

**ANALISIS KOMPARATIF ALGORITMA *NAIVE BAYES* DAN
DECISION TREE UNTUK PREDIKSI PENYAKIT DIABETES
BERBASIS WEB
TUGAS AKHIR**

Diajukan Untuk Memenuhi Salah Satu Syarat Mencapai Gelar Sarjana Komputer
Program Studi Informatika
Universitas Sahid Surakarta



Di Susun Oleh :

XYLA SYARIFATUZZAHRA ARYVENANDA

NIM 2024062006

**PROGRAM STUDI INFORMATIKA
FAKULTAS SAINS, TEKNOLOGI DAN KESEHATAN
UNIVERSITAS SAHID SURAKARTA
2026**

LEMBAR PERSETUJUAN

ANALISIS KOMPARATIF ALGORITMA *NAIVE BAYES* DAN *DECISION TREE* UNTUK PREDIKSI PENYAKIT DIABETES BERBASIS WEB

Disusun Oleh:

XYLA SYARIFATUZZAHRA ARYVENANDA

NIM. 2024062006

Tugas Akhir ini telah disetujui untuk disusun dan ditindaklanjuti

pada tanggal : 22 Juli 2020

Pembimbing I



Farid Fitriyadi, S.Kom., M.Kom
NIDN. 0607058902

Pembimbing II



Ir. Dahlan Susilo, M.Kom
NIDN. 0613016701

Mengetahui,

Ketua Program Studi Informatika



Diyah Ruswanti, S.Kom., M.Kom
NIDN. 0027018101

LEMBAR PENGESAHAN

ANALISIS KOMPARATIF ALGORITMA *NAIVE BAYES* DAN *DECISION TREE* UNTUK PREDIKSI PENYAKIT DIABETES BERBASIS WEB

Disusun Oleh:

XYLA SYARIFATUZZAHRA ARYVENANDA

NIM. 2024062006

Tugas Akhir ini telah diterima dan disahkan
pada hari Palm tanggal: 22 Juli 2020

Dewan Penguji

Penguji 1

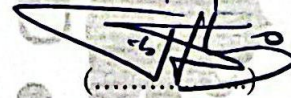
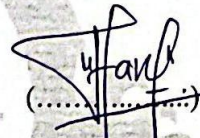
Farid Fitriyadi, S.Kom., M.Kom
NIDN. 0607058902

Penguji 2

Ir. Dahlan Susilo, M.Kom
NIDN. 0614016701

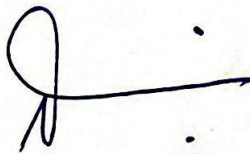
Penguji 3

Dwi Retnoningsih, ST.MT.
NIDN. 0529057501



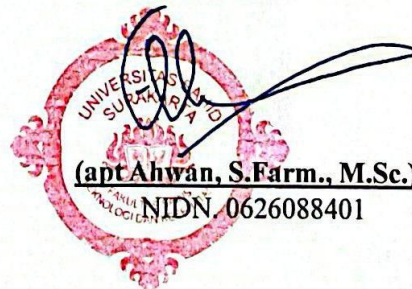
Mengetahui,

Kepala Program Studi



(Diyah Ruswanti, S.Kom., M.Kom)
NIDN. 0027018101

Dekan Fakultas Sains ,
Teknologi dan Kesehatan


(apt Ahwan, S.Farm., M.Sc.)
NIDN. 0626088401

SURAT PERNYATAAN ORISINIL KARYA ILMIAH

Saya yang bertanda tangan dibawah ini:

Nama : Xyla Syarifatuzzahra Aryvenanda

NIM : 2024062006

Program Studi : Informatika

Fakultas : Sains, Teknologi, dan Kesehatan

Menyatakan dengan sesungguhnya bahwa tugas akhir yang berjudul ANALISIS KOMPARATIF ALGORITMA *NAIVE BAYES* DAN *DECISION TREE* UNTUK PREDIKSI PENYAKIT DIABETES BERBASIS WEB. Adalah benar karya yang saya susun sendiri. Apabila dikemudian hari saya terbukti melakukan Tindakan menyalin atau meniru hasil karya orang lain seolah -olah hasil pemikiran saya sendiri, saya bersedia dan menerima sanksi sesuai peraturan yang berlaku di Universitas Sahid Surakarta termasuk pencabutan gelar yang telah saya peroleh.

Demikian surat pernyataan saya buat dengan sebenar-benarnya dan apabila dikemudian hari terbukti melakukan kebohongan maka saya sanggup menanggung segala konsekuensinya.

Surakarta, 17 Juli 2026

Yang Pembuatan Pernyataan



Xyla Syarifatuzzahra Aryvenanda

NIM: 2024062006

PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Sebagai Sivitas akademis Universitas Sahid Surakarta, saya yang bertanda tangan dibawah ini:

Nama : Xyla Syarifatuzzahra Aryvenanda

NIM : 2024062006

Program Studi : Informatika

Jenis Karya : Tugas Akhir

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Sahid Surakarta hak bebas royalti non-eksklusif atas tugas akhir saya yang berjudul: " ANALISIS KOMPARATIF ALGORITMA *NAIVE BAYES* DAN *DECISION TREE* UNTUK PREDIKSI PENYAKIT DIABETES BERBASIS WEB ". Beserta desain atau perangkat (jika ada). Berhak menyimpan, mengalihkan bentuk, mengalihmediakan, mengelola dalam bentuk pangkalan data (*basis data*), merawat serta mempublikasikan karya ilmiah saya selama tetap mencantumkan nama saya sebagai penulis (*author*) dan pembimbing sebagai *co-author* atau pencipta dan juga sebagai pemilik hak cipta. Demikian pernyataan yang saya buat dengan sesungguhnya secara sehat dan sadar tanpa ada paksaan dari pihak manapun.

Surakarta, 17 Juli 2026

Yang Pembuatan Pernyataan


Syarifatuzzahra Aryvenanda

NIM: 2024062006

KATA PENGANTAR

Puji syukur penulis panjatkan ke hadirat Allah SWT atas segala rahmat, taufik, dan hidayah-Nya sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul “Analisis Komparatif Algoritma *Naive Bayes* dan *Decision Tree* untuk Prediksi Penyakit Diabetes Berbasis Web” dengan baik.

Penyusunan Tugas Akhir ini merupakan salah satu syarat untuk memperoleh gelar Sarjana Komputer pada Program Studi Informatika, Fakultas Sains, Teknologi dan Kesehatan, Universitas Sahid Surakarta. Dalam proses penyusunan Tugas Akhir ini, penulis menghadapi berbagai tantangan, baik dari segi pengumpulan data, pemahaman materi, maupun pengembangan sistem. Namun berkat bimbingan, dukungan, dan doa dari berbagai pihak, Tugas Akhir ini akhirnya dapat diselesaikan.

Pada kesempatan ini, penulis menyampaikan penghargaan dan terima kasih yang sebesar-besarnya kepada:

1. Ibu Dr. Sri Huning Anwariningsih, S.T., M.Kom., selaku Rektor Universitas Sahid Surakarta, atas segala kebijakan, fasilitas, dan dukungan yang telah diberikan kepada penulis selama menempuh pendidikan hingga terselesaikannya penyusunan Tugas Akhir ini.
2. Bapak apt. Ahwan, S.Farm., M.Sc., selaku Dekan Fakultas Sains Teknologi dan Kesehatan, yang telah memberikan izin dan kemudahan kepada penulis dalam melaksanakan penelitian serta penyusunan Tugas Akhir ini.
3. Ibu Diyah Ruswanti, S.Kom., M.Kom., selaku Ketua Program Studi Informatika, atas bimbingan dan dukungan yang diberikan kepada penulis.
4. Bapak Farid Fitriyadi, S.Kom., M.Kom., selaku Penguji I, Bapak Ir. Dahlan Susilo, M.Kom., selaku Penguji II, dan Ibu Dwi Retnoningsih, S.T., M.T., selaku Penguji III, yang dengan penuh kesabaran telah membimbing, mengarahkan, serta memberikan saran dan kritik kepada penulis selama proses penelitian hingga penyusunan Tugas Akhir ini dapat terselesaikan dengan baik.
5. Kedua orang tua tercinta, yang senantiasa memberikan doa, dukungan, kasih sayang, serta motivasi yang tidak pernah terputus. Terutama Papa, sosok terbaik bagi penulis, yang tanpa henti mendoakan, memprioritaskan pendidikan, serta mengedepankan kebahagiaan penulis

di setiap langkah. Dari ketulusan merekalah penulis memperoleh kekuatan untuk terus berjuang hingga Tugas Akhir ini dapat diselesaikan.

6. Seluruh keluarga tercinta yang senantiasa memberikan doa, dukungan moral, serta perhatian yang tulus, sehingga penulis dapat menjalani proses perkuliahan dan penyusunan Tugas Akhir ini dengan lebih kuat dan sabar.
7. Viki Ramadhan, yang selalu mendampingi, memberikan semangat, pengertian, serta setia menemani dan mendengarkan keluh kesah di setiap proses yang penulis jalani hingga Tugas Akhir ini dapat terselesaikan.
8. Nadya Saltsa, yang selalu menjadi tempat berbagi keluh kesah selama proses pengerjaan Tugas Akhir, serta dengan senang hati menemani penulis bermain dan sejenak mengalihkan penat ketika mulai merasa jenuh dan lelah.
9. Adinda Putri, yang senantiasa menemani penulis selama menjalani masa perkuliahan dan selalu hadir memberikan dukungan di setiap situasi.
10. Teman-teman saya yang lain, yaitu Eka, Hana, Selvi, Satrio, Coeg, Kai, Gilang, Wiam, dan Icu, yang selalu bersedia meluangkan waktunya untuk menemani penulis ketika merasa jenuh, serta menjadi tempat berbagi cerita dan melepas penat di sela-sela pengerjaan Tugas Akhir.
11. Seperdua Kopi Laweyan, yang selalu hadir dan menjadi tempat penulis mengerjakan Tugas Akhir, menemani setiap proses dengan suasana yang nyaman dan menyenangkan.
12. Semua pihak yang tidak dapat disebutkan satu per satu, yang telah membantu baik secara langsung maupun tidak langsung dalam penyusunan Tugas Akhir ini.

Penulis menyadari bahwa Tugas Akhir ini masih jauh dari kesempurnaan dan memiliki banyak kekurangan, baik dari segi penulisan maupun substansi. Oleh karena itu, penulis sangat mengharapkan kritik dan saran yang membangun demi perbaikan di masa mendatang. Semoga Tugas Akhir ini dapat memberikan manfaat bagi pembaca, khususnya bagi pengembangan ilmu pengetahuan di bidang Informatika.

Surakarta, 17 Juli 2026

Penulis

DAFTAR ISI

LEMBAR PERSETUJUAN	i
LEMBAR PENGESAHAN	ii
SURAT PERNYATAAN ORISINIL KARYA ILMIAH	iii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH	iv
KATA PENGANTAR	v
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
ABSTRAK	xi
ABSTRACT	xii
BAB I PENDAHULUAN	1
1.1 Latar Belakang Masalah	1
1.2 Rumusan Masalah	3
1.3 Tujuan Penelitian	4
1.4 Manfaat Penelitian	4
1.5 Batasan Masalah	5
1.6 Metode Penelitian	5
1.7 Sistematika Penulisan	7
BAB II LANDASAN TEORI	8
2.1 Tinjauan Pustaka	8
2.2 Kerangka Pemikiran.....	9
2.3 Landasan Teori	10
BAB III ANALISIS DAN PERANCANGAN SISTEM	16
3.1 Analisis Sistem.....	16
3.1.1 Analisis Sistem yang Berjalan Saat Ini	16
3.1.2 Analisis Sistem yang Diusulkan	17
3.2 Pengumpulan Data (Dataset)	18
3.3 Pembagian Data	20

3.4	Tahap <i>Preprocessing</i>	20
3.5	Perancangan Baseline model	21
3.6	Perancangan Optimasi	22
3.7	Perancangan Komparasi Algoritma	22
3.8	Perancangan sistem berbasis web	23
3.9	Use Case Sistem	25
3.10	ERD	26
3.11	Desain UI/UX	27
BAB IV IMPLEMENTASI DAN ANALISIS HASIL		32
4.1	Hasil Pembagian Data	32
4.2	Hasil <i>Preprocessing</i>	32
4.3	Hasil <i>Baseline Model</i>	35
4.4	Hasil <i>K-Fold Cross Validation</i>	41
4.5	Analisis Komparatif Performa Model	45
4.6	Implementasi Sistem Berbasis Web	48
4.7	Hasil Pengujian Sistem BlackBox	53
BAB V SIMPULAN DAN SARAN		55
5.1	Simpulan	55
5.2	Saran	56
DAFTAR PUSTAKA		57
Lampiran		59

DAFTAR GAMBAR

Gambar1. 1 Metode Penelitian.....	6
Gambar 2. 1 Kerangka Pemikiran penelitian.....	10
Gambar 3. 1 Analisis berjalan.....	16
Gambar 3. 2 Analisis disulkan	18
Gambar 3. 3 <i>Flowchart</i> Sistem	25
Gambar 3. 4 <i>Use Case</i>	26
Gambar 3. 5 <i>ERD</i>	27
Gambar 3. 6 <i>UI/UX</i> Login	28
Gambar 3. 7 <i>UI/UX</i> Dashboard.....	29
Gambar 3. 8 <i>UI/UX</i> Prediksi.....	29
Gambar 3. 9 <i>UI/UX</i> Hasil Prediksi	30
Gambar 3. 10 <i>UI/UX</i> Data Pasien.....	30
Gambar 3. 11 <i>UI/UX</i> Komparasi Algoritma	31
Gambar 4. 1 <i>Confusion Matrix</i> NB.....	38
Gambar 4. 2 <i>Confusion Matrix</i> DT	39
Gambar 4. 3 <i>Confusion Matrix</i> NB Tuning	43
Gambar 4. 4 <i>Confusion Matrix</i> DT Tuning.....	44
Gambar 4. 5 Grafik Perbandingan Performa	46
Gambar 4. 6 Implementasi Login	49
Gambar 4. 7 Implementasi Dashboard.....	49
Gambar 4. 8 Implementasi Prediksi.....	50
Gambar 4. 9 Implementasi Hasil Prediksi	50
Gambar 4. 10 Implementasi Prediksi Non-Diabetes.....	51
Gambar 4. 11 Implementasi Data Pasien	51
Gambar 4. 12 Implementasi Komparasi Algoritma	52

DAFTAR TABEL

Tabel 2. 1 Perbandingan Penelitian	8
Tabel 2. 2 <i>Confusion Matrix</i>	15
Tabel 3. 1 Kolom <i>Dataset</i>	19
Tabel 4. 1 Split Data.....	32
Tabel 4. 2 Imputasi <i>Missing Value</i>	33
Tabel 4. 3 <i>Preprocessing Data</i>	39
Tabel 4. 4 Standardisasi Data.....	34
Tabel 4. 5 Inisialisasi NB.....	35
Tabel 4. 6 Inisialisasi DT.....	35
Tabel 4. 7 Hasil Training Model.....	36
Tabel 4. 8 Karakteristik NB.....	36
Tabel 4. 9 Karakteristik DT.....	37
Tabel 4. 10 Prediksi Model.....	38
Tabel 4. 11 Evaluasi Baseline Model.....	40
Tabel 4. 12 <i>GridSearchCV</i> NB.....	41
Tabel 4. 13 <i>GridSearchCV</i> DT.....	42
Tabel 4. 14 Evaluasi Metrik <i>5-Fold Cross Validation</i>	43
Tabel 4. 15 Perbandingan performa.....	45
Tabel 4. 16 Pengujian BlackBox.....	54

ABSTRAK

Diabetes merupakan salah satu penyakit kronis dengan jumlah penderita yang terus meningkat, sementara sebagian kasus masih belum teridentifikasi secara klinis, sehingga diperlukan pendekatan berbasis teknologi untuk membantu proses *screening* awal dalam mengenali risiko diabetes. Penelitian ini bertujuan menerapkan algoritma *Naive Bayes* dan *Decision Tree* untuk mengklasifikasikan status diabetes berdasarkan lima fitur, yaitu *Age*, *Glucose*, *BloodPressure*, *Insulin*, dan *BMI*, serta mengembangkan sistem prediksi berbasis *web* yang menampilkan perbandingan kinerja kedua algoritma. Penelitian menggunakan *Pima Indians Diabetes Dataset* sebanyak 768 data yang dibagi menjadi data latih dan data uji dengan rasio 80:20 menggunakan metode *stratified*. Tahapan penelitian meliputi *preprocessing* data, penanganan *missing value*, standardisasi fitur, pembangunan model *baseline*, optimasi menggunakan *Stratified 5-Fold Cross Validation* dan *GridSearchCV*, serta evaluasi model melalui *confusion matrix* dengan metrik akurasi, sensitivitas, dan spesifisitas. Hasil penelitian menunjukkan bahwa *Decision Tree* setelah optimasi memberikan kinerja lebih baik dibandingkan *Naive Bayes*. Pada proses validasi silang, *Decision Tree* memperoleh Best CV Sensitivitas sebesar 80,43%, sedangkan *Naive Bayes* memperoleh 64,51%. Pada pengujian akhir menggunakan data testing, *Decision Tree* menghasilkan akurasi 85,06%, sensitivitas 83,33%, dan spesifisitas 86,00%, sedangkan *Naive Bayes* menghasilkan akurasi 72,73%, sensitivitas 64,81%, dan spesifisitas 77,00%. Berdasarkan hasil tersebut, *Decision Tree* dipilih sebagai model terbaik karena memperoleh performa pengujian yang lebih tinggi, menghasilkan jumlah kesalahan klasifikasi yang lebih rendah, serta memiliki aturan keputusan yang mudah dipahami. Sistem berbasis *web* yang dikembangkan berhasil mengintegrasikan model tersebut untuk membantu tenaga kesehatan melakukan *screening* awal risiko diabetes secara cepat dan terdokumentasi, sebagai alat bantu pendukung keputusan tanpa menggantikan proses diagnosis klinis.

Kata kunci: *diabetes*, *Naive Bayes*, *Decision Tree*, klasifikasi, sistem berbasis *web*

ABSTRACT

Diabetes is a chronic disease with a continuously rising number of cases, yet many remain clinically undiagnosed; consequently, a technology-based approach is needed to assist in the initial screening process for identifying diabetes risk. This study aims to apply Naive Bayes and Decision Tree algorithms to classify diabetes status based on five features: Age, Glucose, Blood Pressure, Insulin, and BMI and to develop a web-based prediction system that compares the performance of both algorithms. The study utilized the Pima Indians Diabetes Dataset, comprising 768 records split into training and testing sets at an 80:20 ratio using a stratified method. Research stages included data preprocessing, handling missing values, feature standardization, baseline model construction, optimization via Stratified 5-Fold Cross Validation and GridSearchCV, and model evaluation using a confusion matrix with accuracy, sensitivity, and specificity metrics. The results indicate that the optimized Decision Tree outperformed Naive Bayes. During cross-validation, the Decision Tree achieved a best CV sensitivity of 80.43%, whereas Naive Bayes achieved 64.51%. In final testing, the Decision Tree yielded an accuracy of 85.06%, sensitivity of 83.33%, and specificity of 86.00%, while Naive Bayes yielded an accuracy of 72.73%, sensitivity of 64.81%, and specificity of 77.00%. Based on these results, the Decision Tree was selected as the best model due to its superior testing performance, lower misclassification rate, and easily interpretable decision rules. The developed web-based system successfully integrates this model to assist healthcare professionals in conducting rapid, documented initial screenings for diabetes risk, serving as a decision-support tool rather than a replacement for the clinical diagnostic process.

Keywords: Diabetes, Naive Bayes, Decision Tree, Classification, Web Based System

