	: <i>Internasional Diabetes Federation</i>
kg	: <i>Kilogram</i>
KNN	: <i>K-Nearest Neighbors</i>
mm	: <i>Milimeter</i>
Mmg/dL	: <i>Miligram per desiliter</i>
mmHg	: <i>Millimetrum Hydrargyr</i>
mu U/ml	: <i>Mikro Unit per Mililiter</i>
NIDDK	: <i>National Institute of Diabetes and Digestive and Kidney Diseases</i>
NIK	: <i>Nomor Induk Kependudukan</i>
OA	: <i>Overall Accuracy</i>
PIDD	: <i>Dataset Pima Indians Diabetes Database</i>
RBF	: <i>Radial Basis Function</i>
RF	: <i>Random Forest</i>
ROC	: <i>Reciver Operting Characteristic</i>
SMOTE	: <i>Synthetic Minority Oversampling Technique</i>
SVM	: <i>Support Vector Machine</i>
TN	: <i>True Negatif</i>
TP	: <i>True Positif</i>
TPR	: <i>True Positive Rate</i>
WHO	: <i>World Health Organization</i>

