



SKRIPSI

PREDIKSI RISIKO DIABETES MENGGUNAKAN ENSEMBLE WEIGHTED VOTING DENGAN ALGORITMA LOGISTIC REGRESSION, SVM, DAN RANDOM FOREST

ACH.DIKI PRASETYO
NPM 21081010055

DOSEN PEMBIMBING

Fetty Tri Anggraeny, S.Kom., M.Kom
Retno Mumpuni, S.Kom., M.Sc

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**



SKRIPSI

PREDIKSI RISIKO DIABETES MENGGUNAKAN ENSEMBLE WEIGHTED VOTING DENGAN ALGORITMA LOGISTIC REGRESSION, SVM, DAN RANDOM FOREST

ACH.DIKI PRSETYO
NPM 21081010055

DOSEN PEMBIMBING

Fetty Tri Anggraeny, S.Kom, M.Kom
Retno Mumpuni, S.Kom., M.Sc

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER
PROGRAM STUDI INFORMATIKA
SURABAYA
2025**

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

PREDIKSI RISIKO DIABETES MENGGUNAKAN ENSEMBLE WEIGHTED VOTING DENGAN ALGORITMA LOGISTIC REGRESSION, SVM DAN RANDOM FOREST

Oleh :

ACH.DIKI PRASETYO

NPM 21081010055

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada tanggal 05 Juni 2025

Menyetujui

Fetty Tri Anggraeny, S.Kom. M.Kom

(Pembimbing I)

NIP. 19820211 202121 2 005

Retno Mumpuni, S.Kom., M.Sc

(Pembimbing II)

NPT. 172198 70 716054

Dr. Ir. Kartini, S.Kom. MT.

(Ketua Penguji)

NIP. 19611110 199103 2 001

Eka Prakarsa Mandyartha, S.T., M.Kom

(Anggota Penguji)

NIP. 19880525 201803 1 001

Mengetahui

Dekan Fakultas Ilmu Komputer

Prof. Dr. Ir. Novirina Hendrasarie, MT

NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

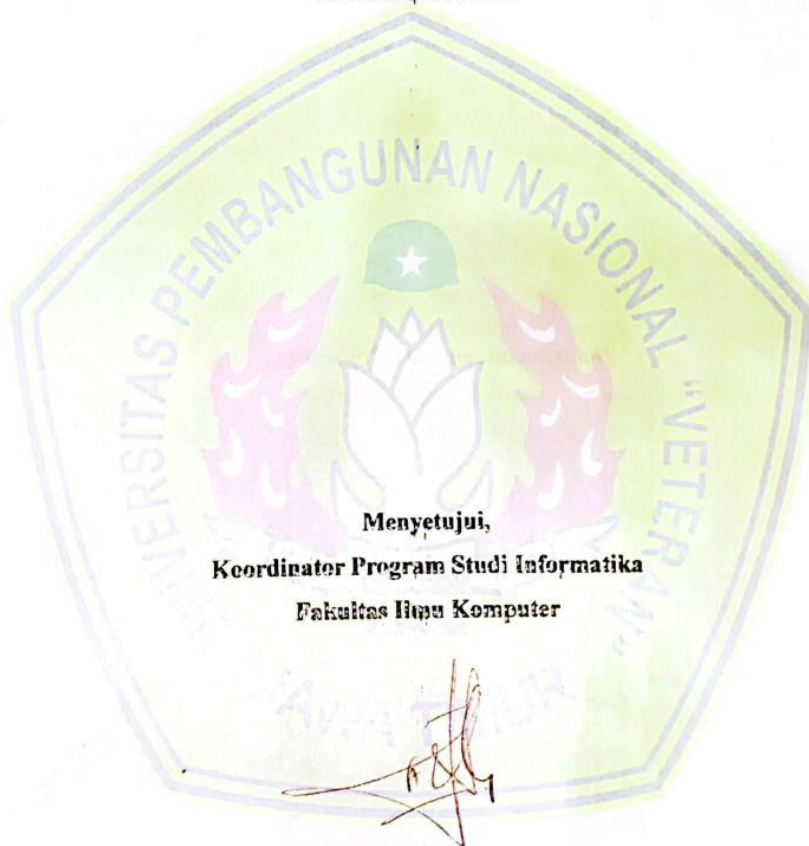
LEMBAR PERSETUJUAN

PREDIKSI RISIKO DIABETES MENGGUNAKAN ENSEMBLE WEIGHTED VOTING DENGAN ALGORITMA LOGISTIC REGRESSION, SVM DAN RANDOM FOREST

Oleh :

ACH.DIKI PRASETYO

NPM. 21081010055



**Menyetujui,
Koordinator Program Studi Informatika
Fakultas Ilmu Komputer**

Fetty Tri Anggraeny, S.Kom., M.Kom.

NIP. 19820211 202121 2 005

Halaman ini sengaja dikosongkan

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan di bawah ini:

Nama Mahasiswa : ACH.DIKI PRASETYO
NPM : 21081010055
Program Studi : Sarjana (S1)
Program Studi : Informatika
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Skripsi ini tidak terdapat bagian dari karya ilmiah lain yang telah diajukan untuk memperoleh gelar akademik di suatu lembaga Pendidikan Tinggi, dan juga tidak terdapat karya atau pendapat yang pernah ditulis atau disebutkan secara lengkap dalam daftar pustaka.

Dan saya menyatakan bahwa dokumen ilmiah ini bebas dari unsur-unsur plagiasi. Apabila dikemudian hari ditemukan indikasi plagiat pada Skripsi ini, saya bersedia menerima sanksi sesuai dengan peraturan perundang-undangan yang berlaku.

Demikian surat pernyataan ini saya buat dengan sesungguhnya tanpa ada paksaan dari siapapun juga dan untuk dipergunakan sebagaimana mestinya.



Surabaya, 12 Juni 2025

Yang Membuat Pernyataan,



ACH.DIKI PRASETYO
NPM. 21081010055

Halaman ini sengaja dikosongkan

ABSTRAK

Nama Mahasiswa / NPM	Ach.Diki Prasetyo / 21081010055
Judul Skripsi	PREDIKSI RISIKO DIABETES MENGUNAKAN ENSEMBLE WEIGHTED VOTING DENGAN ALGORITMA LOGISTIC REGRESSION, SVM, DAN RANDOM FOREST
Dosen Pembimbing	1. Fetty Tri Anggraeny, S.Kom., M.Kom. 2. Retno Mumpuni, S.Kom, M.Sc

Diabetes merupakan penyakit kronis yang terus meningkat jumlah penderitanya secara global, sehingga dibutuhkan sebuah solusi pencegahan seperti penggunaan sistem prediksi dini risiko diabetes. Penelitian ini bertujuan untuk membuat dan menguji model prediksi risiko diabetes menggunakan metode *Ensemble Weighted Voting* yang menggabungkan tiga algoritma *machine learning*, yaitu *Logistic Rregression*, *Support Vector Machine*, dan *Random Forest*. Data yang digunakan berasal dari survei "*Behavioral Risk Factor Surveillance System*" (BRFSS) tahun 2021 yang diambil dari *Kaggle*, serta data primer yang diambil di Indonesia menggunakan *google form*. Pengujian model dilakukan menggunakan *confusion matriks* dengan metrik evaluasi *accuracy*, *precision*, *recall*, dan *F1-score*. Dari hasil pengujian, model *Ensemble* memberikan hasil akurasi yang lebih tinggi dibandingkan dengan model tunggal. Pada skenario 1 dan 2, model *ensemble* mendapatkan akurasi tertinggi 74,89% dan 74,57%. Pada skenario 4, yang dilatih menggunakan 5% data *kaggle* dan di uji menggunakan data lokal, model mendapatkan akurasi tertinggi yaitu 90,00%. Meskipun pada skenario 3 model *random forest* mendapatkan akurasi yang lebih tinggi. Secara keseluruhan model yang menggunakan *ensemble* memberikan hasil yang lebih baik, dengan akurasi yang lebih tinggi dibandingkan model tunggal pada berbagai skenario. Sistem prediksi risiko diabetes juga telah berhasil diimplementasikan dalam bentuk website dengan menggunakan *Flask* sebagai *backend* dan *Laravel* sebagai *frontend*.

Kata kunci : *Logistic Regression*, *Random forest*, *SVM*, *Ensemble*, Prediksi Diabetes

Halaman ini sengaja dikosongkan

ABSTRACT

Student Name / NPM
Thesis Title

Ach.Diki Prasetyo / 21081010055
DIABETES RISK PREDICTION USING
ENSEMBLE WEIGHTED VOTING WITH
LOGISTIC REGRESSION, SVM, AND
RANDOM FOREST ALGORITHMS

Advisors

1. Fetty Tri Anggraeny, S.Kom., M.Kom.
2. Retno Mumpuni, S.Kom, M.Sc

Diabetes mellitus is a chronic disease that continues to increase in number of sufferers globally, so a preventive solution is needed such as the use of an early prediction system for diabetes risk. This study aims to develop and test a diabetes risk prediction model using the Ensemble Weighted Voting method that combines three machine learning algorithms, namely Logistic Regression, Support Vector Machine, and Random Forest. The data used comes from the 2021 "Behavioral Risk Factor Surveillance System" (BRFSS) survey taken from Kaggle, as well as primary data taken in Indonesia using Google Form. Model testing was carried out using a confusion matrix with evaluation metrics of accuracy, precision, recall, and F1-score. The test results show that the Ensemble model is able to provide better accuracy results compared to a single model. In scenarios 1 and 2, the ensemble model gets the highest accuracy with values of 74,89% and 74,57%. Then, in scenario 4, which is only trained using 5% of Kaggle data and tested using local data, the model gets the highest accuracy of 90,00%. Although in scenario 3 the random forest model obtained higher accuracy, overall the model using the ensemble showed better performance, with higher accuracy values than the single model in most scenarios. The diabetes prediction risk system has also been successfully implemented in the form of a website using Flask as the backend and Laravel as the frontend.

Keywords : Logistic Regression, Random forest, SVM, Ensemble, Diabetes Prediction

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Dengan segenap kerendahan hati, penulis memanjatkan puji syukur kehadiran Allah SWT atas limpahan rahmat, hidayah, dan karunia-Nya, sehingga penulis berhasil menyelesaikan skripsi yang berjudul "Prediksi Risiko Diabetes Menggunakan Ensemble Weighted Voting Dengan Algoritma Logistic Regression, SVM, dan Random Forest".

Skripsi ini disusun sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata Satu (S1) pada Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur. Dalam prosesnya, penulis menyadari sepenuhnya bahwa kelancaran penulisan skripsi ini tidak lepas dari dukungan, bantuan, dan bimbingan dari berbagai pihak, baik secara moril, spiritual, maupun materiil. Oleh karena itu, pada kesempatan ini, penulis ingin menyampaikan ucapan terima kasih yang tulus kepada:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT., selaku Rektor Universitas Pembangunan Nasional “Veteran” Jawa Timur, atas kesempatan dan fasilitas yang telah diberikan.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, M.T., selaku Dekan Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur, atas dukungan dan arahan yang diberikan.
3. Ibu Fetty Tri Anggraeny, S.Kom., M.Kom. selaku Koordinator Program Studi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur, dosen pembimbing pertama dan dosen wali, yang dengan sabar memberikan bimbingan, arahan, dan dukungan penuh selama proses pengerjaan laporan skripsi.
4. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom., selaku koordinator skripsi yang telah banyak membantu dalam urusan administrasi.
5. Ibu Retno Mumpuni, S.Kom., M.ScEva Yulia Puspaningrum, S.Kom., M.Kom., selaku dosen pembimbing kedua, yang juga telah memberikan dukungan dan masukan berharga selama penyusunan laporan skripsi ini.

6. Ibu Dr. Ir. Kartini, S.Kom. MT., selaku Dosen Penguji Pertama, yang telah memberikan masukan dan saran-saran yang sangat berharga dalam proses ujian dan penyempurnaan laporan skripsi.
7. Bapak Eka Prakarsa Mandyartha, S.T., M.Kom, selaku Dosen Penguji Kedua, yang dengan penuh perhatian dalam memberikan evaluasi, pertanyaan dan saran yang sangat membantu dalam pengembangan penelitian ini.
8. Seluruh staf dan karyawan Tata Usaha Fakultas Ilmu Komputer yang telah sigap membantu dalam proses administrasi dan pengurusan surat-menyurat yang diperlukan selama pelaksanaan skripsi.
9. Ayah, Ibu, dan Adik tercinta, sebagai keluarga yang senantiasa memberikan dukungan tak ternilai, baik berupa materi, semangat, maupun doa, selama penulis menjalani penelitian dan penyusunan skripsi ini.
10. Semua pihak yang tidak dapat penulis sebutkan satu per satu, atas segala bentuk dukungan dan kontribusi yang telah diberikan demi kelancaran penelitian ini.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna dan memiliki berbagai kekurangan. Oleh karena itu, saran dan kritik yang membangun sangat penulis harapkan demi perbaikan ke depannya. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi penulis pribadi dan para pembaca secara umum.

Surabaya, 30 Mei 2025

Penulis

(Ach. Diki Prasetyo)

DAFTAR ISI

LEMBAR PENGESAHAN	iii
LEMBAR PERSETUJUAN.....	v
SURAT PERNYATAAN BEBAS PLAGIASI.....	vii
ABSTRAK.....	ix
ABSTRACT.....	xi
KATA PENGANTAR.....	xiii
DAFTAR ISI.....	xv
DAFTAR GAMBAR	xix
DAFTAR TABEL.....	xxi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penelitian	3
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	4
BAB II TINJAUAN PUSTAKA.....	5
2.1 Penelitian Terdahulu	5
2.2 Diabetes.....	7
2.3 <i>Data Mining</i>	8
2.3.1 Proses Data Mining.....	9
2.3.2 Teknik <i>Data Mining</i>	11
2.4 <i>Logistic Regression</i>	12
2.5 <i>Support Vector Machine</i>	14
2.6 <i>Random forest</i>	17

2.7	<i>Ensemble Weighted Voting</i>	19
2.8	Confusion matriks	20
2.8.1	<i>Accuracy</i>	21
2.8.2	<i>Precision</i>	21
2.8.3	<i>Recall</i>	22
2.8.4	<i>F1-Score</i>	22
2.9	<i>Flask</i>	23
2.10	<i>Laravel</i>	23
BAB III DESAIN DAN IMPLEMENTASI SISTEM.....		25
3.1	Tahapan Penelitian	25
3.2	Pengumpulan Data	25
3.3	<i>Preprocessing</i>	30
3.3.1	Pembersihan Data	30
3.3.2	<i>Encoding Data</i>	31
3.3.3	Pembagian Data	32
3.3.4	Normalisasi	33
3.4	Implementasi Algoritma	34
3.4.1	<i>Logistic Regression</i>	35
3.4.2	<i>Random forest</i>	37
3.4.3	<i>SVM</i>	41
3.4.4	<i>Ensemble Weighted Voting</i>	45
3.5	Evaluasi	46
3.6	Skenario Uji Coba	47
3.7	Arsitektur Sistem	50
3.7.1	Backend	51

3.7.2	Frontend	52
3.7.3	Alur Pengguna.....	53
3.8	Rancangan dan Implementasi Desai	53
3.8.1	Rancangan Desain.....	54
3.8.2	Implementasi Desain	56
3.9	Implementasi Sistem	60
3.9.1	Unduh Model <i>Ensemble</i>	60
3.9.2	Implementasi Backend.....	61
3.9.3	Integrasi API dengan Frontend	62
BAB IV	PENGUJIAN DAN ANALISA.....	63
4.1	Pengumpulan Dataset	63
4.2	<i>Preprocessing</i>	64
4.3	Model <i>Logistic Regression</i>	69
4.4	Model <i>Random forest</i>	71
4.5	Model SVM.....	72
4.6	Model <i>Ensemble Weighted Voting</i>	73
4.7	Skenario Penelitian	76
4.7.1	Pengujian Skenario Penelitian Pertama	76
4.7.2	Pengujian Skenario Peneltian Kedua	79
4.7.3	Pengujian Skenario Penelitian Ketiga	81
4.7.4	Pengujian Skenario Keempat	84
4.8	Evaluasi Akhir.....	86
4.9	Pengujian Sistem	88
BAB V	KESIMPULAN.....	93
5.1	Kesimpulan.....	93

5.2	Saran	94
DAFTAR PUSTAKA.....		95

DAFTAR GAMBAR

Gambar 2. 1 Alur <i>Data Mining</i>	9
Gambar 2. 2 Konsep <i>Logistic Regression</i> [18].	12
Gambar 2. 3 Konsep Algoritma SVM [20].	14
Gambar 2. 4 Konsep Algoritma Random forest [24].	18
Gambar 3. 1 Alur penelitian	25
Gambar 3. 2 Alur <i>Preprocessing</i>	30
Gambar 3. 3 Alur perhitungan <i>Logistic Regression</i>	35
Gambar 3. 4 Alur perhitungan <i>Random forest</i>	37
Gambar 3. 5 Alur perhitungan SVM	41
Gambar 3. 6 Alur perhitungan <i>Weighted Voting</i>	45
Gambar 3. 7 Arsitektur sistem prediksi risiko diabetes	50
Gambar 3. 8 Flowchart Alur Proses Backend	51
Gambar 3. 9 Flowchart Alur Proses Frontend	52
Gambar 3. 10 Use case penggunaan sistem	53
Gambar 3. 11 Rancangan Halaman Form Prediksi	54
Gambar 3. 12 Rancangan pop-up setelah melakukan prediksi	55
Gambar 3. 13 Rancangan Halaman Dashboard	55
Gambar 3. 14 Rancangan Semua Hasil Prediski	56
Gambar 3. 15 Desain form input data bagian 1	57
Gambar 3. 16 Desain form input data bagian 2	57
Gambar 3. 17 Pop-up risiko rendah diabetes	58
Gambar 3. 18 Pop-up berisiko tinggi diabetes	59
Gambar 3. 19 Dashboard Admin	59
Gambar 3. 20 Implementasi Desain List data hasil prediksi pengguna	60
Gambar 4. 1 Data <i>Kaggle</i> sebelum <i>preprocessing</i>	63
Gambar 4. 2 Data lokal (primer)	64
Gambar 4. 3 Data <i>Kaggle</i> setelah <i>dipreprocessing</i>	65
Gambar 4. 4 Dataset sebelum di <i>encoding</i>	68
Gambar 4. 5 Dataset setelah di <i>encoding</i>	69
Gambar 4. 6 <i>Confusion matiks</i> model <i>ensemble</i> skenario 1	77

Gambar 4. 7 <i>Confusion matriks</i> model <i>ensemble</i> skenario 2.....	80
Gambar 4. 8 Confusion matiks model ensembe skenario 3.....	83
Gambar 4. 9 <i>Confusion matiks</i> model <i>ensemble</i> skenario 4.....	85
Gambar 4. 10 Hasil prediksi tidak berisiko diabetes.....	89
Gambar 4. 11 Hasil prediksi berisiko diabetes.....	89
Gambar 4. 12 Dashboard Sebelum penambahan data	90
Gambar 4. 13 Dasboard setelah penambahan data.....	91
Gambar 4. 14 Hasil Export data prediksi ke <i>excel</i>	91

DAFTAR TABEL

Tabel 3. 1 Dataset BRFSS 2021	26
Tabel 3. 2 Hasil pembersihan data.....	30
Tabel 3. 3 Contoh data duplikat yang akan dihapus	31
Tabel 3. 4 Contoh data yang didapatkan dari survei.....	32
Tabel 3. 5 Hasil setelah di <i>encoding</i>	32
Tabel 3. 6 Pembagian Dataset.....	32
Tabel 3. 7 Variabel yang perlu di normalisasi	33
Tabel 3. 8 Data sebelum dinormalisasi.....	33
Tabel 3. 9 Data setelah dinormalisasi.....	34
Tabel 3. 10 Sample data untuk perhitungan <i>Logistic Regression</i>	35
Tabel 3. 11 Nilai koefisien pada fitur.....	36
Tabel 3. 12 Hasil prediksi <i>Logistic Regression</i>	37
Tabel 3. 13 Data sample yang didapatkan dari bootstrap sampling	38
Tabel 3. 14 Nilai gini pada masing-masing fitur.....	40
Tabel 3. 15 Data sample untuk perhitungan setelah normalisasi	42
Tabel 3. 16 Hasil prediksi SVM.....	44
Tabel 3. 17 Sample hasil probabilitas masing-masing model	45
Tabel 3. 18 Sample hasil confusion matriks.....	46
Tabel 3. 19 Hasil Confusion matriks.....	47
Tabel 3. 20 Skenario Penelitian.....	48
Tabel 4. 1 <i>Encoding</i> variabel sesuai dengan dataset.....	67
Tabel 4. 2 Evaluasi Model <i>Logistic Regression</i>	76
Tabel 4. 3 Evaluasi <i>Random forest</i>	77
Tabel 4. 4 Evaluasi SVM	77
Tabel 4. 5 Classification report <i>ensemble</i> skenario 1.....	78
Tabel 4. 6 Evaluasi model <i>Logistic Regression</i>	79
Tabel 4. 7 Evaluasi Model <i>Random forest</i>	79
Tabel 4. 8 Evaluasi SVM	80
Tabel 4. 9 Classification Report <i>Ensemble</i> skenario 2.....	81
Tabel 4. 10 Evaluasi model <i>Logistic Regression</i>	81

Tabel 4. 11 Evaluasi model <i>Random forest</i>	82
Tabel 4. 12 Evaluasi model SVM	82
Tabel 4. 13 Classification report <i>ensemble</i> pada skenario 3.....	84
Tabel 4. 14 Evaluasi model <i>Logistic Regression</i>	84
Tabel 4. 15 Evaluasi model <i>Random forest</i>	85
Tabel 4. 16 Evaluasi model SVM	85
Tabel 4. 17 Classification report skenario 4	86
Tabel 4. 18 Tabel evaluasi semua model diberbagai skenario.....	87