

# Application of the K-Nearest Neighbor Method for Hypertension Disease Classification

Fadillah Fauziah Diqti<sup>1\*</sup>, Al Khaidar<sup>2</sup>, Muhammad Fikry<sup>3</sup>, Asrianda<sup>4</sup>

<sup>1,2,3,4</sup> Program Studi Magister Teknologi Informasi, Fakultas Teknik, Universitas Malikussaleh, Lhokseumawe, 24355, Indonesia

## Informasi Artikel

Diterima : 21 Desember 2025  
Revisi : 27 Desember 2025  
Publikasi : 31 Desember 2025

## Kata Kunci:

Hipertensi  
Klasifikasi  
K- Nearest Neighbor

## ABSTRAK

Hipertensi merupakan salah satu penyakit tidak menular dengan prevalensi tinggi dan sering disebut sebagai silent killer karena sering tidak menunjukkan gejala. Penelitian ini bertujuan untuk mengklasifikasikan penyakit hipertensi menggunakan metode *K-Nearest Neighbor (KNN)*. Data yang digunakan berjumlah 478 data pasien RSUD H. Sahudin dengan delapan atribut, yaitu usia, tekanan darah sistolik, tekanan darah diastolik, asam urat, kadar glukosa, kolesterol, berat badan, dan tinggi badan. Data dibagi menjadi 70% data latih dan 30% data uji. Hasil pengujian menunjukkan bahwa metode KNN dengan nilai  $K = 5$  menghasilkan tingkat akurasi sebesar 81,25%. Hasil penelitian ini menunjukkan bahwa algoritma KNN efektif digunakan dalam proses klasifikasi penyakit hipertensi dan dapat membantu pengambilan keputusan di bidang kesehatan.

## ABSTRACT

*Hypertension is a non-communicable disease with a high prevalence and is often referred to as a silent killer due to the absence of early symptoms. This study aims to classify hypertension using the K-Nearest Neighbor (KNN) method. The dataset consists of 478 patient records from RSUD H. Sahudin with eight attributes, including age, systolic blood pressure, diastolic blood pressure, uric acid level, glucose level, cholesterol, body weight, and height. The data were divided into 70% training data and 30% testing data. Experimental results show that KNN with  $K = 5$  achieved an accuracy of 81.25%. These results indicate that the KNN algorithm is effective for hypertension classification and can support decision-making in the healthcare sector.*

This is an open-access article under the [CC BY-SA](#) license



## \*Penulis Koresponden

Email: fadillah.257110201018@mhs.unimal.ac.id

Cara sitasi IEEE::

F. F. Diqti, A. Khaidar, M. Fikry, A. Asrianda, "Application of the K-Nearest Neighbor Method for Hypertension Disease Classification," *Journal of Artificial Intelligence and Software Engineering (JAISE)*, vol. 5, no. 4, p. 1401-1412, Bulan tahun. 10.30811/jaise.v5i4.8517

## 1. PENDAHULUAN

Perkembangan teknologi informasi sangat penting dalam proses kehidupan manusia. Seiring dengan pertumbuhan teknologi informasi, membuka peluang terhadap pertumbuhan data yang relatif banyak. Sebelum semua data memiliki nilai dan divalidasi, diperlukan cara untuk memproses, menyimpan, dan mengelola semua

data tersebut. Salah satu cara untuk memfasilitasi hal tersebut dengan adanya teknologi basis data yang dikenal dengan data mining.

Data mining adalah proses mengekstrak data menjadi sebuah informasi yang memungkinkan pengguna untuk memproses data dalam jumlah yang besar dan cepat. Setiap bagian data dalam data mining terdiri dari kelas tertentu sehingga dapat ditentukan kelas variabel data yang dimilikinya [1],[2],[3]. Data mining juga memberikan hasil yang terbaik. Data mining merupakan proses pengolahan data yang menggunakan teknik statistik, matematika, kecerdasan buatan dan machine learning untuk mengekstrak dan mengidentifikasi informasi yang berguna dan pengetahuan yang terkait dari basis data besar [4],[5]. Terdapat beberapa teknik yang dapat kita gunakan dalam data mining untuk menemukan sebuah informasi baru, salah satu teknik dari data mining adalah klasifikasi.

Klasifikasi adalah metode pengelompokan data secara sistematis menurut aturan dan kaidah yang telah ditetapkan sebelumnya. Klasifikasi juga dapat diartikan sebagai pengelompokan data atau objek baru yang kelasnya belum diketahui atau kategorinya [6]. Klasifikasi adalah salah satu metode statistik untuk pengelompokan atau pengkategorian data yang tersusun secara sistematis. Klasifikasi merupakan suatu tahap dalam menemukan model atau fungsi yang mendeskripsikan atau membedakan suatu konsep atau kelas data, tujuannya adalah untuk memperkirakan kelas dari suatu objek yang labelnya tidak diketahui [7],[8].

Ada beberapa algoritma yang biasa digunakan untuk klasifikasi data, antara lain *K-Nearest Neighbor*, *Decision Tree*, *Neighbor Weighted K-Nearest Neighbor* (NWKNN) dan algoritma Genetika. Banyak peneliti menerapkan algoritma *K-Nearest Neighbor* dalam pengklasifikasian mengenai klasifikasi penyakit kanker payudara menggunakan metode *K-Nearest Neighbor*. Pada penelitian ini penghitungan jarak kemiripan menggunakan jarak minkowski, Jarak minkowski adalah jarak di dalam ruang vektor yang telah ditentukan yang bisa dianggap sebagai generalisasi dari kedua jarak Euclidean dan jarak Manhattan. Dengan metode KNN diperoleh hasil paling akurat sebesar 93% [4] [9]. Kemudian juga mengenai penelitian yang berjudul Penerapan Metode *Decesion Tree* dan Algoritma Genetika untuk Klasifikasi Risiko Hipertensi. Pada penelitian ini menghasilkan nilai akurasi pengujian algoritma DT-GA rata-rata akurasi tertinggi 84% [5][10]. Kemudian pada penelitian yang berjudul klasifikasi risiko hipertensi menggunakan metode *Neighbor Weighted K-Nearest Neighbor* (NWKNN). Hasil dari penelitian ini menunjukkan saat data latih yang digunakan sebanyak 100 data dan data uji sebanyak 25 data, nilai  $K=10$ , dan nilai  $E=4$  dengan tingkat akurasi mencapai 88% [6] [11]. Tiga perbandingan tersebut menentukan hasil keputusan dari metode *K-Nearest Neighbor* lebih akurat dibandingkan dengan hasil keputusan metode *Decision Tree*, Algoritma Genetika dan *Neighbor Weighted K-Nearest Neighbor* (NWKNN), sehingga dari itu peneliti mengambil metode algoritma *K-Nearest Neighbor*.

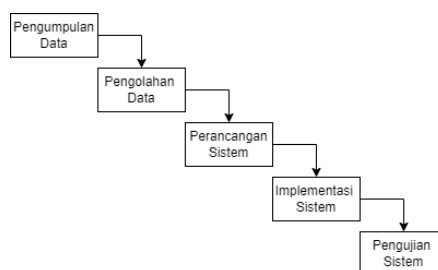
Metode yang digunakan dalam klasifikasi data adalah algoritma *K-Nearest Neighbor*. Algoritma *K-Nearest Neighbor* (KNN) adalah algoritma yang sangat sederhana dengan cara mengelompokkan data baru dengan  $K$  tetangga terdekat [12],[13].

Hipertensi adalah peningkatan tekanan darah sistolik lebih besar atau sama dengan 140 mmHg, dan peningkatan tekanan diastolik lebih besar atau sama dengan 90 mmHg [14],[15]. Hipertensi juga sering disebut Silent Killer karena tidak semua penderita hipertensi menyadari penyakit yang dideritanya. Hipertensi adalah penyebab utama terjadinya gagal jantung, stroke, dan gagal ginjal. Banyak faktor yang dapat mempengaruhi hipertensi yaitu faktor yang dapat dikendalikan atau tidak dikendalikan. Berdasarkan data nasional hasil Riskesdas 2018 menunjukkan bahwa prevalensi penduduk dengan tekanan darah tinggi sebesar 34,1%. Prevalensi hipertensi di provinsi Aceh 26,45% dan prevalensi di Aceh Tenggara sebesar 24,75%.

## 2. METODE

### 2.1 Langkah Penelitian

Penelitian klasifikasi penyakit hipertensi ini menggunakan algoritma *K Nearest Neighbor* yang dilakukan di RSUD H.Sahudin Kutacane. Penelitian ini dilaksanakan pada bulan Maret 2022 sampai Agustus 2022. Penelitian ini melibatkan beberapa tahapan metodologi pengumpulan data, pengolahan data, perancangan sistem, implementasi sistem, pengujian sistem. Tahapan metode penelitian dapat dilihat pada Gambar berikut.



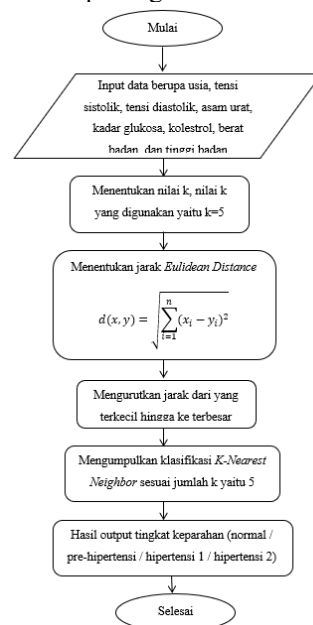
Gambar 1. Tahapan Penelitian

Pada tahap ini terjadi proses pengumpulan data yang diperlukan untuk merancang dan membangun sistem. Adapun jenis data yang digunakan berupa data skunder. Data skunder merupakan data yang berupa data pasien penyakit hipertensi yang didapatkan dari RSUD H. Sahudin Kutacane. Terdapat 8 atribut yaitu jenis kelamin, usia, pengukuran tekanan darah, asam urat, glukosa, kolesterol, berat badan dan tinggi badan. Pengumpulan data pada penelitian ini juga dilakukan dengan observasi yaitu metode pengumpulan data dengan melakukan tinjauan langsung ke lapangan yang menjadi tempat penelitian. Ada juga studi literatur yaitu merupakan kajian terhadap masalah melalui jurnal, internet untuk melengkapi data yang diperlukan dalam penelitian.

Pengolahan data menggunakan berbagai metode statistik dasar mencari mean, modus, median dan persentase, hasil klasifikasi menggunakan berbagai software seperti Excel. Perancangan sistem menggunakan Diagram Konteks, UML. Perancangan ini akan membantu proses pengembangan aplikasi yang menggunakan bahasa pemrograman. Implementasi sistem adalah proses pengembangan aplikasi menggunakan bahasa pemrograman. Dalam hal ini akan digunakan bahasa pemrograman *python*. Pengujian sistem dilakukan dengan melakukan tahapan-tahapan testing dan debugging program untuk memastikan apakah dapat berjalan dengan baik sesuai dengan rancangan yang sudah di buat sebelumnya.

## 2.2 Skema Sistem

Berikut adalah skema sistem klasifikasi yang akan dibangun dan dirancang dengan menggunakan metode algoritma *K-Nearest Neighbor*, seperti terlihat pada gambar dibawah ini:



Gambar 2. Skema Sistem

Berikut proses alur sistem akan dijelaskan sebagai berikut:

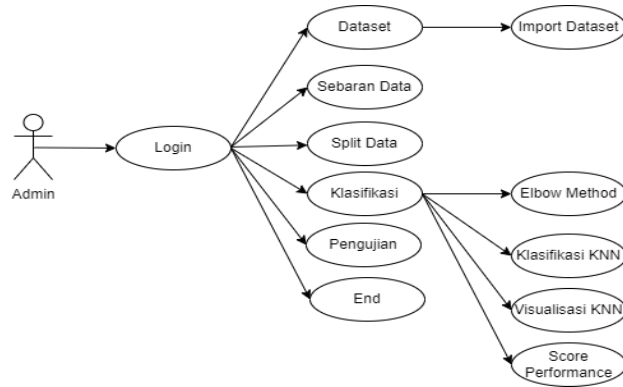
Proses sistem dimulai. *User* menginput atributnya berupa usia, tensi sistolik, tensi diastolik, asam urat, kadar glukosa, kolesterol, berat badan dan tinggi badan. Lalu *user* menentukan nilai k, nilai k yang digunakan yaitu k=5. Lalu *user* melakukan hitung jarak *euclidean distance*. Kemudian *user* mengurutkan jarak *euclidean distance* dari yang terkecil sampai ke terbesar. Selanjutnya *user* mengumpulkan klasifikasi *K-Nearest Neighbor* sesuai jumlah k yaitu 5. Setelah itu tampil hasil tingkat keparahan yang berupa normal, pre-hipertensi, hipertensi 1, hipertensi 2 tergantung risiko yang ditunjukkan. Proses kerja sistem selesai.

## 2.3 Analisis Proses

Sistem ini di rancang menggunakan digram konteks (*Context Diagram*), *DFD* (*Data Flow Diagram*) guna menjelaskan proses sistem sehingga mudah dimengerti dan membantu implementasi ke bahasa pemrograman.

### 2.3.1 Use Case Diagram

Diagram *use case* yang ingin dibuat dalam penelitian ini ditunjukkan pada Gambar dibawah ini:

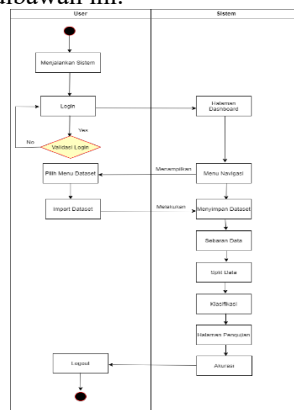


Gambar 3. Use Case Diagram

Dapat dilihat dari gambar diatas bahwasanya disimpulkan yaitu admin membutuhkan *username* dan *password* untuk *login* agar dapat mengakses sistemnya. Setelah admin berhasil *login*, admin bisa mengakses dataset didalam dataset terdapat *import dataset* lalu bisa mengakses sebaran data, split data, klasifikasi dimana klasifikasi di dalam nya terdapat *elbow method*, klasifikasi KNN, visualisasi KNN dan *score performance*. Admin juga bisa mengakses pengujian dan admin juga dapat dapat mengakhiri sistem.

2.3.2 Activity Diagram

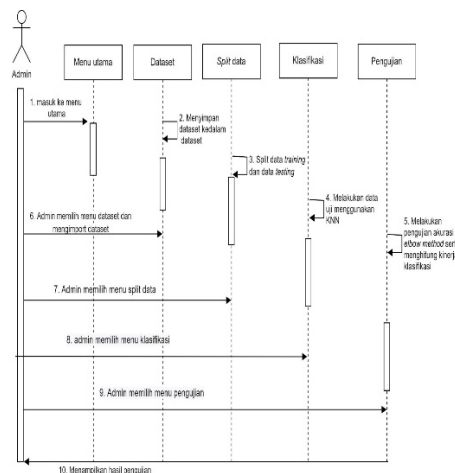
Berikut Gambar activity diagram dibawah ini.



Gambar 4. Activity Diagram

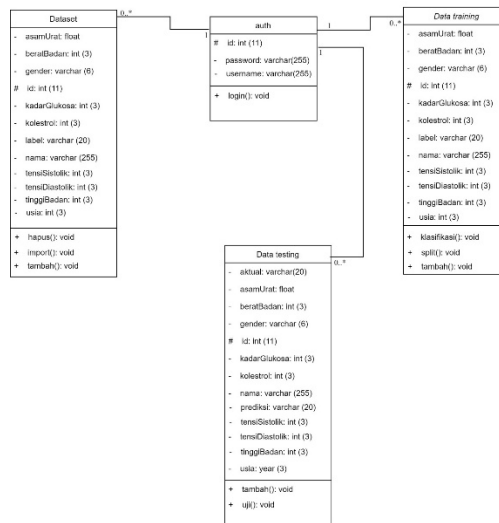
2.3.2 Sequence Diagram

Sequence diagram untuk mendeskripsikan interaksi antar objek di dalam maupun di sekitar sistem yang berupa pesan yang digambarkan terhadap waktu. Berikut *sequence diagram* dibawah ini.



Gambar 5. Sequence Diagram

### 2.3.3 Class Diagram



Gambar 6. Class Diagram

Seorang admin dapat menginput banyak data pasien yang berupa nama, usia, jenis kelamin, tensi sistolik, tensi diastolik, asam urat, kadar kolestrol, kadar glukosa, berat badan dan tinggi badan. Seorang admin juga bisa *split* data yaitu data *training* dan data *testing* dan admin juga bisa melakukan pengujian.

## 3 HASIL DAN PEMBAHASAN

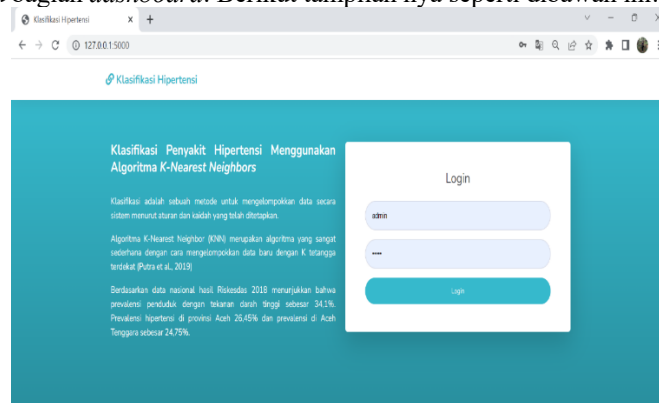
Penelitian ini bertujuan untuk mengklasifikasikan penyakit hipertensi menggunakan algoritma *K-Nearest Neighbor* (KNN) berdasarkan delapan atribut, yaitu usia, tensi sistolik, tensi diastolik, asam urat, kadar glukosa, kolesterol, berat badan, dan tinggi badan. Data yang digunakan berasal dari rekam medis RSUD H. Sahudin tahun 2021–2022 dengan jumlah 478 data pasien. Hasil klasifikasi menunjukkan bahwa kategori hipertensi normal sebesar 72%, pre-hipertensi 1%, hipertensi tingkat 1 sebesar 55%, dan hipertensi tingkat 2 sebesar 16%. Pengujian algoritma KNN menghasilkan tingkat akurasi sebesar 81,25%, sehingga metode ini dinilai efektif dalam mengklasifikasikan penyakit hipertensi.

### 3.1 Hasil Implementasi Sistem

Adapun hasil implementasi sitem yang telah dirancang terdapat beberapa tampilan sebagai berikut:

#### 1. Tampilan Login

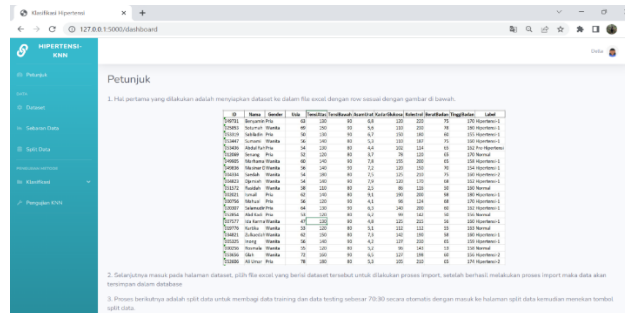
Pertama admin memasukkan *username* dan *password* agar bisa mengakses sistem tersebut. Dalam sistem ini *username* nya berupa “*admin*” dan *password* nya “*admin*”. Kemudian setelah admin memasukkan username beserta passwordnya admin perlu menekan bagian *login* dan sistemnya otomatis menuju bagian *dashboard*. Berikut tampilannya seperti dibawah ini:



Gambar 7. Tampilan Login

#### 2. Tampilan Dashboard

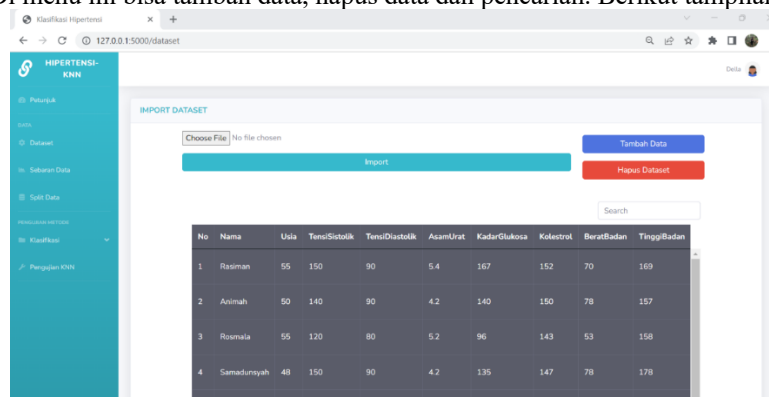
Setelah admin *login*, sistem akan ke halaman *dashboard*. Dalam bagian ini admin bisa melihat dan mmembaca petunjuk dalam menggunakan sistem tersebut. Tampilannya sebagai berikut:



Gambar 8. Tampilan Dashboard

3. Halaman Dataset

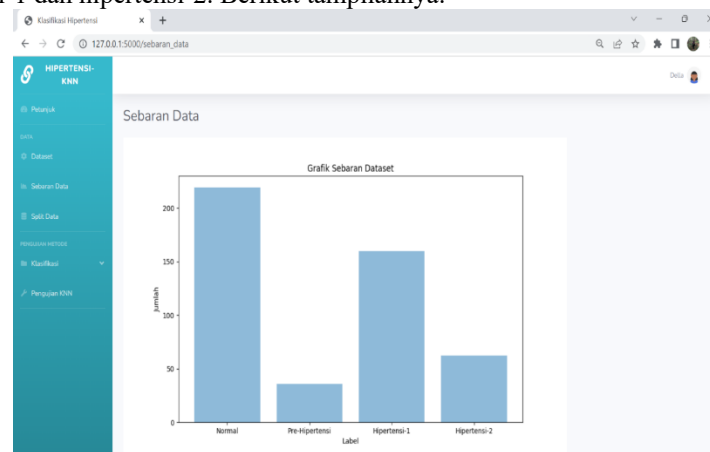
Dalam bagian dataset, admin bisa mengimport dataset berupa *excel* yang ingin dipakai bagi data latihnya serta data ujinya. Dalam menu ini data yang sudah di *import* akan tersimpan dan tampil dengan berbentuk tabel. Di menu ini bisa tambah data, hapus data dan pencarian. Berikut tampilannya:



Gambar 9. Tampilan Dataset

4. Halaman Sebaran Data

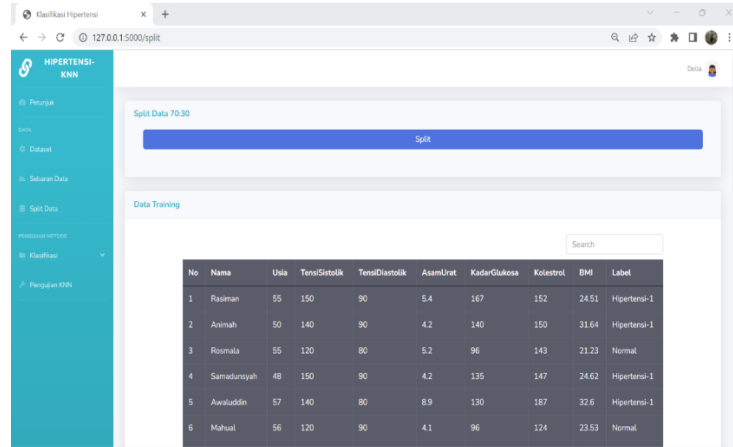
Didalam halaman ini sebaran data yaitu menampilkan berapa banyak jumlah yang normal, pre-hipertensi, hipertensi-1 dan hipertensi-2. Berikut tampilannya:



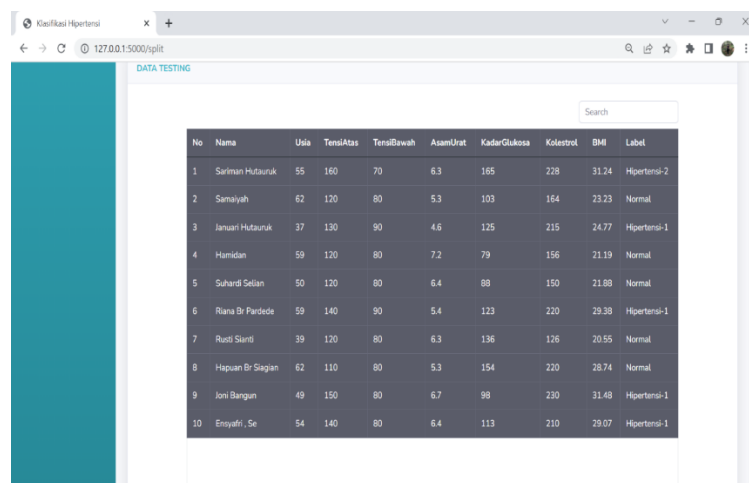
Gambar 10. Tampilan Sebaran Data

5. Halaman Split Data

Di dalam halaman ini split data digunakan untuk meng split data *training* dan data uji dengan split data 70:30. Dimana tabel data *training* dan data *testing* terpisah. Berikut tampilannya:



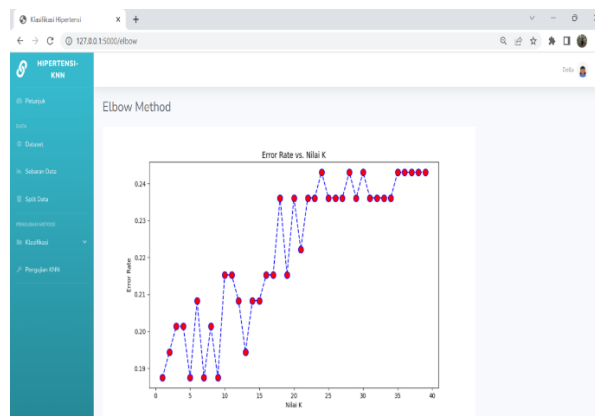
Gambar 11. Tampilan Split Data



Gambar 12. Tampilan Split Data (Lanjutan)

6. Halaman *Elbow Method*

Di dalam halaman *elbow method* berupa grafik yang digunakan untuk mencari nilai k jarak terdekat. Berikut tampilannya:



Gambar 13. Tampilan Elbow Method

7. Halaman Klasifikasi KNN

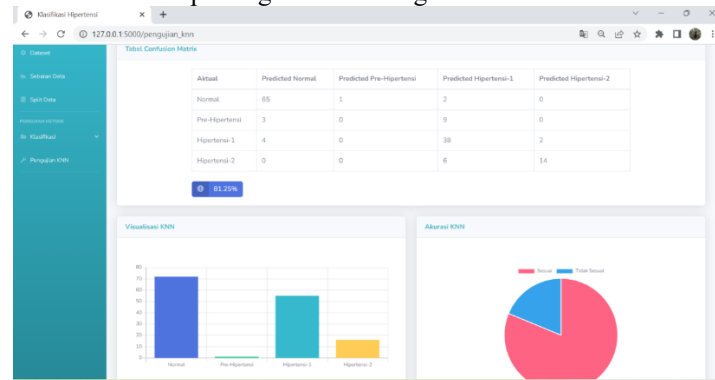
Di dalam halaman ini klasifikasi KNN digunakan untuk mengklasifikasi dari usia, tensi sistolik, tensi diastolik, asam urat, kadar glukosa, kolesterol, berat badan, tinggi badan dan BMI sehingga dapat di klasifikasikan. Berikut tampilannya:

No	Nama	Usia	TensiSistolik	TensiDiastolik	AsamUrut	KadarGlukosa	Kolesterol	BMI	Aktual	Prediksi
1	Sariman Hutauruk	55	160	70	6.3	165	228	31.24	Hipertensi-2	Hipertensi-2
2	Samsyah	62	120	80	5.3	103	164	23.23	Normal	Normal
3	Jansari Hutauruk	37	130	90	4.6	135	215	24.77	Hipertensi-1	Hipertensi-1
4	Hamdan	59	120	80	7.2	79	156	21.19	Normal	Normal
5	Suhardi Solian	50	120	80	6.4	88	150	21.88	Normal	Normal
6	Riana Br-Paradisi	59	140	90	5.4	123	220	29.38	Hipertensi-1	Hipertensi-1
7	Rusti Siant	39	120	80	6.3	138	126	20.55	Normal	Normal
8	Hapsun Br-Sigalan	62	110	80	5.3	154	220	28.74	Pre-Hipertensi	Pre-Hipertensi
9	Jani Bangun	49	150	80	6.7	98	230	31.48	Hipertensi-1	Hipertensi-1

Gambar 14. Tampilan Klasifikasi KNN

8. Halaman Visualisasi KNN

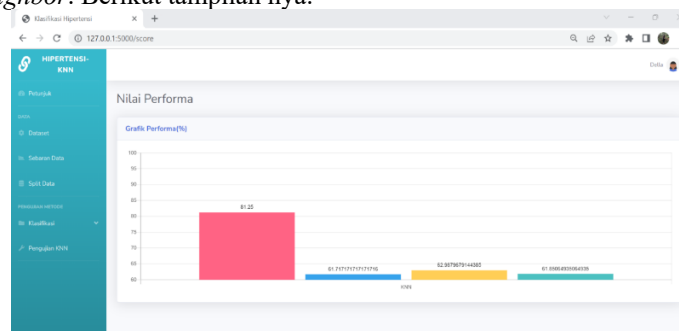
Di dalam halaman ini visualisasi KNN terdapat tabel *confusion matrix* yang digunakan untuk pengujian dari metode *K-Nearest Neighbor*. Terdapat juga persentase berapa persen jumlah normal, pre-hipertensi, hipertensi 1 dan hipertensi 2 dan terdapat tingkat akurasi algoritma tersebut. Berikut tampilannya:



Gambar 15. Tampilan Visualisasi KNN

9. Halaman *Score Performance*

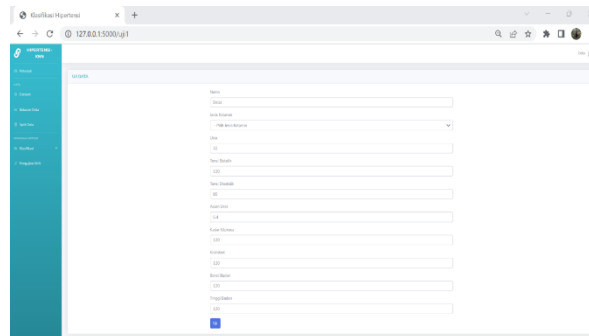
Di dalam halaman ini terdapat grafik performa hasil perbandingan dari akurasi, presisi dan *recall* dari metode *K-Nearest Neighbor*. Berikut tampilannya:



Gambar 16. Tampilan *Score Performance*

10. Pengujian KNN

Di dalam halaman pengujian KNN disini admin bisa menginput pasien dengan nama, jenis kelamin, usia, asam urat, kadar glukosa, kolesterol, berat badan dan tinggi badan lalu admin bisa mengklik uji untuk mengetahui pasien di klasifikasi apa. Berikut tampilannya:



Gambar 17. Tampilan Pengujian KNN

**3.2 Evaluasi Model Performa KNN**

Berikut evaluasi model performa KNN yang melakukan pengujian untuk memperkirakan objek yang benar dan salah.

Tabel 1. Evaluasi Model Performa KNN

Aktual	Prediksi			
	Normal	Pre-Hipertensi	Hipertensi 1	Hipertensi 2
Normal	65	1	2	0
Pre-hipertensi	3	0	9	0
Hipertensi 1	4	0	38	2
Hipertensi 2	0	0	6	14

**3.3 Analisa Pembahasan**

Berdasarkan hasil penelitian, algoritma *K-Nearest Neighbor* (KNN) mampu mengklasifikasikan data pasien hipertensi dengan baik menggunakan atribut yang tersedia. Proses klasifikasi dilakukan berdasarkan perhitungan jarak antara data uji dan data latih untuk menentukan kelas terdekat. Hasil pengujian menunjukkan bahwa algoritma KNN efektif digunakan sebagai metode pendukung dalam klasifikasi penyakit hipertensi.

**3.4 Perhitungan Algoritma K-Nearest Neighbor**

Perhitungan algoritma *K-Nearest Neighbor* (KNN) dilakukan menggunakan 30 data training dan 1 data testing dengan delapan atribut yang digunakan dalam perhitungan jarak. Untuk mempermudah proses perhitungan manual, setiap atribut diberikan kode sesuai dengan nama atributnya, berikut atributnya :

Tabel 2. Kode Atribut Penyakit Hipertensi

Kode Atribut	Atribut
U	Usia
TS	Tensi Sistolik
TD	Tensi Diastolik
Au	Asam Urat
Kgl	Kadar Glukosa
Ch	Kolestrol
BB	Berat Badan
TB	Tinggi Badan

Selanjutnya untuk data training dan data testing masing-masing data dapat dilihat pada Tabel 3 dan 4.

Tabel 3. Data Training

No	Id	Nama	JK	U	TD	TS	Au	Kgl	Ch	BB	TB	BMI	label
1	165	Rasiman	Pria	55	150	90	5,4	167	152	70	169	24,50895	Hipertensi-1
2	174	Animah	Wanita	50	140	90	4,2	140	150	78	157	31,64429	Hipertensi-1
3	256	Rosmala	Wanita	55	120	80	5,2	96	143	53	158	21,23057	Normal
4	390	Samadunyah	Pria	48	150	90	4,2	135	147	78	178	24,6181	Hipertensi-1
5	690	Awaluddin	Pria	57	140	80	8,9	130	187	92	168	32,59637	Hipertensi-1

...	...	...	...	...	...	...	...	...	...	...	...	...	...
30	2022	Ismail Bin Boruddin	Pria	60	130	70	4,5	166	234	94	180	29,01235	Pre-Hipertensi

Tabel 4. Data Testing

No	Id	Nama	JK	U	TD	TS	Au	Kgl	Ch	BB	TB	BMI	label
1	1153	Sariman Hutauruk	Pria	55	160	70	6,3	165	228	83	163	31,23941	Hipertensi-2

Berdasarkan data pada Tabel 2. dan 3. dilakukan perhitungan dengan langkah-langkahnya sebagai berikut:

a. Menentukan parameter K.

Dalam penyelesaian ini diambil beberapa nilai K untuk menentukan parameter K yang akan digunakan dalam penelitian ini. Berikut ini Tabel pemilihan nilai K:

Tabel 5. Pemilihan Nilai K

Keterangan	K = 1	K = 2	K = 3	K = 4	K = 5
Berhasil	116	112	115	112	117
Gagal	28	32	29	32	27

Berdasarkan tabel diatas penulis memilih nilai K = 5 dikarenakan lebih optimal dan tingkat kegagalannya lebih rendah.

b. Kemudian menghitung kuadrat jarak Euclidean.

Kuadrat jarak *euclidean* masing-masing objek terhadap data sampel yang diberikan, dapat dihitung dengan menggunakan persamaan (2.1). Berikut perhitungan jarak *euclidean* untuk 1 data masing-masing pasien:

$$\begin{aligned}
 d1 &= \sqrt{(55 - 55)^2 + (150 - 160)^2 + (90 - 70)^2 + (5,4 - 6,3)^2 +} \\
 &\quad \sqrt{(167 - 165)^2 + (152 - 228)^2 + (24,50895 - 31,23941)^2} \\
 &= \sqrt{0 + 100 + 400 + 0,81 + 4 + 5.776 + 45,29909} \\
 &= 79,53684
 \end{aligned}$$

Hasil perhitungan d1 sampai d30 dapat dilihat pada Tabel berikut:

Tabel 6. Hasil Perhitungan Jarak Euclidean Distance

No	Id	Jarak Euclidean	Label
1	165	79,53684	Hipertensi-1
2	174	86,82496	Hipertensi-1
3	256	117,4197	Normal
4	390	89,76777	Hipertensi-1
5	690	58,46881	Hipertensi-1
...	...	...	...
30	2022	31,14803	Pre-Hipertensi

c. Mengurutkan objek-objek dari jarak yang terkecil. Untuk mengurutkannya kita membuat urutan dari terkecil hingga ke yang terbesar. Berikut Tabel nya sebagai berikut:

Tabel 7. Hasil Pengurutan Jarak dari yang Terkecil

No	Id	Jarak Euclidean	Label
11	1232	30,15687	Hipertensi-2
30	2022	31,14803	Pre-Hipertensi
20	1424	43,77403	Hipertensi-1
21	1667	44,84645	Hipertensi-2
10	1118	46,80763	Hipertensi-1
...	...	...	...
27	2014	146,1605	Normal

d. Mengumpulkan klasifikasi K-Nearest Neighbor, pada tahap ini kita hanya mengambil data sesuai jumlah k yang kita tentukan yaitu 5. Jadi penulis memilih 5 data terdekat saja. Hasil nya seperti Tabel berikut:

Tabel 8. Jarak Terdekat K 5

Perhitungan jarak			
Id	Jarak	Klasifikasi	K
1232	30,15687	Hipertensi-2	1
2022	31,14803	Pre-Hipertensi	2
1424	43,77403	Hipertensi-1	3
1667	44,84645	Hipertensi-2	4
1118	46,80763	Hipertensi 1	5

Pada tahap ini kita dapat mengelompokkan hasil class data uji dengan K=5 berdasarkan hasil class yang di dapat. Dengan menggunakan 333 data training dan 144 data testing pada tahun 2021 sampai 2022 sehingga diperoleh hasil yang dapat dilihat pada Tabel berikut:

Tabel 9. Hasil Pengujian Nilai K= 5

No	Id	Nama	Actual	Prediksi
1	1153	Sariman Hutauruk	Hipertensi-2	Hipertensi-2
2	1401	Samaiyah	Normal	Normal
3	1402	Januari Hutauruk	Hipertensi-1	Hipertensi-1
4	1407	Hamidan	Normal	Normal
5	1425	Suhardi Selian	Normal	Normal
...	...	...	...	...
144	58730	Hami	Hipertensi-2	Hipertensi-1

Tingkat presisi, recall dan akurasi dapat dihitung sebagai berikut:

$$\text{Presisi X} = \frac{\text{Nilai kelas X yang aktual}}{\text{Total nilai kelas X yang di prediksi}}$$

$$\text{Presisi normal} = \frac{65}{65+3+4+0} = \frac{65}{72} = 0,902$$

$$\text{Presisi pre-hipertensi} = \frac{0}{1+0+0+0} = \frac{0}{1} = 0$$

$$\text{Presisi hipertensi 1} = \frac{38}{2+9+38+6} = \frac{38}{55} = 0,690$$

$$\text{Presisi hipertensi 2} = \frac{14}{0+0+2+14} = \frac{14}{16} = 0,875$$

$$\text{Macro average presisi} = \frac{0,902+0+0,690+0,875}{4} = 0,6171$$

$$\text{Recall X} = \frac{\text{Nilai kelas X yang aktual}}{\text{Total nilai kelas X yang seharusnya aktual}}$$

$$\text{Recall normal} = \frac{65}{65+1+2+0} = \frac{65}{68} = 0,955$$

$$\text{Recall pre-hipertensi} = \frac{0}{3+0+9+0} = \frac{0}{12} = 0$$

$$\text{Recall hipertensi 1} = \frac{38}{4+0+38+2} = \frac{38}{44} = 0,863$$

$$\text{Recall hipertensi 2} = \frac{14}{0+0+6+14} = \frac{14}{20} = 0,7$$

$$\text{Macro average recall} = \frac{0,955+0+0,863+0,7}{4} = 0,6298$$

$$\text{Akurasi} = \frac{\text{true normal+true pre-hipertensi+true hipertensi 1+true hipertensi 2}}{\text{Total data yang di uji}}$$

$$\text{Akurasi} = \frac{64+0+40+13}{144} = \frac{117}{144} = 0,8125$$

Tabel 10. Hasil Presisi, Recall, dan akurasi

Kelas	Presisi	Recall	Akurasi
Normal	0,902	0,955	0,8125
Pre-hipertensi	0	0	
Hipertensi 1	0,690	0,863	

Hipertensi 2	0,875	0,7
Macro	0,6171	0,6298

Dari Tabel diatas untuk algoritma K-Nearest Neighbor dengan pendekatan Eulidean Distance pada penelitian ini didapatkan persentase presisi sebesar 64% recall sebesar 62% dan akurasi sebesar 81.25%.

#### 4 KESIMPULAN

Berdasarkan hasil penelitian dan pembahasan, dapat disimpulkan bahwa sistem klasifikasi penyakit hipertensi berhasil dibangun menggunakan bahasa pemrograman Python dengan MySQL sebagai basis data. Penelitian ini menggunakan 478 data pasien RSUD H. Sahudin, terdiri dari 333 data training dan 144 data testing. Penerapan metode K-Nearest Neighbor (KNN) menghasilkan klasifikasi normal, pre-hipertensi, hipertensi tingkat 1, dan hipertensi tingkat 2. Nilai K = 5 dipilih karena memberikan kinerja terbaik dengan tingkat kesalahan yang rendah. Hasil pengujian menunjukkan distribusi klasifikasi sebesar 72% normal, 1% pre-hipertensi, 55% hipertensi tingkat 1, dan 16% hipertensi tingkat 2, dengan akurasi sebesar 81,25%.

#### REFERENSI

- [1] Dinata, R. K., Akbar, H., & Hasdyna, N. (2020). Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus. *ILKOM Jurnal Ilmiah*, 12(2), 104–111. <https://doi.org/10.33096/ilkom.v12i2.539.104-111>
- [2] A. Khaidar and S. Kurnia, "Classification of health indicators for diabetes mellitus prediction using a TabTransformer model on clinical tabular data," *Journal of Computer Science and Research (JoCoSiR)*, vol. 2, no. 2, pp. 8–15, 2024.
- [3] A. Khaidar, M. Arhami, and M. Abdi, "Application of machine learning algorithms for student classification," *Journal of Artificial Intelligence and Software Engineering*, vol. 4, no. 2, pp. 94–103, 2024.
- [4] Sari, S. K., & Mahmudy, W. F. (2019). *Penerapan Metode Decision Tree dan Algoritme Genetika Untuk Klasifikasi Risiko Hipertensi*. 3(3), 2867–2873.
- [5] P2PTM Kemenkes RI. (2019). Hipertensi, The Silent Killer - Direktorat P2PTM. In 2018. <http://p2ptm.kemkes.go.id/infographic-p2ptm/hipertensi-penyakit-jantung-dan-pembuluh-darah/hipertensi-the-silent-killer> Pt, D. I., & Lestari, P. (2018). 1) , 2). 9(September), 67–78.
- [6] J. Han, J. Pei, and H. Tong, "Data mining: Concepts and techniques," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 2, pp. 1–18, 2020, doi: 10.1109/TKDE.2019.2941637.
- [7] M. Eslami, A. Ghasemaghaei, and K. Hassanein, "Data mining techniques for extracting knowledge from large datasets," *Expert Systems with Applications*, vol. 154, pp. 1–12, 2020, doi: 10.1016/j.eswa.2020.113476.
- [8] S. A. Salloum, M. Alshurideh, A. Elnagar, and K. Shaalan, "Machine learning and data mining techniques for healthcare data classification," *IEEE Access*, vol. 8, pp. 191–205, 2020, doi: 10.1109/ACCESS.2020.2968740.
- [9] Athalla, I. N., Jovandy, A., & Habibie, H. (2018). Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K-Nearest Neighbor. *Prosiding Annual Research Seminar*, 4(1), 978–979.
- [10] Dinata, R. K., Akbar, H., & Hasdyna, N. (2020). Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus. *ILKOM Jurnal Ilmiah*, 12(2), 104–111. <https://doi.org/10.33096/ilkom.v12i2.539.104-111>
- [11] Agustinus, I., Santoso, E., & Rahayudi, B. (2018). Klasifikasi Risiko Hipertensi Menggunakan Metode Learning Vector Quantization ( LVQ ). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(8), 2947–2955. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1725/654>
- [12] Agustinus, I., Santoso, E., & Rahayudi, B. (2018). Klasifikasi Risiko Hipertensi Menggunakan Metode Learning Vector Quantization ( LVQ ). *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer (J-PTIIK) Universitas Brawijaya*, 2(8), 2947–2955. <http://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/1725/654>
- [13] Bustami, 2012, Teknik, D. I., & Bayes, N. (2018). "Penerapan Algoritma Naive Bayes Untuk Mengklasifikasi Data Nasabah." *Jurnal Penelitian Teknik Informatika Universitas Malikussaleh*, 146(Klasifikasi), 128–146.
- [14] A. A. Alyahya, M. A. Al-Turki, and S. A. Alghamdi, "Hypertension prediction using machine learning algorithms," *Journal of Healthcare Engineering*, vol. 2021, pp. 1–9, 2021, doi: 10.1155/2021/6638436.