



SKRIPSI

OPTIMASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN ALGORITMA GENETIKA PADA DETEKSI PENYAKIT DIABETES MELLITUS

MARCELLINUS ADITYA VITRO DARMAWAN
NPM 21081010107

DOSEN PEMBIMBING

Muhammad Muharrom Al Haromainy, S.Kom, M.Kom.
Achmad Junaidi, S.Kom., M.Kom.

**KEMENTERIAN PENDIDIKAN TINGGI, SAINS, DAN TEKNOLOGI
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FAKULTAS ILMU KOMPUTER**



SKRIPSI

OPTIMASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN ALGORITMA GENETIKA PADA DETEKSI PENYAKIT DIABETES MELLITUS

MARCELLINUS ADITYA VITRO DARMAWAN
NPM 21081010107

DOSEN PEMBIMBING

Muhammad Muharrom Al Haromainy, S.Kom, M.Kom.
Achmad Junaidi, S.Kom., M.Kom.

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

Halaman ini sengaja dikosongkan

LEMBAR PENGESAHAN

OPTIMASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN ALGORITMA GENETIKA PADA DETEKSI PENYAKIT DIABETES MELLITUS

Oleh :

MARCELLINUS ADITYA VITRO DARMAWAN
NPM. 21081010107

Telah dipertahankan dihadapan dan diterima oleh Tim Penguji Skripsi Prodi Informatika
Fakultas Ilmu Komputer Universitas Pembangunan Nasional Veteran Jawa Timur Pada
tanggal 10 Maret 2025

Muhammad Muharrom Al Haromayn, S.Kom,
M.Kom.
NIP. 19950601 202203 1 006



(Pembimbing I)

Achmad Junaldi, S.Kom., M.Kom.
NPT. 3 7811 04 0199 1



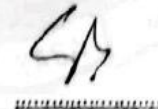
(Pembimbing II)

Budi Nugroho, S.Kom. M.Kom.
NIP. 19800907 202121 1 005



(Penguji I)

Eva Yulia Puspaningrum, S.Kom., M.Kom
NIP. 19890705 202121 2 002



(Penguji II)

Mengetahui,
Dekan Fakultas Ilmu Komputer



Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

Halaman ini sengaja dikosongkan

LEMBAR PERSETUJUAN

**OPTIMASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN
ALGORITMA GENETIKA PADA DETEKSI PENYAKIT DIABETES
MELLITUS**

Oleh :

MARCELLINUS ADITYA VITRO DARMAWAN

NPM. 21081010107

Telah disetujui untuk mengikuti Ujian Skripsi

Menyetujui,

Koordinator Program Studi Informatika

Fakultas Ilmu Komputer



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

NIP. 198202112021212005

SURAT PERNYATAAN BEBAS PLAGIASI

Saya yang bertanda tangan dibawah ini :

Nama : Marcellinus Aditya Vitro Darmawan
NPM : 21081010107
Program : Sarjana(S1))
Program Studi : Informatika
Fakultas : Ilmu Komputer

Menyatakan bahwa dalam dokumen ilmiah Tugas Akhir/Skripsi/Tesis/Disertasi* 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 diterbitkan oleh orang/lembaga lain, kecuali yang secara tertulis disitasi dalam dokumen ini dan 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/Tesis/Desertasi 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, 15 Mei 2025.....

Yang Membuat pernyataan



MARCELLINUS ADITYA
NPM 21081010107

SURAT PERNYATAAN ORISINALITAS

Yang bertandatangan di bawah ini:

Nama Mahasiswa : MARCELLINUS ADITYA VITRO DARMAWAN

Program Studi : Informatika

Dosen Pembimbing : 1. Muhammad Muharrom Al Haromainy, S.Kom,
M.Kom
2. Achmad Junaidi, S.Kom., M.Kom.

dengan ini menyatakan bahwa isi sebagian maupun keseluruhan disertasi dengan judul:

OPTIMASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN ALGORITMA GENETIKA PADA DETEKSI PENYAKIT DIABETES MELLITUS

adalah benar-benar hasil karya intelektual mandiri, diselesaikan tanpa menggunakan bahan-bahan yang tidak diizinkan dan bukan merupakan karya pihak lain yang saya akui sebagai karya sendiri. Semua referensi yang dikutip maupun dirujuk telah ditulis secara lengkap pada daftar pustaka. Apabila ternyata pernyataan ini tidak benar, saya bersedia menerima sanksi sesuai peraturan yang berlaku.



Surabaya, Mei 2025
Yang Membuat Pernyataan,



MARCELLINUS ADITYA VITRO DARMAWAN
NPM. 21081010107

ABSTRAK

Nama Mahasiswa / NPM : Marcellinus Aditya Vitro Darmawan/ 21081010107
Judul Skripsi : OPTIMASI ALGORITMA K-NEAREST NEIGHBOR
(KNN) DENGAN ALGORITMA GENETIKA PADA
DETEKSI PENYAKIT DIABETES MELLITUS
Dosen Pembimbing : 1. Muhammad Muharrom Al Haromainy, S.Kom,
M.Kom
2. Achmad Junaidi, S.Kom., M.Kom.

Penelitian ini membahas optimasi Algoritma K-Nearest Neighbor (KNN) menggunakan Algoritma Genetika (GA) dalam mendeteksi penyakit diabetes mellitus. Penelitian mencakup tahapan pengumpulan dataset gejala penyakit diabetes mellitus, preprocessing data melalui normalisasi dan penyesuaian dataset, implementasi model, serta pengujian dengan berbagai skenario untuk mendapatkan akurasi tertinggi. Data yang digunakan adalah data *Pima Indians Diabetes Database* sebagai dataset 1 dan data *Early Stage Diabetes Risk Prediction Dataset* sebagai dataset 2. Evaluasi dilakukan dengan membandingkan hasil akurasi antara KNN tanpa optimasi dan KNN yang telah dioptimasi menggunakan Algoritma Genetika. Hasil penelitian menunjukkan bahwa optimasi dilakukan dengan mencari kombinasi optimal dari nilai k dan fitur-fitur yang digunakan dalam klasifikasi. Algoritma Genetika menghasilkan individu dengan fitness terbaik berdasarkan kombinasi nilai k dan fitur yang mampu memberikan akurasi tertinggi. Pengujian dilakukan pada dua dataset dengan dua nilai fold yang berbeda. Hasil akurasi terbaik didapat pada pengujian fold 10, dengan akurasi pada dataset 1 dari 74.2% menjadi 79.1% setelah dioptimasi. Sedangkan pada dataset 2, akurasi yang didapat adalah 97.5% menjadi 98.2% setelah dioptimasi. Terdapat peningkatan akurasi pada dataset 1, sedangkan pada dataset 2, tidak terjadi peningkatan akurasi yang signifikan. Kesimpulan dari penelitian ini adalah bahwa optimasi Algoritma KNN menggunakan Algoritma Genetika terbukti meningkatkan akurasi deteksi diabetes mellitus, terutama pada dataset numerik dengan fitur yang lebih kompleks.

Kata kunci : *Diabetes Mellitus, K-Nearest Neighbor (KNN), Algoritma Genetika, Optimasi parameter, Deteksi Penyakit*

Halaman ini sengaja dikosongkan

ABSTRACT

Name / NPM : Marcellinus Aditya Vitro Darmawan/ 21081010107
Thesis : *OPTIMIZATION OF THE K-NEAREST NEIGHBOR (KNN) ALGORITHM WITH GENETIC ALGORITHM IN THE DETECTION OF DIABETES MELLITUS*
Advisor : 1. Muhammad Muharrom Al Haromainy, S.Kom, M.Kom
2. Achmad Junaidi, S.Kom., M.Kom.

This study discusses the optimization of the K-Nearest Neighbor (KNN) algorithm using Genetic Algorithm (GA) in detecting diabetes mellitus. The research includes stages of collecting datasets on diabetes mellitus symptoms, data preprocessing through normalization and dataset alignment, model implementation, and testing with various scenarios to achieve the highest accuracy. The data used consists of the Pima Indians Diabetes Database as dataset 1 and the Early Stage Diabetes Risk Prediction Dataset as dataset 2. The evaluation is conducted by comparing the accuracy results between KNN without optimization and KNN optimized using Genetic Algorithm. The study's results indicate that optimization is performed by finding the optimal combination of the k-value and the features used in classification. The Genetic Algorithm produces individuals with the best fitness based on the combination of k-values and features that yield the highest accuracy. Testing was conducted on two datasets with two different fold values. The best accuracy was obtained in the 10-fold test, where the accuracy for dataset 1 increased from 74.2% to 79.1% after optimization. Meanwhile, for dataset 2, the accuracy improved from 97.5% to 98.2% after optimization. There was an increase in accuracy for dataset 1, whereas for dataset 2, the improvement was not significant. The conclusion of this study is that optimizing the KNN algorithm using Genetic Algorithm has proven to enhance the accuracy of diabetes mellitus detection, especially in numerical datasets with more complex features.

Keywords : *Diabetes Mellitus, K-Nearest Neighbor (KNN), Genetic Algorithm, Parameter optimization, Disease detection*

Halaman ini sengaja dikosongkan

KATA PENGANTAR

Puji syukur saya panjatkan kepada Tuhan YME atas segala rahmat dan karunia-Nya kepada penulis sehingga skripsi dengan judul **“OPTIMASI ALGORITMA K-NEAREST NEIGHBOR (KNN) DENGAN ALGORITMA GENETIKA PADA DETEKSI PENYAKIT DIABETES MELLITUS”** mampu untuk diselesaikan dengan baik.

Penyusunan skripsi telah melalui banyak sekali proses dan waktu untuk diselesaikan. Kerja keras selama proses penyusunan skripsi bertujuan untuk memenuhi salah satu syarat kelulusan prodi Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa Timur. Adapula pihak-pihak telah berjasa dalam pembuatan skripsi ini:

1. Bapak Prof. Dr. Ir. Akhmad Fauzi, M.MT selaku Rektor Universitas.
2. Ibu Prof. Dr. Ir. Novirina Hendrasarie, MT. selaku Dekan Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
3. Ibu Fetty Anggraeny, S.Kom, M.Kom. selaku Koordinator Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
4. Bapak Andreas Nugroho Sihananto, S.Kom., M.Kom selaku koordinator skripsi Program Studi Informatika Fakultas Ilmu Komputer Universitas Pembangunan Nasional “Veteran” Jawa Timur.
5. Bapak Christia Aji Putra S.Kom M.T. selaku dosen wali saya
6. Bapak Muhammad Muharrom Al Haromainy, S.Kom, M.Kom. selaku dosen pembimbing pertama yang dengan sabar membimbing saya dan selalu arahan dengan baik dalam pembuatan skripsi.
7. Bapak Achmad Junaidi, S.Kom., M.Kom. selaku dosen pembimbing pertama yang dengan sabar membimbing saya dan selalu arahan dengan baik dalam pembuatan skripsi.
8. Ibu Yisti Vita Via, S.ST. M.Kom.. selaku dosen penguji pertama sidang proposal yang telah meluangkan waktu untuk menguji proposal skripsi saya.
9. Ibu Afina Lina Nurlaili, S.Kom., M.Kom. selaku dosen penguji kedua sidang proposal yang telah meluangkan waktu untuk menguji proposal skripsi saya.

10. Bapak Budi Nugroho, S.Kom. M.Kom. selaku dosen penguji pertama sidang skripsi yang telah meluangkan waktu untuk menguji dan memberikan revisi pada skripsi saya.
11. Ibu Eva Yulia Puspaningrum, S.Kom., M.Kom. selaku dosen penguji kedua sidang skripsi yang telah meluangkan waktu untuk menguji dan memberikan revisi pada skripsi saya.
12. Bapak/Ibu Karyawan yang berkerja pada bagian tata usaha yang selalu melayani untuk keperluan-keperluan dokumen persyaratan ujian skripsi.
13. Orang tua dan keluarga besar saya yang selalu mendukung, memberi arahan, dan membuat kondisi yang nyaman di rumah selama pembuatan skripsi
14. Teman-teman dekat saya turut mendukung saya dalam pengerjaan skripsi ini hingga selesai

Penulis menyadari bahwa skripsi ini masih memiliki berbagai kekurangan. Oleh karena itu, kritik dan saran yang konstruktif dari berbagai pihak sangat diharapkan guna menyempurnakan penulisan ini. Dengan segala keterbatasan yang ada, penulis berharap laporan ini dapat memberikan manfaat bagi semua pihak secara umum, serta bagi penulis secara khusus.

Surabaya, 8 Mei 2025

Penulis

Halaman ini sengaja dikosongkan

DAFTAR ISI

LEMBAR PENGESAHAN	
LEMBAR PERSETUJUAN	
SURAT PERNYATAAN ORISINALITAS	
ABSTRAK	
KATA PENGANTAR.....	i
DAFTAR ISI.....	iv
DAFTAR GAMBAR.....	vii
DAFTAR TABEL	viii
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	3
1.3. Tujuan Penelitian	4
1.4. Manfaat Penelitian	4
1.5. Batasan Masalah.....	4
BAB II TINJAUAN PUSTAKA.....	5
2.1. Penelitian Terdahulu	5
2.2. Landasan Teori.....	9
2.2.1. Diabetes Mellitus.....	9
2.2.2. K-Nearest Neighbor	11
2.2.3. Algoritma Genetika	13
2.2.4. Optimasi Algoritma KNN dengan Algoritma Genetika.....	17
2.2.5. Dataset	18
2.2.6. Preprocessing Data	19
2.2.7. Evaluasi Model.....	21
2.2.8. Cross Validation	23
2.2.9. Python.....	25
BAB III DESAIN DAN IMPLEMENTASI SISTEM	27
3.1. Analisis Kebutuhan	27
3.1.1. Lingkungan Penelitian.....	27
3.2. Tahapan Peneltian	27

3.2.1. Identifikasi Permasalahan.....	28
3.2.2. Studi Literatur.....	28
3.2.3. Pengumpulan Data	28
3.2.4. Preprocess Data	28
3.2.5. Perancangan Program.....	28
3.2.6. Pembuatan Program	29
3.2.7. Pengujian & Evaluasi	29
3.3. Perancangan Sistem	29
3.4. Pengumpulan Data	31
3.5. Preprocess Data.....	32
3.6. Populasi Awal	33
3.7. K-Fold Cross Validation	34
3.8. Penghitungan Fitness dengan Algoritma KNN.....	35
3.8.1. Inisialisasi Parameter K	35
3.8.2. Menghitung Jarak	35
3.8.3. Mengurutkan Hasil Penghitungan Jarak.....	36
3.8.4. Menentukan Kategori yang Sering Muncul	37
3.8.5. Penentuan Akurasi Sebagai Fitness.....	38
3.9. Seleksi	39
3.10. Crossover	41
3.11. Mutasi.....	43
3.12. Populasi Baru	44
3.13. Fitness Terbaik.....	45
3.14. Skenario Pengujian.....	46
3.14.1. Skenario Pengujian KNN dengan Fold 5	47
3.14.2. Skenario Pengujian KNN + Genetika dengan Fold 5.....	47
3.14.3. Skenario Pengujian KNN dengan Fold 10	48
3.14.4. Skenario Pengujian KNN + Genetika dengan Fold 10.....	49
BAB IV PENGUJIAN DAN ANALISA	50
4.1. Metode Pengujian.....	50
4.2. Pengujian pada Dataset 1	50
4.2.1. Pengujian Fold 5.....	51

4.2.2. Pengujian Fold 10.....	53
4.3. Pengujian pada Dataset 2	55
4.3.1. Pengujian Fold 5.....	55
4.3.2. Pengujian Fold 10.....	58
4.4. Implementasi Deteksi Penyakit Diabetes Mellitus.....	60
4.5. Evaluasi Akhir.....	62
BAB V PENUTUP	65
5.1. Kesimpulan	65
5.2. Saran.....	66
DAFTAR PUSTAKA	67
LAMPIRAN.....	70

DAFTAR GAMBAR

Gambar 2.1. Contoh Crossover	15
Gambar 2.2. Contoh Mutasi	16
Gambar 2.3. Contoh K-Fold	24
Gambar 3.1. Diagram Alur Penelitian.....	27
Gambar 3.2. Diagram Alur Rancangan Sistem.....	30
Gambar 3.3. Dataset 1: <i>Pima Indians Diabetes Database</i>	31
Gambar 3.4. Dataset 2: <i>Early Stage Diabetes Risk Prediction Dataset</i>	31
Gambar 3.5. Dataset 1 Preprocess.....	32
Gambar 3.5. Dataset 2 Preprocess.....	33
Gambar 4.1. Grafik Hasil Pengujian KNN dengan Fold 5 Dataset 1	51
Gambar 4.2. Grafik Hasil Pengujian KNN + Genetika dengan Fold 5 Dataset 1. 52	
Gambar 4.3. Grafik Hasil Pengujian KNN dengan Fold 10 Dataset 1	53
Gambar 4.4. Grafik Hasil Pengujian KNN + Genetika dengan Fold 10 Dataset 154	
Gambar 4.5. Grafik Hasil Pengujian KNN dengan Fold 5 Dataset 2	56
Gambar 4.6. Grafik Hasil Pengujian KNN + Genetika dengan Fold 5 Dataset 2. 57	
Gambar 4.7. Grafik Hasil Pengujian KNN dengan Fold 10 Dataset 2	58
Gambar 4.8. Grafik Hasil Pengujian KNN + Genetika dengan Fold 10 Dataset 259	
Gambar 4.9. Tampilan Sistem Deteksi Penyakit Diabetes	60
Gambar 4.10. Tampilan Hasil Prediksi Sistem Deteksi Penyakit Diabetes	61
Gambar 4.11. Tampilan Hasil Generasi Ulang	62

DAFTAR TABEL

Tabel 3.1. Inisialisasi Populasi Awal	34
Tabel 3.2. Jumlah Data Tiap Fold	35
Tabel 3.3. Tabel Data Penghitungan Jarak.....	36
Tabel 3.4. Tabel Pengurutan Data Penghitungan Jarak	37
Tabel 3.5. Tabel Data Terdekat Sejumlah K.....	38
Tabel 3.6. Tabel Confusion Matrix	38
Tabel 3.7. Fitness Tiap Individu	40
Tabel 3.8. Proses Tournament Selection.....	41
Tabel 3.9. Proses Crossover	42
Tabel 3.10. Proses Mutasi	43
Tabel 3.11. Populasi Baru	44
Tabel 3.12. Fitness Tiap Individu	46
Tabel 3.13. Skenario Pengujian KNN dengan Fold 5.....	47
Tabel 3.14. Skenario Pengujian KNN + Genetika dengan Fold 5	48
Tabel 3.15. Skenario Pengujian KNN dengan Fold 10.....	48
Tabel 3.16. Skenario Pengujian KNN + Genetika dengan Fold 10	49
Tabel 4.1. Hasil Pengujian KNN dengan Fold 5 Dataset 1.....	51
Tabel 4.2. Hasil Pengujian KNN + Genetika dengan Fold 5 Dataset 1	52
Tabel 4.3. Hasil Pengujian KNN dengan Fold 10 Dataset 1.....	53
Tabel 4.4. Hasil Pengujian KNN + Genetika dengan Fold 10 Dataset 1	54
Tabel 4.5. Hasil Pengujian KNN dengan Fold 5 Dataset 2.....	55
Tabel 4.6. Hasil Pengujian KNN + Genetika dengan Fold 5 Dataset 2	57
Tabel 4.7. Hasil Pengujian KNN dengan Fold 10 Dataset 2.....	58
Tabel 4.8. Hasil Pengujian KNN + Genetika dengan Fold 10 Dataset 2	59
Tabel 4.9. Perbandingan Hasil Pengujian	63

Halaman ini sengaja dikosongkan