

SKRIPSI

**ANALISIS PERBANDINGAN PREDIKSI PENYAKIT
DIABETES MENGGUNAKAN ALGORITMA *LOGISTIC*
REGRESSION DAN *K-NEAREST NEIGHBOR***



**Muhammad Fajar Islam Imam Mujahid
1519620027**

**PROGRAM STUDI SISTEM DAN TEKNOLOGI INFORMASI
FAKULTAS TEKNIK
UNIVERSITAS NEGERI JAKARTA
2025**

LEMBAR PENGESAHAN PROPOSAL SKRIPSI

Judul : ANALISIS PERBANDINGAN PREDIKSI PENYAKIT
DIABETES MENGGUNAKAN ALGORITMA LOGISTIC
REGRESSION DAN K-NEAREST NEIGHBOR

Penyusun : Muhammad Fajar Islam Imam Mujahid

NIM : 1519620027

Proposal skripsi telah didiskusikan dan diusulkan dari Dosen Pembimbing:

NAMA DOSEN	TANDA TANGAN	TANGGAL
Ali Idrus, M.Kom. (Dosen Pembimbing 1)		12 Desember 2024
Lipur Sugiyanta, Ph.D (Dosen Pembimbing 2)		24 Desember 2024

Telah disetujui oleh:

NAMA DOSEN	TANDA TANGAN	TANGGAL
Fuad Mumtas, M.TI (Dosen Penguji 1)		23 Januari 2025
Irma Permata Sari, M.Eng. (Dosen Penguji 2)		3 Februari 2025

LEMBAR PENGESAHAN SKRIPSI

Judul : ANALISIS PERBANDINGAN PREDIKSI PENYAKIT DIABETES
MENGGUNAKAN ALGORITMA LOGISTIC REGRESSION DAN
K-NEAREST NEIGHBOR
Penyusun : Muhammad Fajar Islam Imam Mujahid
NIM : 1519620027

Skripsi telah didiskusikan dan diusulkan dari Dosen Pembimbing:

NAMA DOSEN	TANDA TANGAN	TANGGAL
Ali Idrus, S.Kom. (Dosen Pembimbing 1)		26 Juni 2025
Lipur Sugiyanta, Ph.D. (Dosen Pembimbing 2)		2 Juli 2025

Telah disetujui oleh:

NAMA DOSEN	TANDA TANGAN	TANGGAL
Murien Nugraheni, M.Cs. (Dosen Penguji 1)		29 Juli 2025
Irma Permata Sari M.Eng.. (Dosen Penguji 2)		23 Juli 2025

LEMBAR PERNYATAAN

Dengan ini saya menyatakan bahwa:

1. Skripsi ini yang berjudul “ANALISIS PERBANDINGAN PREDIKSI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA *LOGISTIC REGRESSION* DAN *K-NEAREST NEIGHBOR*” merupakan karya asli dan belum pernah diajukan untuk mendapatkan gelar akademik sarjana, baik di Universitas Negeri Jakarta maupun di Perguruan Tinggi lain.
2. Skripsi ini belum dipublikasikan, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
3. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan dan ketidakbenaran, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh, serta sanksi lainnya sesuai dengan norma yang berlaku di Universitas Negeri Jakarta.

Jakarta, 30 Juli 2025



Muhammad Fajar Islam Imam Mujahid



KEMENTERIAN PENDIDIKAN, KEBUDAYAAN, RISET DAN TEKNOLOGI
UNIVERSITAS NEGERI JAKARTA
UPT PERPUSTAKAAN

Jalan Rawamangun Muka Jakarta 13220
Telepon/Faksimili: 021-4894221
Laman: lib.unj.ac.id

**LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS**

Sebagai sivitas akademika Universitas Negeri Jakarta, yang bertanda tangan di bawah ini, saya:

Nama : Muhammad Fajar Islam Imam Mujahid
NIM : 1519620027
Fakultas/Prodi : Teknik/Sistem dan Teknologi Informasi
Alamat email : fajar12345678905@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada UPT Perpustakaan Universitas Negeri Jakarta, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah:

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain (... ..)

yang berjudul :

ANALISIS PERBANDINGAN PREDIKSI PENYAKIT DIABETES MENGGUNAKAN

ALGORITMA LOGISTIC REGRESSION DAN K-NEAREST NEIGHBOR

Dengan Hak Bebas Royalti Non-Eksklusif ini UPT Perpustakaan Universitas Negeri Jakarta berhak menyimpan, mengalihmediakan, mengelolanya dalam bentuk pangkalan data (*database*), mendistribusikannya, dan menampilkan/mempublikasikannya di internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan Universitas Negeri Jakarta, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini saya buat dengan sebenarnya.

Jakarta 3 Agustus 2025

Penulis

(Muhammad Fajar Islam)
Imam Mujahid

KATA PENGANTAR

Puji syukur penulis panjatkan kepada Allah SWT atas segala limpahan rahmat dan karunia-Nya, sehingga dapat menyelesaikan penyusunan Skripsi dengan judul **“ANALISIS PERBANDINGAN PREDIKSI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA LOGISTIC REGRESSION DAN K-NEAREST NEIGHBOR”**, yang diajukan untuk memenuhi syarat kelulusan dalam meraih gelar Sarjana Ilmu Komputer di Program Studi Sistem dan Teknologi Informasi, Fakultas Teknik, Universitas Negeri Jakarta.

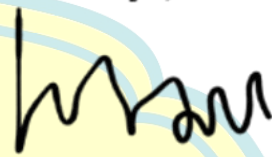
Penulis menyadari bahwa proses penusunan dan penulisan skripsi ini mengalami berbagai kesulitan. Namun berkat ikhtiar, doa, bimbingan, saran dan dukungan dari berbagai pihak, penulis dapat menyelesaikan penulisan skripsi ini. Oleh karena itu, penulis ingin mengucapkan terima kasih sebesar-besarnya kepada:

1. Kedua orang tua yang penulis hormati dan sayangi, Bapak Teguh dan Ibu Mayasari yang senantiasa mendoakan yang terbaik, memberikan dukungan dan semangat baik secara moril dan materiil kepada penulis.
2. Bapak Lipur Sugiyanta, Ph.D, sebagai Koordiantor Program Studi Sistem dan Teknologi Informasi, serta sebagai Dosen Pembimbing II yang telah memberikan kesediaan waktunya kapanpun dalam memberikan bimbingan dan arahan dalam penyusunan skripsi dengan sangat memuaskan.
3. Bapak Ali Idrus, S.Kom., sebagai Dosen Pembimbing I yang memberikan bimbingan, dukungan moril, serta ruang konsultasi baik dalam penyusunan skripsi, proyek, serta lingkup akademik.
4. Ibu Murien Nugraheni, S.T, M.Cs. dan Irma Permatasari, M.Eng. selaku Dosen Penguji yang telah memberikan masukan dan saran dalam perbaikan penyusunan skripsi.
5. Teman-Teman baik di dalam ataupun di luar lingkungan UNJ yang memberikan ruang berinteraksi, bertukar pikiran, serta berekspresi dalam penyusunan skripsi ini.
6. Seluruh pihak yang terlibat secara langsung maupun tidak langsung yang tidak bisa disebutkan satu persatu.

Penulis menyadari bahwa skripsi ini masih jauh dari kata sempurna, maka dari itu penulis menyampaikan permohonan maaf apabila terdapat kekurangan dan kesalahan dalam proses penulisan skripsi ini. Penulis terbuka terhadap kritik dan saran yang diberikan pembaca untuk menyempurnakan skripsi ini. Akhir kata, penulis berharap agar penelitian dan penyusunan skripsi ini dapat bermanfaat bagi pembaca dan khususnya bagi penulis.

Jakarta, 7 Juli 2025

Penyusun,



Muhammad Fajar Islam
Imam Mujahid



ANALISIS PERBANDINGAN PREDIKSI PENYAKIT DIABETES MENGUNAKAN ALGORITMA LOGISTIC REGRESSION DAN K- NEAREST NEIGHBOR

Muhammad Fajar Islam Imam Mujahid
Ali Idrus S.Kom., dan Lipur Sugiyanta, Ph.D.

ABSTRAK

Penyakit diabetes merupakan salah satu penyakit kronis yang prevalensinya terus meningkat setiap tahun, baik secara global maupun nasional. Dalam upaya mendeteksi diabetes secara dini, pemanfaatan teknologi *machine learning* menjadi salah satu alternatif yang menjanjikan. Penelitian ini bertujuan untuk menganalisis dan membandingkan performa algoritma *Logistic Regression* dan *K-Nearest Neighbor* (KNN) dalam klasifikasi data diabetes. Dataset yang digunakan diperoleh dari *National Institute of Diabetes and Digestive and Kidney Diseases* melalui platform *Kaggle*, yang terdiri dari delapan atribut gejala pasien dan satu atribut target (*outcome*).

Proses penelitian dimulai dengan tahap *preprocessing* data, dilanjutkan dengan penerapan masing-masing algoritma baik sebelum maupun sesudah dilakukan *hyperparameter tuning*. Pada *Logistic Regression*, parameter yang dituning mencakup *lambda*, *solver*, dan *maximum iteration*, sedangkan pada KNN, parameter yang diujikan meliputi jumlah tetangga (*k*), *uniform*, dan parameter jarak. Evaluasi performa model dilakukan dengan menggunakan metrik akurasi, *precision*, *recall*.

Hasil penelitian menunjukkan bahwa setelah dilakukan *hyperparameter tuning*, algoritma *Logistic Regression* menghasilkan akurasi sebesar 84% sedangkan KNN mencapai akurasi sebesar 81,43%. Temuan ini mengindikasikan bahwa pilihan parameter berpengaruh signifikan terhadap performa model. Oleh karena itu, pemilihan dan pengujian berbagai kombinasi parameter sangat penting dalam proses perancangan sistem klasifikasi penyakit berbasis *machine learning*.

Kata Kunci : Diabetes, *Logistic Regression*, *K-Nearest Neighbor*, *Machine Learning*, *Hyperparameter Tuning*, Klasifikasi.

COMPARATIVE ANALYSIS OF DIABETES DISEASE PREDICTION USING LOGISTIC REGRESSION AND K-NEAREST NEIGHBOR ALGORITHMS

Muhammad Fajar Islam Imam Mujahid
Ali Idrus S.Kom., dan Lipur Sugiyanta, Ph.D.

ABSTRACT

Diabetes is one of the chronic diseases whose prevalence continues to increase every year, both globally and nationally. In an effort to detect diabetes early, the use of machine learning technology has become a promising alternative. This study aims to analyze and compare the performance of the Logistic Regression and K-Nearest Neighbor (KNN) algorithms in classifying diabetes data. The dataset used was obtained from the National Institute of Diabetes and Digestive and Kidney Diseases via the Kaggle platform, consisting of eight patient symptom attributes and one target attribute (outcome).

The research process began with the data preprocessing stage, followed by the application of each algorithm both before and after hyperparameter tuning. In Logistic Regression, the parameters tuned included lambda, solver, and maximum iteration, while in KNN, the parameters tested included the number of neighbors (k), uniformity, and distance parameters. Model performance was evaluated using accuracy, precision, and recall metrics.

The results show that after hyperparameter tuning, the Logistic Regression algorithm produced an accuracy of 84%, while KNN achieved an accuracy of 81.43%. These findings indicate that parameter selection has a significant effect on model performance. Therefore, selecting and testing various parameter combinations is very important in the process of designing a machine learning-based disease classification system.

Keywords : Diabetes, Logistic Regression, K-Nearest Neighbor, Machine Learning, Hyperparameter Tuning, Classification

DAFTAR ISI

DAFTAR ISI.....	x
DAFTAR TABEL.....	xii
DAFTAR GAMBAR.....	xiii
DAFTAR RUMUS	xvii
DAFTAR LAMPIRAN	xviii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang Masalah.....	1
1.2 Identifikasi Masalah	3
1.3 Pembatasan Masalah	4
1.4 Perumusan Masalah	5
1.5 Tujuan dan Manfaat Penelitian	5
BAB II KERANGKA TEORITIK.....	6
2.1 Kerangka Teoritik	6
2.1.1 Diabetes	6
2.1.2 Machine Learning.....	6
2.1.3 Data Mining.....	8
2.1.4 KDD (Knowlegde Discovery in Databases).....	9
2.1.5 Logistic Regretion	10
2.1.6 K-NN	11
2.1.7 Flowchart	12
2.1.8 Confusion Matrix	13
2.2 Penelitian Relevan.....	14
2.3 Kerangka Berpikir.....	16
BAB III METODE PENELITIAN	18
3.1 Alat dan Bahan Penelitian.....	18
3.1.1 Alat (Hardware (Perangkat Keras)).....	18
3.1.2 Bahan (Software (Perangkat Lunak))	18
3.2 Alur Penelitian	19
3.3 Penjebaran Alur Penelitian.....	21

BAB IV HASIL DAN PEMBAHASAN.....	35
4.1 Hasil Penelitian	35
4.1.1 Perancangan dan pengujian Algoritma <i>Logistic Regression</i>	35
4.1.2 Perancangan dan pengujian Algoritma K-Nearest Neighbor	45
4.1.3 Perancangan dan pengujian Algoritma <i>Logistic Regression</i>	57
4.1.4 Perancangan dan Pengujian Algoritma <i>K-Nearest Neighbor</i>	60
4.2 Pembahasan.....	62
BAB V KESIMPULAN DAN SARAN.....	66
5.1 Kesimpulan	66
5.2 Saran.....	66
DAFTAR PUSTAKA.....	68
LAMPIRAN.....	72

