

**TUGAS AKHIR**

**PERBANDINGAN KLASIFIKASI PENYAKIT DIABETES  
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBORS  
(KNN) DAN NAÏVE BAYES: ANALISIS PERFORMA DAN  
KEAKURATAN**



**Oleh :**

**David Christian Putra**

**1462000251**

**PROGRAM STUDI INFORMATIKA**

**FAKULTAS TEKNIK**

**UNIVERSITAS 17 AGUSTUS 1945 SURABAYA**

**2024**

TUGAS AKHIR

PERBANDINGAN KLASIFIKASI PENYAKIT DIABETES  
MENGUNAKAN ALGORITMA K-NEAREST  
NEIGHBORS (KNN) DAN NAÏVE BAYES: ANALISIS  
PERFORMA DAN KEAKURATAN

Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana  
Komputer di Program Studi Informatika



Oleh :

David Christian Putra

1462000251

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA  
2024

TUGAS AKHIR

COMPARISON OF DIABETES CLASSIFICATION USING  
K-NEAREST NEIGHBORS (KNN) AND NAIVE BAYES  
ALGORITHM: PERFORMANCE AND ACCURACY  
ANALYSIS

Prepared as partial fulfilment of the requirement for the degree of Sarjana  
Komputer at Informatics Departement



Oleh :

David Christian Putra

1462000251

PROGRAM STUDI INFORMATIKA  
FAKULTAS TEKNIK  
UNIVERSITAS 17 AGUSTUS 1945 SURABAYA  
2024

**PROGRAM STUDI INFORMATIKA**  
**FAKULTAS TEKNIK**  
**UNIVERSITAS 17 AGUSTUS 1945 SURABAYA**

---

**LEMBAR PENGESAHAN TUGAS AKHIR**

**Nama** : David Christian Putra  
**NBI** : 1462000251  
**Prodi** : S-1 Informatika  
**Fakultas** : Teknik  
**Judul** : PERBANDINGAN KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (KNN) DAN NAÏVE BAYES: ANALISIS PERFORMA DAN KEAKURATAN

**Mengetahui / Menyetujui**

**Dosen Pembimbing 1**



**Supangat, M.Kom., Ph.D**

0724047106

**Dekan Fakultas Teknik Universitas**

17 Agustus 1945

Surabaya



**Dr. Ir. Sajiyo, M.Kes., IPU., ASEAN Eng.**

NPP. 20410.90.0197

**Ketua Program Studi Informatika**

Universitas 17 Agustus 1945

Surabaya



**Aidil Primasetya Armin, S.ST., M.T.**

NPP. 20460.16.0700

*Halaman ini sengaja dikosongkan*

# **PERNYATAAN KEASLIAN DAN PERSETUJUAN PUBLIKASI TUGAS AKHIR**

Saya yang bertanda tangan di bawah ini

Nama : David Christian Putra

NBI : 1462000251

Fakultas/Program Studi : Teknik/Informatika

Judul Tugas Akhir : Perbandingan Klasifikasi Penyakit Diabetes  
Menggunakan Algoritma K-Nearest Neighbors (Knn)  
Dan Naïve Bayes: Analisis Performa Dan Keakuratan

Menyatakan dengan sesungguhnya bahwa :

1. Tugas Akhir dengan judul diatas bukan merupakan tiruan atau duplikasi dari Tugas Akhir yang sudah dipublikasikan dan atau pernah dipakai untuk mendapatkan gelar Sarjana Teknik di lingkungan Universitas 17 Agustus 1945 Surabaya maupun di Peguruan Tinggi atau Instansi manapun, kecuali bagian yang sumber informasinya dicantumkan sebagaimana mestinya.
2. Tugas Akhir dengan judul diatas bukan merupakan plagiarisme, pencurian hasil karya milik orang lain, hasil kerja orang lain untuk kepentingan saya karena hubungan material maupun non-material, ataupun segala kemungkinan lain yang pada hakekatnya bukan merupakan karya tulis tugas akhir saya secara orisinal dan otentik.
3. Demi pengembangan ilmu pengetahuan, saya memberikan hak Tugas Akhir ini kepada Universitas 17 Agustus 1945 Surabaya untuk menyimpan, mengalihmedia/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan memublikasikan tugas akhir saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan sebagai pemilik Hak Cipta.
4. Pernyataan ini saya buat dengan kesadaran sendiri dan tidak atas tekanan ataupun paksaan dari pihak maupun demi menegakan integritas akademik di institusi ini dan bila kemudian hari diduga kuat ada ketidaksesuaian antara fakta dengan kenyataan ini, saya bersedia diproses oleh tim Fakultas yang dibentuk untuk melakukan verifikasi, dengan sanksi terberat berupa pembatalan kelulusan/kesarjanaan.

Surabaya, 18 Juni 2024

  
DAVID Christian Putra

1462000251

*Halaman ini sengaja dikosongkan*



Universitas

**17 AGUSTUS 1945**

SURABAYA

**BADAN PERPUSTAKAAN**

JL. SEMOLOWARU 45

SURABAYATLP. 031 593 1800

(EX 311)

EMAIL: PERPUS@UNTAG-

SBY.AC.ID.

## **PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH**

Sebagai sivitas akademik Universitas 17 Agustus 1945 Surabaya, Saya yang bertanda tangan di bawah ini:

Nama : David Christian Putra  
NIM : 1462000251  
Fakultas : Teknik  
Program Studi : Informatika  
Jenis Karya : Tugas Akhir/Skripsi

Saya dengan ini menyatakan persetujuan saya untuk memberikan **hak bebas royalti noneksklusif (Nonexclusive Royalty-Free Right)** atas karya saya yang berjudul “**PERBANDINGAN KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (KNN) DAN NAÏVE BAYES: ANALISIS PERFORMA DAN KEAKURATAN**” kepada Badan Perpustakaan Universitas 17 Agustus 1945 Surabaya. Saya menyadari bahwa dengan memberikan hak ini, saya mengizinkan Universitas untuk menyimpan, mengelola, dan mendistribusikan karya saya dalam rangka mendukung tujuan pendidikan dan penelitian. Hak bebas royalti noneksklusif ini memungkinkan Universitas untuk menggunakan karya saya tanpa harus membayar royalti atau mendapatkan izin eksklusif, sehingga dapat diakses dan dimanfaatkan oleh komunitas akademik dan publik yang lebih luas. Dengan demikian, saya berkomitmen untuk mendukung Universitas 17 Agustus 1945 Surabaya dalam upayanya meningkatkan sumber daya informasi dan memajukan pendidikan serta penelitian di institusi ini.

Dibuat di : Universitas 17 Agustus 1945 Surabaya

Pada Tanggal : 13 Juni 2024

Yang Menyatakan



*Halaman ini sengaja dikosongkan*

## KATA PENGANTAR

Puji syukur kepada Tuhan Yesus Kristus atas segala karunia dan anugerah sehingga penulis dapat menyelesaikan tugas akhir yang berjudul “PERBANDINGAN KLASIFIKASI PENYAKIT DIABETES MENGGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (KNN) DAN NAÏVE BAYES: ANALISIS PERFORMA DAN KEAKURATAN” sebagai salah satu syarat untuk memperoleh gelar sarjana Komputer di Program Studi Teknik Informatika Universitas 17 Agustus 1945 Surabaya. Penulis sadar bahwa tanpa bantuan dari Tuhan yesus dan orang tua serta doa dari teman-teman yang ikut berperan dalam membantu penulis untuk menyelesaikan dengan baik.

Karena itu penulis ingin menyampaikan terimakasih sebesar-besarnya kepada:

1. Keluarga tercinta, mama, papa, dan dua adek saya yang selalu mendoakan, bimbingan, dan dukungan moral selama penulis menempuh perkuliahan.
2. Bapak Supangat, M.Kom., Ph.D selaku dosen pembimbing Tugas Akhir yang telah banyak memberi banyak waktu untuk memberikan arahan dan dukungan dan juga terimakasih atas kesabaran, perhatian, petunjuk dan semangat dalam menyelesaikan Tugas Akhir ini.
3. Bapak Agus Hermanto, S.Kom., M.MT., ITIL, COBIT, SFC dan Fridy Mandita, S.Kom., M.Sc. selaku dosen penguji yang telah memberikan banyak saran dan kritik untuk kebaikan tugas akhir penulis.
4. Bapak Aidil Primasetya Armin, S.ST., M.T., selaku Ketua Program Studi Teknik Informatika, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya.
5. Dr. Ir. H. Sajiyo, M.Kes., IPM, selaku Dekan Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya.
6. Bapak Agil Ardi Rahmadi, S.Kom., M.A., selaku Koordinator Tugas Akhir Jurusan Teknik Informatika, Fakultas Teknik, Universitas 17 Agustus 1945 Surabaya.
7. Teman-Teman MOCCA saya yang telah menemani, memberi dukungan moral, membantu, dan juga mendoakan penulis sehingga dapat menyelesaikan tugas akhir.
8. Teman-Teman tenaga pendidik dari Fakultas Kedokteran Universitas Widya Mandala Surabaya yang telah memberi semangat dalam menyelesaikan tugas akhir.
9. dr. Alvin Julian yang telah membantu saya dalam mengerjakan program klasifikasi diabetes.

10. Teman-Teman Teknik Informatika Universitas 17 Agustus 1945 Surabaya Angkatan 2020 yang membantu saya dari awal perkuliahan sampai ada di titik ini.

Akhir kata penulis mengucapkan banyak terimakasih kepada semua pihak yang belum bisa penulis sebutkan karena telah membantu setiap perkuliahan dari awal sampai akhir. Semoga Tugas Akhir ini bisa bermanfaat bagi penulis maupun pembaca.

Surabaya, 18 Juni 2024

## ABSTRAK

Nama : David Christian Putra

Program Studi : Informatika

Judul : PERBANDINGAN KLASIFIKASI PENYAKIT DIABETES  
MENGUNAKAN ALGORITMA K-NEAREST  
NEIGHBORS (KNN) DAN NAÏVE BAYES: ANALISIS  
PERFORMA DAN KEAKURATAN

Diabetes adalah penyakit kronis yang mengganggu metabolisme tubuh dengan tingginya kadar glukosa darah, yang mengakibatkan gangguan respons terhadap insulin. Jumlah penderita diabetes diprediksi mencapai 700 juta orang pada tahun 2045. Selain orang berusia 35 ke atas, anak muda usia 18 tahun ke atas kini juga banyak yang terkena diabetes, terutama karena konsumsi gula yang tinggi. Diabetes dapat menyebabkan komplikasi serius seperti masalah ginjal, amputasi, gangguan penglihatan, kerusakan saraf, serta meningkatkan risiko serangan jantung dan stroke dua hingga tiga kali lipat pada orang dewasa. Deteksi dini melalui pemeriksaan gula darah sangat penting untuk mencegah keparahan penyakit ini. Faktor-faktor yang mempengaruhi diagnosis diabetes meliputi jenis kelamin, umur, tekanan darah, BMI, dan riwayat merokok, hba1c level, dan gulad darah. Data mining memungkinkan penggunaan data historis untuk menemukan pola-pola baru. Algoritma K-Nearest Neighbor (KNN) dan Naïve Bayes adalah dua algoritma data mining yang digunakan untuk klasifikasi. KNN memprediksi kelas data baru berdasarkan kedekatan dengan data pelatihan terdekat, sementara Naïve Bayes menghitung probabilitas berdasarkan teorema Bayes dan menganggap atribut saling bebas. Dari hasil penelitian, algoritma K-Nearest Neighbor menghasilkan akurasi 95.38%, presisi 95.77%, recall 99.33%, dan F1-Score sebesar 97.52. Sedangkan algoritma Naïve Bayes Menghasilkan akurasi 90.77%, 96.78%, 92.99%, dan F1-Score 94.85.

**Kata Kunci:** Data Mining, Diabetes, Klasifikasi, K-Nearest Neighbor, Naïve Bayes

*Halaman ini sengaja dikosongkan*

## ABSTRACT

Nama : David Christian Putra

Program Studi : Informatika

Judul : PERBANDINGAN KLASIFIKASI PENYAKIT DIABETES  
MENGUNAKAN ALGORITMA K-NEAREST NEIGHBORS (KNN) DAN  
NAÏVE BAYES: ANALISIS PERFORMA DAN KEAKURATAN

Diabetes is a chronic disease that disrupts the body's metabolism by causing high blood glucose levels, resulting in impaired response to insulin. The number of diabetes sufferers is predicted to reach 700 million people by 2045. Apart from people aged 35 and over, many young people aged 18 years and over are now also suffering from diabetes, especially due to high sugar consumption. Diabetes can cause serious complications such as kidney problems, amputations, vision problems, nerve damage, and increases the risk of heart attack and stroke two to three times in adults. Early detection through blood sugar checks is very important to prevent the severity of this disease. Factors that influence the diagnosis of diabetes include gender, age, blood pressure, BMI, and smoking history, hba1c level, and blood sugar. Data mining allows the use of historical data to discover new patterns. The K-Nearest Neighbor (KNN) and Naïve Bayes algorithms are two data mining algorithms used for classification. KNN predicts new data classes based on proximity to nearby training data, while Naïve Bayes calculates probabilities based on Bayes' theorem and assumes attributes are independent. From the research results, the K-Nearest Neighbor algorithm produces 95.38% accuracy, 95.77% precision, 99.33% recall, and an F1-Score of 97.52. Meanwhile, the Naïve Bayes algorithm produces accuracy of 90.77%, 96.78%, 92.99%, and F1-Score of 94.85.

**Keywords:** Data Mining, Diabetes, Classification, K-Nearest Neighbor, Naïve Bayes

*Halaman ini sengaja dikosongkan*

# DAFTAR ISI

LEMBAR PENGESAHAN TUGAS AKHIR .....	i
PERNYATAAN KEASLIAN DAN PERSETUJUAN PUBLIKASI TUGAS AKHIR .....	iii
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH .....	v
KATA PENGANTAR.....	vii
ABSTRAK .....	ix
ABSTRACT .....	xi
DAFTAR GAMBAR.....	xv
DAFTAR TABEL .....	xvii
DAFTAR PERSAMAAN .....	xix
BAB 1 PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Perumusan Masalah .....	3
1.3 Batasan Masalah .....	3
1.4 Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
BAB 2 KAJIAN PUSTAKA DAN LANDASAN TEORI.....	5
2.1 Data Mining .....	5
2.2 Klasifikasi .....	6
2.3 Naïve Bayes .....	6
2.4 KNN.....	8
2.5 Diabetes .....	10
2.6 Confusion Matrix .....	10
2.7 Penelitian Sebelumnya.....	10
BAB 3 METODE PENELITIAN .....	37
3.1 Bahan dan Perangkat Penelitian .....	37
3.2 Obyek Penelitian.....	37

3.3	Tahapan Penelitian .....	37
3.4	Skenario Penelitian.....	39
BAB 4 HASIL DAN PEMBAHASAN .....		41
4.1	Hasil dan Pengumpulan data .....	41
4.2	Hasil Pengelolaan Data dengan Algoritma K-Nearest Neighbor 49	
4.3	Hasil Pengelolaan Data dengan Algoritma Naïve Bayes .....	51
4.4	Tahap Melihat Algoritma .....	53
4.5	Tahap Perancangan Website .....	53
4.5.1	Use case.....	53
4.5.2	Activity Diagram.....	54
4.5.3	Entity Relationship Diagram (ERD).....	59
4.6	Tahap Implementasi Web.....	60
4.6.1	Models.....	60
4.6.2	Form Login .....	62
4.6.3	Index.....	65
4.6.4	Pasien .....	69
4.6.5	Klasifikasi .....	72
4.6.6	Report.....	79
4.7	Pengujian Perangkat Lunak.....	79
4.7.1	Form Login .....	79
4.7.2	Form User .....	80
4.7.3	Form Pasien.....	81
4.7.4	Form Klasifikasi.....	82
BAB 5 KESIMPULAN.....		85
5.1	Kesimpulan.....	85
5.2	Saran.....	85
DAFTAR PUSTAKA .....		87

## DAFTAR GAMBAR

<b>Gambar 3. 1</b>	Tahapan penelitian .....	38
<b>Gambar 4. 1</b>	Visualisasi data label.....	47
<b>Gambar 4. 2</b>	Visualisasi data gender.....	48
<b>Gambar 4. 3</b>	Diagram Use case .....	53
<b>Gambar 4. 4</b>	Diagram Activity Login.....	54
<b>Gambar 4. 5</b>	Diagram Activity user.....	55
<b>Gambar 4. 6</b>	Diagram Activity Pasien .....	56
<b>Gambar 4. 7</b>	Diagram Activitiy Klasifikasi .....	57
<b>Gambar 4. 8</b>	Activity Diagram Report.....	58
<b>Gambar 4. 9</b>	Entity Relationship Diagram.....	59
<b>Gambar 4. 10</b>	Form Login .....	62
<b>Gambar 4. 11</b>	Halaman Index .....	65
<b>Gambar 4. 12</b>	Halaman data Pasien.....	69
<b>Gambar 4. 13</b>	Halaman data klasifikasi.....	72
<b>Gambar 4. 14</b>	Form 1 tambah data klasifikasi.....	73
<b>Gambar 4. 15</b>	Form 2 tambah data klasifikasi.....	73
<b>Gambar 4. 16</b>	Gambar halaman report.....	79

*Halaman ini sengaja dikosongkan*

## DAFTAR TABEL

<b>Tabel 2. 1</b> Tabel Metriks Jurnal Penelitian terkait .....	17
<b>Tabel 3. 1</b> Skenario Penelitian .....	40
<b>Tabel 4. 1</b> beberapa Dataset .....	41
<b>Tabel 4. 2</b> Kode Sumber Import Library.....	45
<b>Tabel 4. 3</b> Tabel Import Dataset.....	46
<b>Tabel 4. 4</b> Tabel Kode Sumber Visualisasi Data Label DIabetes.....	46
<b>Tabel 4. 5</b> Kode Sumber Visualisasi Data Atribut Gender.....	47
<b>Tabel 4. 6</b> Kode Sumber Label Encoder .....	49
<b>Tabel 4. 7</b> Tabel Perubahan Data Kolom Gender .....	49
<b>Tabel 4. 8</b> Tabel Perubaha Data Kolom smoking_history .....	49
<b>Tabel 4. 9</b> Kode Sumber Pembagian Data .....	49
<b>Tabel 4. 10</b> Kode Pemanggilan Algoritma K-Nearest Neighbors.....	50
<b>Tabel 4. 11</b> Kode sumber prediksi algoritma .....	50
<b>Tabel 4. 12</b> Tabel Performa Algoritma K-Nearest Neighbors.....	50
<b>Tabel 4. 13</b> Kode Sumber Congusion Matrix .....	51
<b>Tabel 4. 14</b> Tabel Confusion Matrix K-Nearest Neighbors .....	51
<b>Tabel 4. 15</b> Tabel Kode Sumber Pembagian Data .....	51
<b>Tabel 4. 16</b> Kode Sumber Pemanggilan Algoritma Naïve Bayes .....	51
<b>Tabel 4. 17</b> Kode Sumber Prediksi Algoritma .....	52
<b>Tabel 4. 18</b> Tabel Performa Algoritma Naïve Bayes .....	52
<b>Tabel 4. 19</b> Tabel Kode Sumber Confusion Matrix .....	52
<b>Tabel 4. 20</b> Tabel Confusion Matrix Naïve Bayes .....	52
<b>Tabel 4. 21</b> Kode Sumber Model .....	60
<b>Tabel 4. 22</b> Kode Sumber Form Login.html.....	63
<b>Tabel 4. 23</b> Kode Sumber Fungsi Login Pada Views .....	64
<b>Tabel 4. 24</b> Kode Sumber Fungsi Index pada Views .....	65
<b>Tabel 4. 25</b> Kode Sumber index.html .....	67
<b>Tabel 4. 26</b> Kode Sumber fungsi index pasien .....	69
<b>Tabel 4. 27</b> Kode Sumber Pasien.html.....	70

<b>Tabel 4. 28</b> Kode Sumber Fungsi hitung_umur().....	73
<b>Tabel 4. 29</b> Kode Sumber Fungsi load_data .....	74
<b>Tabel 4. 30</b> Kode Sumber knn_model() .....	75
<b>Tabel 4. 31</b> Kode Sumber nb_model() .....	75
<b>Tabel 4. 32</b> Kode Sumber post_tambah_data.....	75
<b>Tabel 4. 33</b> Tabel Pengujian Form Login.....	79
<b>Tabel 4. 34</b> Tabel Pengujian Form User.....	80
<b>Tabel 4. 35</b> Tabel Pengujian Form Pasien.....	81
<b>Tabel 4. 36</b> Tabel Pengujian Form Klasifikasi .....	82

## DAFTAR PERSAMAAN

Persamaan 2.1.....	7
Persamaan 2.2.....	8
Persamaan 2.3.....	8
Persamaan 2.4.....	9
Persamaan 2.5. ....	10

*Halaman ini sengaja dikosongkan*