

SKRIPSI

**IMPLEMENTASI KNN DENGAN OPTIMASI ALGORITMA GENETIKA
SEBAGAI SELEKSI FITUR DAN NILAI K UNTUK KLASIFIKASI
PENYAKIT STROKE**

*IMPLEMENTATION OF K-NEAREST NEIGHBOR WITH GENETIC
ALGORITHM OPTIMIZATION FOR FEATURE SELECTION AND K
VALUE IN STROKE DISEASE CLASSIFICATION*



ASRUL AFANDI

D0220390

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS SULAWESI BARAT

2025

SKRIPSI

**IMPLEMENTASI KNN DENGAN OPTIMASI ALGORITMA GENETIKA
SEBAGAI SELEKSI FITUR DAN NILAI K UNTUK KLASIFIKASI
PENYAKIT STROKE**

Diajukan Sebagai salah satu syarat untuk menyelesaikan pendidikan pada
program studi informatika, Fakultas teknik, Universitas Sulawesi Barat



ASRUL AFANDI

D0220390

PROGRAM STUDI INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS SULAWESI BARAT

2025

LEMBAR PENGESAHAN

**IMPLEMENTASI KNN DENGAN OPTIMASI ALGORITMA GENETIKA
SEBAGAI SELEKSI FITUR DAN NILAI K UNTUK KLASIFIKASI
PENYAKIT STROKE**

SKRIPSI

Untuk memenuhi sebagian persyaratan
memperoleh gelar Sarjana Komputer

Disusun oleh:

ASRUL AFANDI
NIM. DD220390

Skrripsi ini telah diuji dan dinyatakan lulus
pada 30 / 10 / 2025

Telah diperiksa dan disetujui oleh:

Pembimbing I

Arnita Irianti, S.Si., M.Si
NIP. 198708062018032001

Pembimbing II

A. Amirul Asnan Cirua, S.T., M.Kom
NIP. 199804022024061001

Dekan Fakultas Teknik,
Universitas Sulawesi Barat



Dr. Ir. Hafsah Nirwana, M.T
NIP. 196404051990032002

Ketua Program Studi
Informatika,



Muh. Rafli Rasyid, S.Kom., M.T
NIP. 198808182022031006

HALAMAN PERSETUJUAN
PENELITIAN S1
IMPLEMENTASI KNN DENGAN OPTIMASI ALGORITMA GENETIKA
SEBAGAI SELEKSI FITUR DAN NILAI K UNTUK KLASIFIKASI
PENYAKIT STROKE
SKRIPSI

Untuk memenuhi Sebagian persyaratan memperoleh gelar Sarjana Teknik

Telah dipersiapkan dan disusun oleh

ASRUL AFANDI

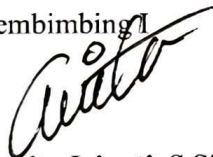
D0220390

Telah dipertahankan didepan Tim Penguji

Pada Tanggal 30 /10 /2025

Susunan Tim Penguji

Pembimbing I



Arnita Irianti, S.Si., M.Si

NIP. 198708062018032001

Penguji I



Dr. Eng. Sulfayanti, S.Si., M.T

NIP. 198903172020122011

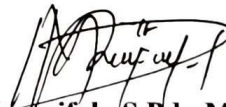
Pembimbing II



A. Amirul Asnan Cirua, S.T., M.Kom

NIP. 199804022024061001

Penguji II



Muskrifah, S.Pd., M.Pd

NIDN. 0014119302

Penguji III



Nurani Natsir, S. Si, M. Si

NUP: 9990549367

PERNYATAAN ORISNILITAS

Dengan ini saya menyatakan bahwa Skripsi ini tidak terdapat karya yang pernah diajukan untuk memperoleh gelar kesarjanaan di suatu Perguruan Tinggi dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam dalam daftar Pustaka.

Tinambung, 28 Oktober 2025



ASRUL AFANDI

NIM.D0220390

ABSTRAK

Stroke merupakan gangguan neurologis serius yang terjadi akibat terganggunya suplai darah ke otak, baik karena penyumbatan pembuluh darah maupun pecahnya pembuluh darah. Penelitian ini mengembangkan model klasifikasi risiko stroke menggunakan metode *K-Nearest Neighbor* (KNN) yang dioptimasi dengan Algoritma Genetika (GA) untuk seleksi fitur dan penentuan nilai K terbaik. Dataset yang digunakan terdiri dari 40.910 data pasien dengan 10 atribut input dan 1 atribut target dari dataset Kaggle yang berjudul *Diabetes, Hypertension and Stroke Prediction*. Melalui proses optimasi, diperoleh nilai K optimal sebesar 4 dengan 8 fitur terpilih, yaitu *hypertension*, *heart_disease*, *ever_married*, *work_type*, *Residence_type*, *avg_glucose_level*, *bmi*, dan *smoking_status*. Model optimal ini mencapai akurasi klasifikasi sebesar 98.15%% pada pembagian data training dan testing 80:20, meningkat signifikan dibandingkan metode KNN konvensional yang menggunakan semua fitur dan nilai K 4 dengan akurasi 93.97%. Pendekatan ini membuktikan efektivitas optimasi GA dalam meningkatkan performa klasifikasi dan efisiensi model melalui pengurangan fitur yang tidak relevan, sekaligus mempertahankan kemampuan deteksi risiko stroke secara akurat.

Kata Kunci—*K-Nearest Neighbors, KNN, Algoritma Genetika, GA, Seleksi Fitur, K Optimal, Penyakit Stroke*

ABSTRACT

Stroke is a serious neurological disorder caused by disrupted blood supply to the brain, either due to blood vessel blockage or rupture. This study develops a stroke risk classification model using the K-Nearest Neighbors (KNN) method optimized with the Genetic Algorithm (GA) for feature selection and determination of the best K value. The dataset used consists of 40,910 patient records with 10 input attributes and 1 target attribute from the Kaggle dataset titled Diabetes, Hypertension and Stroke Prediction. Through optimization, an optimal K value of 4 was obtained along with 8 selected features: hypertension, heart_disease, ever_married, work_type, Residence_type, avg_glucose_level, bmi, and smoking_status. This optimized model achieves a classification accuracy of 98.15%% with an 80:20 training-testing data split, significantly improving over the conventional KNN method which used all features and K=4 with an accuracy of 93.97%. This approach demonstrates the effectiveness of GA optimization in enhancing classification performance and model efficiency by reducing irrelevant features while maintaining accurate stroke risk detection.

Keywords – *K-Nearest Neighbors, Genetic Algorithm, Feature Selection, K Optimal, Stroke.*

BAB I

PENDAHULUAN

1.1 Latar Belakang

Stroke merupakan masalah kesehatan utama bagi masyarakat modern. Pada saat ini, *stroke* semakin menjadi masalah serius yang dihadapi hampir diseluruh dunia. Hal tersebut dikarenakan serangan *stroke* yang mendadak dapat mengakibatkan kematian, kecacatan fisik dan mental baik pada usia produktif maupun usia lanjut (Teknika & Estian Pambudi, 2022). *Stroke* dapat dicegah melalui deteksi dini dan pengendalian faktor risiko. Cara ini butuh biaya dan waktu karena harus ke rumah sakit. Untuk membaantu pasien dan masyarakat, memerlukan suatu model yang dapat mengkasifikasi pasien penyakit *stroke* serta memprediksi resiko penyakit *stroke* agar bisa dihindari (Zuriati & Qomariyah, 2022).

Solusi penelitian ini adalah memanfaatkan data penyakit *stroke* melalui teknik data mining untuk menggali informasi baru. Metode yang digunakan yaitu klasifikasi, yang bertujuan mengelompokkan variabel target secara akurat berdasarkan variabel lain yang sudah diketahui (Zuriati & Qomariyah, 2022) Terdapat beberapa algoritma yang dapat digunakan untuk klasifikasi seperti *K-Nearest Neighbor*, *Decision Tree*, *Naïve Bayes*, dan algoritma lainnya. Pada penelitian ini menggunakan algoritma *K-Nearest Neighbor* untuk klasifikasi penyakit *stroke* (Riany & Testiana, 2023). Metode KNN memiliki beberapa kelebihan yaitu, pelatihan sangat cepat, sederhana, efektif pada data pelatihan yang

besar. Sedangkan kekurangan KNN sendiri adalah nilai k bias harus ditentukan, komputasi kompleks, keterbatasan memori, dan dapat tertipu oleh atribut yang tidak relevan (Amien et al., 2023). Untuk mengatasi hal ini dibutuhkan proses optimasi fitur yang paling relevan untuk digunakan dalam proses klasifikasi, serta pemilihan nilai K yang optimal yang akan menjadi faktor penting untuk meningkatkan kinerja KNN. *Algoritma Genetika* dapat menentukan nilai k dan atribut mana yang paling optimal, *Algoritma Genetika* dan *KNN* juga dapat dan mampu meningkatkan nilai akurasi dari KNN (Ulya et al., 2021) dan.

Pada penelitian terdahulu oleh (Zuriati & Qomariyah, 2022) tentang tentang Klasifikasi Penyakit *Stroke* Menggunakan Algoritma *K-Nearest Neighbor* (KNN). Dengan dataset yang terdiri dari 5110 record. Pada hasil evaluasi kinerja KNN dengan *Confusion matrix* dan perhitungan akurasi, pefornma terbaik dari KNN di dapatkan pada nilai K=5 dengan akurasi sebesar 93.54% . Di penelitian ini algoritma KNN terbukti mampu memberikan akurasi yang cukup baik pada dataset *stroke*.

Penelitian berikutnya oleh (Amien et al., 2023) mengoptimasi KNN menggunakan algoritma genetika untuk deteksi penyakir diabetes mellitus dengan dua dataset, yaitu Pima Indians Diabetes Database dan Early Stage Diabetes Risk Prediction Dataset. Optimasi dilakukan melalui pencarian kombinasi antara nilai K dan fitur menggunakan mekanisme seleksi, *crossover*, serta mutasi pada GA. Hasil penelitian menunjukkan adanya peningkatan akurasi pada dataset 1 dari 74.2% menjadi 79.1% dan pada dataset 2 dari 97.5% menjadi 98.2% Kesimpulannya, penerapan GA ini terbukti efektif dalam meningkatkan peforma KNN.

Pada penelitian berikutnya (Supriana et al., 2024) menunjukkan adanya peningkatan akurasi menggunakan algoritma genetika untuk seleksi nilai k pada metode modified *K-Nearest Neighbors* (MKNN) untuk klasifikasi penyakit ginjal kronis, disini algoritma genetika digunakan untuk mencari nilai K optimal berdasarkan performa klasifikasi (*fitness function*). Dataset yang berisi 24 fitur klinis yang digunakan untuk pengujian. Hasil pengujian menunjukkan nilai K optimal Adalah 3 dengan akurasi 93%, presisi 93.2%, dan recall 93.2%, lebih baik jika dibandingkan dengan knn tanpa genetika algoritma dengan maksimal akurasi 86%. MKNN yang di optimasi dengan GA menghasilkan peningkatan signifikan dalam performa klasifikasi

1.2 Rumusan Masalah

Berdasarkan latar belakang diatas, maka rumusan masalah pada penelitian ini adalah tentang bagaimana mengimplementasikan metode *K-Nearest Neighbors* (KNN) dengan optimasi *Genetic Algorithm* (GA) untuk seleksi fitur dan nilai K pada KNN pada kasus klasifikasi penyakit *stroke*.

1.3 Tujuan Penelitian

Berdasarkan dari rumusan masalah, maka tujuan dari penelitian ini ialah untuk mengetahui hasil implementasi dari klasifikasi penyakit *stroke* dengan metode *K-Nearest Neighbors* (KNN) dengan *Genetic Algorithm* (GA).

1.4 Batasan Masalah

Ruang lingkup penelitian ini dibatasi pada penerapan Algoritma genetika (GA) sebagai metode seleksi fitur pada algoritma *K-Nearest Neighbours* (KNN). Penelitian ini difokuskan untuk meningkatkan performa akurasi klasifikasi KNN dalam mengklasifikasi penyakit stroke. Data yang digunakan merupakan data *Stroke* dari dataset “*Diabetes, Hypertension, and Stroke Prediction*” yang diperoleh dari situs website *open dataset Kaggle*.

1.5 Manfaat Penelitian

Penelitian ini diharapkan memberikan manfaat pada bidang informatika, khususnya pada konteks analisis data. Adapun beberapa manfaatnya:

1. Meningkatkan pengalaman mahasiswa dalam meningkatkan kemampuan generalisasi model, sehingga model yang dihasilkan dapat diterapkan pada penelitian berikutnya.
2. Meningkatkan pemahaman tentang faktor-faktor apa saja yang mempengaruhi penyakit *stroke*.
3. Memberikan kontribusi dalam peningkatan akurasi pada klasifikasi penyakit *stroke*.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil implementasi serta pengujian algoritma K-Nearest Neighbor (KNN) yang dioptimasi menggunakan algoritma genetika sebagai metode seleksi fitur dan penentuan nilai K, diperoleh bahwa pendekatan ini efektif dalam meningkatkan performa klasifikasi data. Algoritma genetika mampu mengidentifikasi fitur-fitur yang paling relevan, mengurangi dimensi data, dan sekaligus menentukan nilai K yang optimal untuk menghasilkan tingkat akurasi terbaik. Kombinasi fitur yang terpilih sebanyak 8 atribut, yaitu *hypertension*, *heart_disease*, *ever_married*, *work_type*, *Residence_type*, *avg_glucose_level*, *bmi*, dan *smoking_status*. dengan nilai K optimal sebesar 4 menghasilkan akurasi tertinggi sebesar 98.15% pada rasio data 80:20.

Selisih akurasi antara hasil optimasi dengan KNN konvensional hanya sekitar 4,18%, menunjukkan bahwa algoritma genetika mampu meningkatkan akurasi tanpa menambah kompleksitas model secara signifikan. Model ini dapat dijadikan pilihan yang efisien dan andal untuk klasifikasi data dengan jumlah fitur besar, karena mampu menyeimbangkan antara akurasi tinggi dan efisiensi komputasi.

5.2 Saran

Berdasarkan hasil penelitian yang telah dilakukan, terdapat beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya. Pertama, algoritma

genetika dapat dikembangkan lebih lanjut dengan menyesuaikan parameter seperti ukuran populasi, jumlah generasi, crossover rate, dan mutation rate agar proses optimasi menjadi lebih adaptif terhadap variasi data yang lebih kompleks. Kedua, perlu dilakukan pengujian menggunakan algoritma pembelajaran lain seperti *Support Vector Machine (SVM)*, *Random Forest*, atau *Neural Network* sebagai pembandingan untuk mengetahui sejauh mana efektivitas algoritma genetika dalam meningkatkan performa klasifikasi.

Selain itu, penelitian berikutnya juga disarankan untuk menggunakan metode validasi silang (*cross-validation*) dalam evaluasi agar hasil yang diperoleh lebih stabil dan tidak bergantung pada satu kali pembagian data. Terakhir, penerapan metode KNN yang dioptimasi dengan algoritma genetika dapat diperluas ke dataset dengan domain berbeda, seperti kesehatan, keuangan, atau pendidikan, guna menguji konsistensi dan generalisasi model terhadap berbagai jenis data.

DAFTAR PUSTAKA

- Amien, I. L. F., Astuti, W., & Lhaksamana, K. M. (2023). Perbandingan Metode Naïve Bayes dan KNN (K-Nearest Neighbor) dalam Klasifikasi Penyakit Diabetes. *E-Proceeding of Engineering*, 10(2), 1911–1920.
- Amir Mahmud Husein, Haifa Muntaza, L., Grescilya Sinaga, Khoirulliza, & Maychilia Lamerossa Br Sembiring. (2025). Model CNN-LSTM Untuk Klasifikasi Tingkat Stres Mahasiswa Dalam Menghadapi Ujian Menggunakan Data Elektrokardiogram (EKG). *Decode: Jurnal Pendidikan Teknologi Informasi*, 5(1), 191–201. <https://doi.org/10.51454/decode.v5i1.1080>
- Arapah, R. N., & Fitriyani. (2024). Optimasi Algoritma Genetika Untuk Memprediksi Penyakit Hepatitis Dengan Menggunakan Metode K-Nearest Neighbor (Knn). *Kohesi: Jurnal Multidisiplin Saintek*, 3(1), 11–21.
- Darmawan, M. A. V., Haromainy, M. M. Al, & Junaidi, A. (2025). Optimasi Algoritma K-Nearest Neighbor Dengan Algoritma Genetika Pada Deteksi Penyakit Diabetes Mellitus. *JATISI (Jurnal Teknik Informatika Dan Sistem Informasi)*, 12(2), 247–259. <https://doi.org/10.35957/jatisi.v12i2.11353>
- Kecerdasan Buatan, J., dan Teknologi Informasi, K., Rizal, F., Nurul Jadid, U., Artikel, R., & Kunci Mangga, K. (2021). *Perbandingan Algoritma K-Nearest Neighbor Untuk Klasifikasi Jenis Mangga Menggunakan Berdasarkan Fitur Gray Level Co-Occurrence Matric dan Fitur Warna*. 2(2). <https://ejournal.unuja.ac.id/index.php/core>
- Khotiah, T., Sugianela, Y., Setiowati, D., & Arianto, F. (2022). MODEL

KLASIFIKASI DATA REMISI NARAPIDANA MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS DENGAN OPTIMASI ALGORITMA GENETIKA Classification Model of Prisoner Remission Data Using K-Nearest Neighbors Algorithm with Genetic Algorithm Optimization. *Jurnal Ilmiah Intech : Information Technology Journal of UMUS*, 4(01), 87–96. <http://jurnal.umus.ac.id/index.php/intech/article/download/709/445>

Prasetio, R. T. (2020). SELEKSI FITUR DAN OPTIMASI PARAMETER k-NN BERBASIS ALGORITMA GENETIKA PADA DATASET MEDIS. *JURNAL RESPONSIF*, 2(2), 213–221. <http://ejournal.ars.ac.id/index.php/jti>

Riany, A. F., & Testiana, G. (2023). Penerapan Data Mining untuk Klasifikasi Penyakit Stroke Menggunakan Algoritma Naïve Bayes. *Jurnal SAINTEKOM*, 13(1), 42–54. <https://doi.org/10.33020/saintekom.v13i1.352>

Supriana, I. W., Pramatha, C., Putra, I. G. N. A. C., Raharja, M. D. A., & Wiguna, P. P. K. (2024). *Modified K-Nearest Neighbor Optimization with Genetic Algorithm in Chronic Kidney Disease Classification* (Issue Icamsac). Atlantis Press International BV. https://doi.org/10.2991/978-94-6463-413-6_20

Teknika, J., & Estian Pambudi, R. (2022). klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C4.5. *Jurnal Teknika*, x, No.x(x), 1–5.

Ulya, S., Soeleman, M. A., & Budiman, F. (2021). Optimasi Parameter K Pada Algoritma K-NN Untuk Klasifikasi Prioritas Bantuan Pembangunan Desa. *Techno.Com*, 20(1), 83–96. <https://doi.org/10.33633/tc.v20i1.4215>

Zuriati, Z., & Qomariyah, N. (2022). Klasifikasi Penyakit Stroke Menggunakan Algoritma K-Nearest Neighbor (KNN). *ROUTERS: Jurnal Sistem Dan Teknologi Informasi*, 1(1), 1–8. <https://doi.org/10.25181/rt.v1i1.2665>