

SKRIPSI

**PERBANDINGAN METODE K-NEAREST NEIGHBOR (KNN)
DAN NAÏVE BAYES: KLASIFIKASI RESIKO DIABETES
(STUDI KASUS POSBINDU DESA BULUPITU KABUPATEN
MALANG)**



Oleh :

Rizki Alifia Safitri

NIM : 20083000082

**PROGRAM S1 SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA
MALANG
2024**

**PERBANDINGAN METODE K-NEAREST NEIGHBOR (KNN)
DAN NAÏVE BAYES: KLASIFIKASI RESIKO DIABETES
(STUDI KASUS POSBINDU DESA BULUPITU KABUPATEN
MALANG)**



SKRIPSI

Diajukan Sebagai Salah Satu Persyaratan
Mendapatkan Gelar Sarjana Sistem Informasi (S. Kom)

Oleh :

Rizki Alifia Safitri

NIM : 20083000082

**PROGRAM S1 SISTEM INFORMASI
FAKULTAS TEKNOLOGI INFORMASI
UNIVERSITAS MERDEKA
MALANG**

2024

MOTTO

Setiap langkah dalam hidup adalah kesempatan untuk belajar dan tumbuh. Aku percaya bahwa setiap tantangan yang kuhadapi adalah batu loncatan menuju kesuksesan yang lebih besar. Dengan tekad dan kerja keras, tidak ada impian yang terlalu tinggi untuk dicapai. Keberanian untuk bermimpi besar dan ketekunan untuk terus bergerak maju, itulah yang menjadi pendorong dalam setiap perjuanganku.


LEMBAR PERSETUJUAN SKRIPSI

Nama : Rizki Alifia Safitri
NIM : 20083000082
Program Studi : Sistem Informasi
Bidang Kajian : Data mining
Judul Tugas Akhir : Perbandingan Metode K-Nearest Neighbor (KNN) Dan Naïve Bayes: Klasifikasi Resiko Diabetes (Studi Kasus Posbindu Desa Bulupitu Kabupaten Malang)


Malang, 08 Mei 2024

DISETUJUI DAN DITERIMA

Ketua Program Studi S1 Sistem
Informasi


Luthfi Indana, S.Pd., M.Pd
0711059203

Dosen Pembimbing


Rahmatina Hidayati, S. Kom, M. T.
0720028902

Dekan
Fakultas Teknologi Informasi

Dr. Mardiana Andarwati, M.Si
0716037601

LEMBAR PENGESAHAN SKRIPSI

Dipersiapkan dan disusun oleh

Nama : Rizki Alifia Safitri
NIM : 20083000082
Judul : Perbandingan Metode K-Nearest Neighbor (KNN) Dan Naïve Bayes: Klasifikasi Resiko Diabetes (Studi Kasus Posbindu Desa Bulupitu Kabupaten Malang)

Telah dipertahankan di depan Dosen penguji pada

Hari : Kamis
Tanggal : 25 Juli 2024
Tempat : Fakultas Teknologi Informasi
Universitas Merdeka Malang

Susunan Dewan Penguji

Ketua Penguji



Firnanda Al-Islama Achyunda Putra,
S.Kom., M.Kom.
0706129401

Sekretaris Penguji



Rahmatina Hidayati, S. Kom, M. T.
0720028902

Anggota Penguji



Himawan Pramaditya, S.Kom., M.Kom.
0716098405

Skrripsi ini Telah Diterima Sebagai Salah Satu Persyaratan untuk Memperoleh
Gelar Sarjana / Ahli Madya Komputer

Malang, 20 Agustus 2024
Dekan Fakultas Teknologi Informasi
Universitas Merdeka Malang

Dr. Mardiana Andarwati, M.Si
0716037601

LEMBAR PERNYATAAN ORISINALITAS SKRIPSI

Saya yang bertandatangan di bawah ini:

Nama : RIZKI ALIFIA SAFITRI
NPM : 20083000082
Program Studi : S1 Sistem Informasi
Judul : Perbandingan Metode K-Nearest Neighbor (KNN) Dan Naïve Bayes: Klasifikasi Resiko Diabetes (Studi Kasus Posbindu Desa Bulupitu Kabupaten Malang)
Fakultas : Teknologi Informasi
Perguruan Tinggi : Universitas Merdeka Malang

Menyatakan dengan sesungguhnya bahwa skripsi saya yang berjudul:

“PERBANDINGAN METODE K-NEAREST NEIGHBOR (KNN) DAN NAÏVE BAYES: KLASIFIKASI RESIKO DIABETES (STUDI KASUS POSBINDU DESA BULUPITU KABUPATEN MALANG)”

adalah benar-benar hasil karya saya dan bukan karya orang lain (plagiasi/jiplakan) serta tidak didasarkan pada data palsu, baik sebagian maupun keseluruhan kecuali dalam bentuk kutipan daftar pustaka.

Demikian pernyataan ini saya buat dengan sebenar-benarnya dan apabila ternyata di dalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiat, maka saya bersedia mendapatkan sanksi akademik berupa pencabutan gelar akademik sesuai dengan peraturan perundang-undangan yang berlaku.

Malang, 22 Agustus 2024

Yang membuat pernyataan,



Rizki Alifia Safitri

ABSTRAK

Diabetes melitus termasuk masalah kesehatan dengan pertumbuhan paling cepat di abad ke-21. Salah satu penyebabnya karena kurangnya kesadaran Masyarakat untuk *check-up* kesehatan secara berkala, sedangkan pola hidup yang dijalani terbilang tidak sehat. Pemeriksaan hemoglobin A1c (*HbA1c*) sangat dianjurkan untuk mendeteksi diabetes. Tapi layanan tersebut belum ada di Posbindu Desa Bulupitu. Oleh karena itu perlu pendekatan lain untuk mendeteksi dini resiko Masyarakat terkena diabetes yakni dengan data mining. Metode data mining yang digunakan dalam penelitian ini adalah metode klasifikasi *Naïve Bayes* dan *KNN*. Variabel untuk menentukan resiko diabetes antara lain: jenis kelamin, usia, keturunan diabetes, sering kencing, *Body Mass Index (BMI)*, kadar gula, dan output resiko diabetes. Pembagian dataset *testing* dan *training* menggunakan *cross-validation* dan *rasio* (60:40, 70:30, 80:20, dan 90:10). Akurasi terbaik metode *Naïve Bayes* diperoleh diperoleh dengan pembagian dataset *cross-validation* k-fold=2 sebesar 96,1%. Sedangkan pada metode *KNN* hasil terbaik diperoleh dari rasio dataset 80:30. *Manhattan distance* menjadi perhitungan jarak terbaik dalam penelitian ini dibandingkan dengan *Euclidean distance* dan *Chebyshev distance*.

Kata kunci: Diabetes, *KNN*, *Naïve Bayes*

ABSTRACT

Diabetes mellitus is among the fastest-growing health problems in the 21st century. One of the causes is the lack of public awareness for regular health check-ups, while the lifestyle they lead is generally unhealthy. The hemoglobin A1c (HbA1c) test is highly recommended for detecting diabetes, but this service is not yet available at the Posbindu in Bulupitu Village. Therefore, an alternative approach is needed to detect the risk of diabetes early, using data mining. The data mining methods used in this study are the Naïve Bayes and k-Nearest Neighbor (kNN) classification methods. The variables used to determine diabetes risk include gender, age, family history of diabetes, frequent urination, Body Mass Index (BMI), blood sugar level, and the output of diabetes risk. The dataset was split for testing and training using cross-validation and various ratios (60:40, 70:30, 80:20, and 90:10). The best accuracy for the Naïve Bayes method was obtained with a k-fold cross-validation split of 2, yielding 96.1%. Meanwhile, the best results for the kNN method were obtained with an 80:20 dataset ratio, with the Manhattan distance being the most effective distance calculation compared to Euclidean distance and Chebyshev distance.

Keywords: *Diabetes, KNN, Naïve Bayes*

LEMBAR PERSEMBAHAN

Skripsi ini saya persembahkan kepada:

1. Allah SWT, Tuhan Yang Maha Esa, atas rahmat, hidayah, dan karunia-Nya yang tak terhingga, sehingga saya dapat menyelesaikan skripsi ini dengan baik.
2. Orang tua tercinta, Ayah dan Ibu, yang selalu memberikan doa, dukungan, kasih sayang, dan motivasi tanpa henti. Terima kasih atas segala pengorbanan dan cinta yang tiada batas.
3. Saudara-saudara dan keluarga besar, yang selalu memberikan dukungan moral dan semangat dalam setiap langkah perjalanan saya.
4. Dosen pembimbing, yang dengan sabar dan penuh dedikasi membimbing dan memberikan arahan dalam penyusunan skripsi ini. Terima kasih atas waktu, ilmu, dan nasihat yang telah diberikan.
5. Sahabat-sahabat dan teman-teman seperjuangan, yang selalu memberikan dukungan, semangat, dan keceriaan. Terimakasih
6. Pasanganku yang selalu ada disampingku memberikan dukungan dan pengertian. Terima kasih atas kesabaran, semangat, dan dorongan yang tiada henti, yang telah memberiku kekuatan dalam menyelesaikan skripsi ini.
7. Almamater tercinta, yang telah memberikan wadah dan kesempatan untuk belajar, berkembang, dan meraih mimpi.

Semoga semua kebaikan yang diberikan kepada saya mendapatkan balasan yang berlipat dari Tuhan Yang Maha Esa. Skripsi ini saya persembahkan dengan penuh rasa hormat dan terima kasih.

UCAPAN TERIMAKASIH

Saya mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah mendukung dan membantu dalam penyelesaian skripsi ini. Terima kasih kepada dosen pembimbing yang telah memberikan bimbingan, saran, dan motivasi yang sangat berharga selama proses penelitian ini. Saya juga mengapresiasi keluarga, pasangan dan teman-teman yang selalu memberikan dorongan dan dukungan moral. Tak lupa, saya menyampaikan terima kasih kepada seluruh responden yang telah berpartisipasi dalam penelitian ini. Semoga hasil dari penelitian ini dapat memberikan kontribusi positif dan bermanfaat bagi perkembangan ilmu pengetahuan.

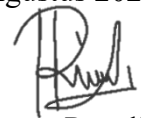
KATA PENGANTAR

Segala puji dan syukur kepada Tuhan Yang Maha Esa atas limpahan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan tugas akhir dengan judul: "**PERBANDINGAN METODE K-NEAREST NEIGHBOR (KNN) DAN NAÏVE BAYES: KLASIFIKASI RESIKO DIABETES (STUDI KASUS POSBINDU DESA BULUPITU KABUPATEN MALANG)** ." Pada kesempatan ini, penulis ingin mengucapkan terima kasih yang sebesar-besarnya kepada semua pihak yang telah membantu dalam penyelesaian tugas akhir ini. Adapun ucapan terima kasih tersebut disampaikan kepada:

1. Ibu Dr. Mardiana Andarwati, M.Si, Dekan Fakultas Teknologi Informasi Universitas Merdeka Malang.
2. Ibu Luthfi Indana, S.Pd., M.Pd, Ketua Program S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
3. Ibu Rahmatina Hidayati, S.Kom., M.T., dosen pembimbing yang telah memberikan bimbingan dan motivasi dalam penyelesaian tugas akhir ini.
4. Para Dosen Program S1 Sistem Informasi Fakultas Teknologi Informasi Universitas Merdeka Malang.
5. Kedua orang tua dan seluruh keluarga yang senantiasa memberikan dukungan moral dan spiritual.
6. Ibu Dr. Sunaria dan seluruh staf Posbindu Desa Bulupitu yang telah membantu selama penelitian.
7. Ayatulloh Alkursi, yang selalu ada di sampingku, memberikan dukungan, cinta, dan pengertian yang luar biasa. Terima kasih atas kesabaran dan semangat yang terus diberikan.
8. Semua pihak yang telah memberikan bantuan, baik secara langsung maupun tidak langsung, yang tidak dapat disebutkan satu per satu.

Dengan segala kerendahan hati, penulis menyadari bahwa penelitian ini memiliki banyak kekurangan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi kesempurnaan penelitian ini. Semoga karya sederhana ini dapat memberikan manfaat bagi semua pihak.

Malang, 22 Agustus 2024



Penulis

DAFTAR ISI

JUDUL	i
MOTTO.....	ii
LEMBAR PERSETUJUAN SKRIPSI.....	iii
LEMBAR PENGESAHAN SKRIPSI	iv
LEMBAR PERNYATAAN ORISINALITAS SKRIPSI	v
ABSTRAK	vi
ABSTRACT.....	vii
LEMBAR PERSEMBAHAN	viii
UCAPAN TERIMAKASIH	ix
KATA PENGANTAR.....	x
DAFTAR ISI	xi
DAFTAR TABEL.....	xiii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
DAFTAR RIWAYAT HIDUP.....	xvi
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang	1
1.2. Rumusan Masalah	2
1.3. Tujuan.....	2
1.4. Manfaat Penelitian.....	2
1.5. Batasan Penelitian	3
BAB II KAJIAN PUSTAKA	4
2.1 Landasan Teori	4
2.2 Penelitian Terdahulu	5
BAB III METODE PENELITIAN.....	8
3.1 Desain Penelitian	8
3.2. Lokasi Penelitian	10
3.3 Populasi dan <i>Sample</i>	10
3.4 Teknik Pengumpulan Data.....	11
3.5 Teknik Analisis Data.....	12
BAB IV HASIL DAN PEMBAHASAN.....	13
4.1 Hasil Pengembangan Produk.....	13

4.2 Uji Coba Produk.....	14
Uji coba produk dibagi menjadi beberapa bagian sebagai berikut:.....	14
4.3 Pembahasan	15
BAB V PENUTUP.....	18
5.1 Kesimpulan.....	18
5.2 Saran.....	18
DAFTAR PUSTAKA	20
LAMPIRAN.....	22

DAFTAR TABEL

Tabel 2.1 Penelitian Terdahulu	5
Tabel 3.1 <i>Dataset</i>	8
Tabel 3.2 <i>Confusion Matrix</i>	9
Tabel 3.3 Variabel Penelitian	11
Tabel 4.1 <i>k-fold Cross Validation</i> metode <i>Naïve Bayes</i>	13
Tabel 4.2 <i>k-fold Cross Validation</i> metode <i>K-Nearest Neighbour (KNN)</i>	13
Tabel 4.3 Percobaan <i>Naïve Bayes</i> berdasarkan perbandingan rasio datanya (<i>training:testing</i>)	14
Tabel 4.4 <i>K-Nearest Neighbour (KNN)</i> berdasarkan <i>Manhattan</i>	14
Tabel 4.5 <i>K-Nearest Neighbour (KNN)</i> berdasarkan <i>Chebyshev</i>	15
Tabel 4.6 Hasil Klasifikasi <i>Naive Bayes</i>	16
Tabel 4.7 Hasil Klasifikasi Tertinggi Masing-Masing <i>Distance</i>	16
Tabel 4.8 <i>Confusion Matrix KNN</i>	16

DAFTAR GAMBAR

Gambar 3.1 Desain Penelitian	8
Gambar 4.1 Evaluasi <i>Naïve Bayes</i> dengan <i>Cross-Validation</i>	16
Gambar 4.2 <i>Scatter Plot</i> antara <i>BMI</i> dan kadar gula darah pasien.....	17

DAFTAR LAMPIRAN

Lampiran 1	Daftar Pasien Posbindu Januari – Februari 2024	22
Lampiran 2	<i>Screenshot Aplikasi Orange</i> Hasil Prediksi Penelitian	26