

**KLASIFIKASI PENYAKIT KARDIOVASKULAR DENGAN  
ALGORITMA K-NEAREST NEIGHBOR**

**LAPORAN TUGAS AKHIR**

Laporan Tugas Akhir ini Disusun Sebagai Salah Satu Syarat Untuk Memperoleh  
Gelar Sarjana (S1) Pada Program Studi Teknik Informatika  
Universitas Islam Sultan Agung Semarang



**OLEH :**  
**IKHSANUL KARIM**  
**32601601046**

**PROGRAM STUDI TEKNIK INFORMATIKA  
FAKULTAS TEKNOLOGI INDUSTRI  
UNIVERSITAS ISLAM SULTAN AGUNG  
SEMARANG  
2023**

***FINAL PROJECT***

***CLASSIFICATION OF CARDIOVASCULAR DISEASE USING THE K-  
NEAREST NEIGHBOR ALGORITHM***

*This Final Project Report Is Compiled As One Of The Requirements To Obtain  
Bachelor Degree (S1) in Informatics Engineering Study Program  
Sultan Agung Islamic University Semarang*



**32601601046**

**MAJORING OF INFORMATICS ENGINEERING  
INDUSTRIAL TECHNOLOGY FACULTY  
SULTAN AGUNG ISLAMIC UNIVERSITY  
SEMARANG  
AUGUST 2023**

## LEMBAR PENGESAHAN PEMBIMBING

Laporan Tugas Akhir dengan judul “**Klasifikasi Penyakit Kardiovaskular Dengan Algoritma K-Nearest Neighbor**” ini disusun oleh :

Nama : Ikhsanul Karim

Nim : 32601601046

Program Studi : Teknik Informatika.

Telah disetujui oleh Dosen Pembimbing pada :

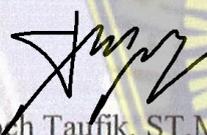
Hari : **selasa** .....

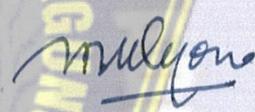
Tanggal : **5. september 2023** .....

Menyetujui,

Pembimbing I

Pembimbing II

  
Moch Taufik, ST.M.IT  
NIDN. 0622037502

  
Ir. Sri Mulyono, M.Eng  
NIDN. 0626066601

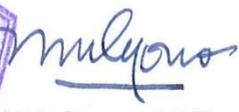
Mengetahui,

Ketua Program Studi Teknik Informatika

Fakultas Teknologi Industri

Universitas Islam Sultan Agung



  
Ir. Sri Mulyono, M.Eng  
NIDN. 0626066601

## LEMBAR PENGESAHAN PENGUJI

Laporan tugas akhir dengan judul “Klasifikasi Penyakit Kardiovaskular Dengan Algoritma *K-Nearest Neighbor*” ini telah dipertahankan di depan tim penguji proposal Tugas Akhir pada :

Hari : Selasa .....

Tanggal : 5 September 2023 .....

### TIM PENGUJI

Penguji I



Andi Riansyah, ST., M.Kom

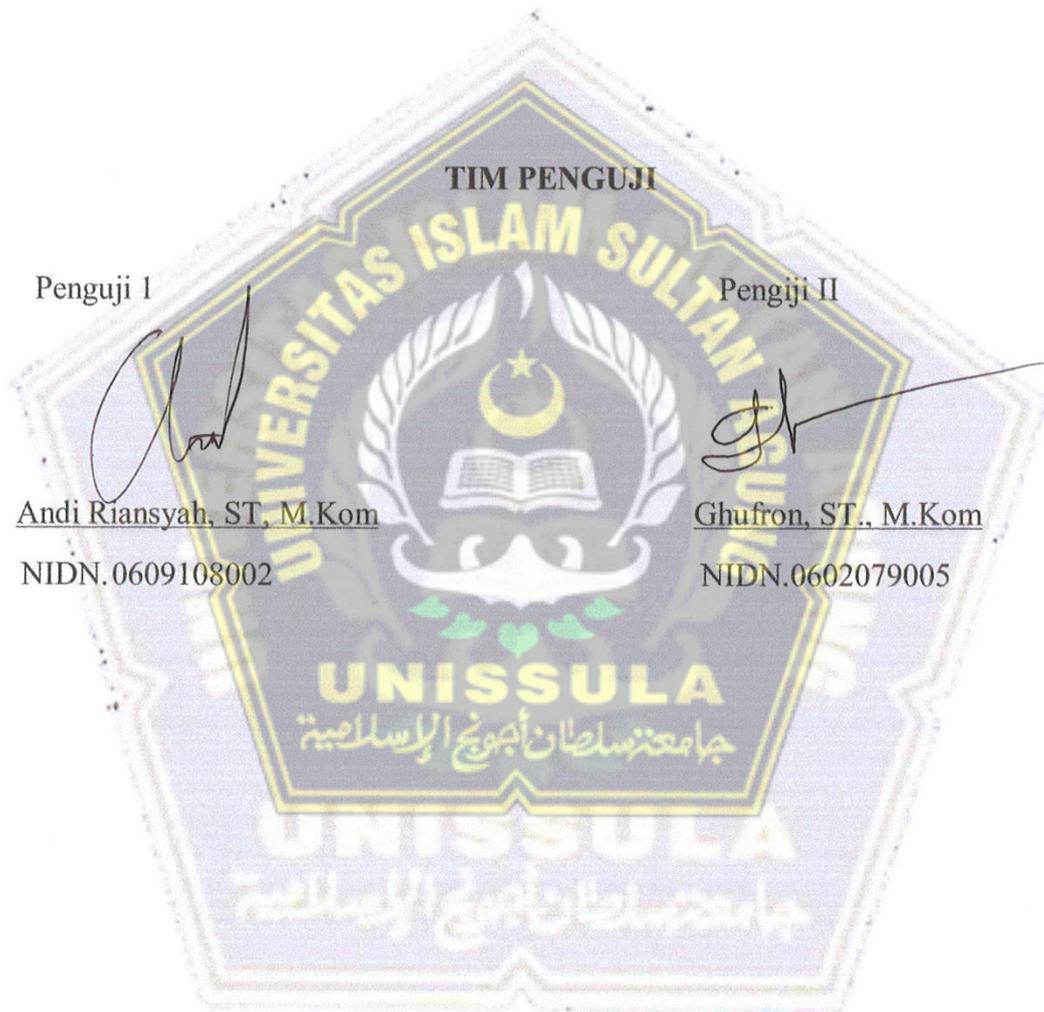
NIDN.0609108002

Penguji II



Ghufron, ST., M.Kom

NIDN.0602079005



## SURAT PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan dibawah ini :

Nama : Ikhsanul Karim

NIM : 32601601046

Judul Tugas Akhir : Klasifikasi Penyakit Kardiovaskular Dengan Algoritma *K-Nearest Neighbor*

Dengan bahwa ini saya menyatakan bahwa judul dan isi Tugas Akhir yang saya buat dalam rangka menyelesaikan Pendidikan Strata Satu (S1) Teknik Informatika tersebut adalah asli dan belum pernah diangkat, ditulis ataupun dipublikasikan oleh siapapun baik keseluruhan maupun sebagian, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka, dan apabila di kemudian hari ternyata terbukti bahwa judul Tugas Akhir tersebut pernah diangkat, ditulis ataupun dipublikasikan, maka saya bersedia dikenakan sanksi akademis. Demikian surat pernyataan ini saya buat dengan sadar dan penuh tanggung jawab.

Semarang, ..... September 2023  
Yang Menyatakan,



Ikhsanul Karim

## PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH

Saya yang bertanda tangan dibawah ini :

Nama : Ikhsanul Karim

NIM : 32601601046

Program Studi : Teknik Informatika

Fakultas : Teknologi industri

Alamat Asal : Desa Rau RT 03 RW 02 Kec. Kedung Kab.Jepara

Dengan ini menyatakan Karya Ilmiah berupa Tugas akhir dengan Judul : **Klasifikasi Penyakit Kardiovaskular Dengan Algoritma *K-Nearest Neighbor*** Menyetujui menjadi hak milik Universitas Islam Sultan Agung serta memberikan Hak bebas Royalti Non-Eksklusif untuk disimpan, dialihmediakan, dikelola dan pangkalan data dan dipublikasikan diinternet dan media lain untuk kepentingan akademis selama tetap menyantumkan nama penulis sebagai pemilik hak cipta. Pernyataan ini saya buat dengan sungguh-sungguh. Apabila dikemudian hari terbukti ada pelanggaran Hak Cipta/Plagiarisme dalam karya ilmiah ini, maka segala bentuk tuntutan hukum yang timbul akan saya tanggung secara pribadi tanpa melibatkan Universitas Islam Sultan agung.

Semarang, 5 September 2023

menyatakan,



Ikhsanul Karim

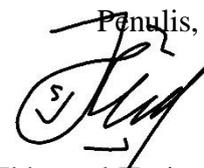
## KATA PENGANTAR

Puji syukur kehadiran Allah SWT atas limpahan Rahmat dan Karunia-Nya, sehingga penulis dapat menyelesaikan Tugas Akhir yang berjudul **“Klasifikasi Penyakit Kardiovaskular Dengan Algoritma K-Nearest Neighbor”**. Penyusunan Laporan Tugas Akhir merupakan syarat untuk memperoleh gelar Sarjana Strata 1 (S1) Teknik Informatika Universitas Islam Sultan Agung Semarang. Dengan selesainya penyusunan Laporan Tugas Akhir ini penulis menyampaikan terimakasih kepada:

1. Bapak Prof. Dr. H. Gunarto, SH., M.Hum selaku Rektor Universitas Islam Sultan Agung .
2. Ibu Dr. Novi Marlyana, ST, MT. selaku Dekan Fakultas Teknologi Industri Universitas Islam Sultan Agung Semarang.
3. Bapak Ir. Sri Mulyono, M.Eng, selaku Ketua Program Studi Teknik Informatika Universitas Islam Sultan Agung Semarang.
4. Bapak Moch Taufik, ST.M.IT selaku Dosen Pembimbing I.
5. Bapak Ir. Sri Mulyono, M.Eng selaku Dosen Pembimbing II.
6. Orang tua dan keluarga yang selalu memberi suportnya kepada penulis.
7. Teman – Teman dan semua pihak yang tidak dapat di sebutkan satu persatu.

Penulis menyadari masih banyak kesalahan dan kekurangan dalam penyusunan Laporan Tugas Akhir ini. Oleh karena itu, penulis mengharapkan kritik, saran dan bimbingan akan sangat membantu bagi pengembangan dan perbaikan tugas akhir ini dimasa yang akan datang. Akhir kata apabila ada uraian dan penjelasan yang kurang berkenan, penulis mengucapkan permohonan maaf yang sebesar-besarnya.

Semarang, 5 September 2023

Penulis,  


Ikhsanul Karim

NIM. 32601601046

## DAFTAR ISI

KLASIFIKASI PENYAKIT KARDIOVASKULAR DENGAN ALGORITMA K-NEAREST NEIGHBOR .....	i
LEMBAR PENGESAHAN PEMBIMBING .....	iii
LEMBAR PENGESAHAN PENGUJI .....	iv
SURAT PERNYATAAN KEASLIAN TUGAS AKHIR.....	v
PERNYATAAN PERSETUJUAN PUBLIKASI KARYA ILMIAH .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI .....	viii
DAFTAR GAMBAR .....	x
DAFTAR TABEL .....	xi
ABSTRAK .....	xii
BAB I PENDAHULUAN .....	1
1.1. Latar Belakang.....	1
1.2. Perumusan Masalah.....	2
1.3. Pembatasan Masalah.....	2
1.4. Tujuan.....	3
1.5. Manfaat .....	3
1.6. Sistematika Penulisan .....	3
BAB II TINJAUAN PUSTAKA DAN DASAR TEORI.....	5
2.1. Tinjauan Pustaka.....	5
2.2. Dasar Teori.....	6
2.2.1 Data Mining.....	6
2.2.2 Klasifikasi .....	8
2.2.3 KNN.....	9
2.2.4 UML ( <i>Unified Modelling Language</i> ) .....	10
2.2.5 PHP .....	17
2.2.6 <i>MySQL</i> .....	18
2.2.7 <i>Confusion Matrix</i> .....	19
BAB III METODE PENELITIAN .....	21

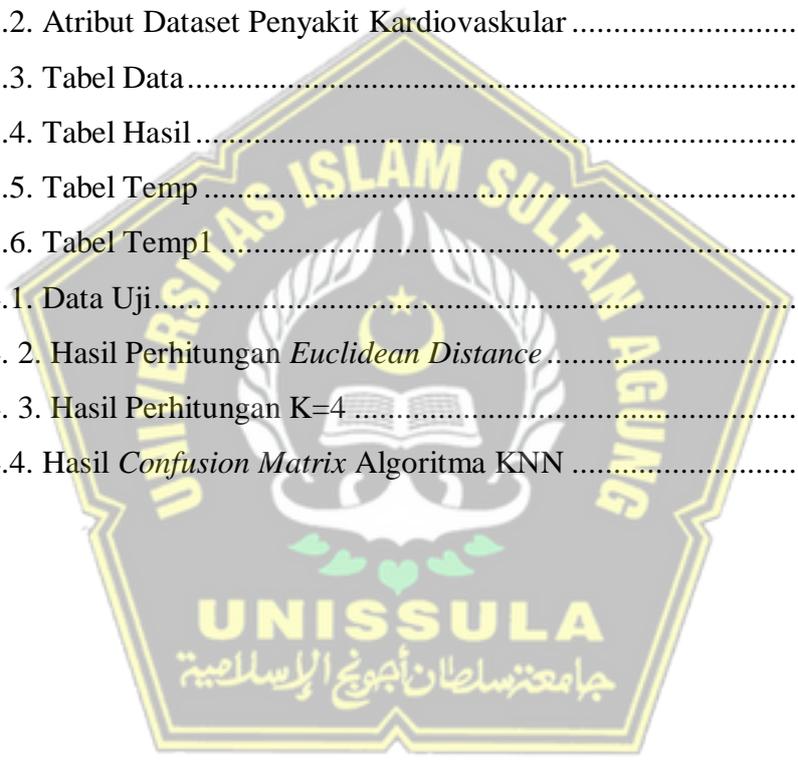
3.1	Metode Penelitian .....	21
3.1.1	Metode Pengumpulan Data .....	21
3.1.2	Metode Pengembangan Sistem .....	21
3.2	Identifikasi Masalah.....	22
3.3	Identifikasi Perangkat Keras ( <i>Hardware</i> ) .....	22
3.4	Identifikasi Kebutuhan Perangkat Lunak ( <i>Software</i> ) .....	22
3.5	Analisis Kebutuhan Data .....	23
3.6	<i>Flowchart</i> Sistem.....	24
3.7	<i>Use Case Diagram</i> .....	26
3.8	<i>Activity Diagram</i> .....	26
3.9	<i>Sequence Diagram</i> .....	30
3.10	<i>Class Diagram</i> .....	32
3.11	Rancangan Database .....	33
3.12	Rancangan <i>Interface</i> .....	35
3.13	Metode Pengujian Sistem.....	38
BAB IV	HASIL DAN ANALISIS PENELITIAN .....	40
4.1	<i>Login</i> .....	40
4.2	<i>Home</i> .....	40
4.3	Pengujian.....	41
4.4	Akurasi .....	42
4.5	Klasifikasi .....	43
4.6	<i>Admin</i> .....	44
4.7	Perhitungan KNN .....	45
BAB V	PENUTUP .....	51
5.1.	Kesimpulan.....	51
5.2.	Saran .....	51
DAFTAR PUSTAKA	.....	52
LAMPIRAN-LAMPIRAN		

## DAFTAR GAMBAR

Gambar 3.1. <i>Flowchart</i> Sistem.....	25
Gambar 3.2 . <i>Use Case Diagram</i> .....	26
Gambar 3.3. <i>Activity Diagram Login</i> .....	27
Gambar 3. 4. <i>Activity Diagram</i> Mengelola Kardiovaskular.....	28
Gambar 3. 5. <i>Activity Diagram</i> Pengujian KNN .....	28
Gambar 3. 6. <i>Activity Diagram</i> Klasifikasi Kardiovaskular .....	29
Gambar 3.7. <i>Activity Diagram Logout</i> .....	29
Gambar 3. 8. <i>Sequence Diagram Login</i> .....	30
Gambar 3. 9. <i>Sequence Diagram</i> Mengeloa Kardiovaskular.....	31
Gambar 3.10. <i>Sequence Diagram</i> Pengujian KNN .....	31
Gambar 3.11. <i>Sequence Diagram</i> Klasifikasi Kardiovaskular.....	32
Gambar 3. 12. <i>Sequence Diagram Logout</i> .....	32
Gambar 3. 13. <i>Class Diagram</i> .....	33
Gambar 3. 14. <i>Login</i> .....	35
Gambar 3.15. Input Data Kardiovaskular .....	36
Gambar 3. 16. Pengujian.....	36
Gambar 3. 17. Hasil .....	37
Gambar 3. 18. Klasifikasi.....	37
Gambar 3.19. <i>Admin</i> .....	38
Gambar 4.1. <i>Login</i> .....	40
Gambar 4.2. Home.....	41
Gambar 4.3. Pengujian.....	42
Gambar 4.4. Akurasi .....	42
Gambar 4.5. Klasifikasi .....	43
Gambar 4.6. Hasil Klasifikasi .....	44
Gambar 4.7. <i>Admin</i> .....	44
Gambar 4. 8. Hasil Akurasi K=4 .....	49
Gambar 4. 9. Hasil Akurasi K=3 .....	49
Gambar 4. 10. Hasil Akurasi K=1 .....	50

## DAFTAR TABEL

Tabel 2. 1. Simbol <i>Use Case</i> .....	11
Tabel 2.2. Simbol <i>Class Diagram</i> .....	13
Tabel 2. 3. Simbol <i>Activity Diagram</i> .....	15
Tabel 2. 4. Simbol <i>Sequence Diagram</i> .....	16
Tabel 2. 5. <i>Confusion Matrix</i> .....	19
Tabel 3.1. Dataset Penyakit Kardiovaskular .....	23
Tabel 3.2. Atribut Dataset Penyakit Kardiovaskular .....	24
Tabel 3.3. Tabel Data.....	34
Tabel 3.4. Tabel Hasil.....	34
Tabel 3.5. Tabel Temp .....	34
Tabel 3.6. Tabel Temp1 .....	35
Tabel 4.1. Data Uji.....	45
Tabel 4. 2. Hasil Perhitungan <i>Euclidean Distance</i> .....	47
Tabel 4. 3. Hasil Perhitungan K=4 .....	47
Tabel 4.4. Hasil <i>Confusion Matrix</i> Algoritma KNN .....	48



## ABSTRAK

Penyakit kardiovaskular merupakan salah satu masalah kesehatan utama di negara maju maupun berkembang. Penyakit ini menjadi penyebab nomor satu kematian di dunia setiap tahunnya. Sejak tahun 2008 diperkirakan sebanyak 17,3 juta kematian disebabkan oleh penyakit kardiovaskular. Tingginya faktor kematian akibat penyakit kardiovaskular dapat dicegah dan ditekan faktor risikonya jika masyarakat memiliki pengetahuan tentang gejala penyakit kardiovaskular. Banyaknya faktor-faktor yang dikumpulkan guna mendeteksi seseorang apakah menderita penyakit kardiovaskular atau tidak membutuhkan sistem olah data yang cukup besar. Data yang digunakan berjumlah 1.000 data yang didapatkan dari *website* Kaggle Data Set yang terdiri dari atribut *age, height, weight, gender, systolic, diastolic, cholesterol, glucose, smooking, alcohol, physical activity*. Penelitian yang dilakukan akan mengubah nilai parameter K dari metode KNN menjadi 3 varian nilai saja yaitu 1,3 dan 4. Nilai akurasi tertinggi dari perubahan nilai parameter K yang terdiri dari 1,3 dan 4 pada metode KNN didapatkan hasil akurasi tertinggi pada nilai parameter K = 4 dengan akurasi 88 %

**Kata Kunci :** Data Mining, Kardiovaskular, KNN

## ABSTRACT

*Cardiovascular disease is a major health problem in both developed and developing countries. This disease is the number one cause of death in the world every year. Since 2008 an estimated 17.3 million deaths are caused by cardiovascular disease. The high mortality factor from cardiovascular disease can be prevented and the risk factors reduced if people have knowledge about the symptoms of cardiovascular disease. The many factors that are collected to detect whether a person has cardiovascular disease or not require a large data processing system. The data used amounted to 1,000 data obtained from the Kaggle Data Set website consisting of the attributes of age, height, weight, gender, systolic, diastolic, cholesterol, glucose, smoking, alcohol, physical activity. The research carried out will change the value of the K parameter from the KNN method to only 3 value variants, namely 1.3 and 4. The highest accuracy value for changing the K parameter values consisting of 1.3 and 4 in the KNN method produces the highest accuracy results at the parameter value K = 4 with an accuracy of 88.*

**Keyword :** Data Mining, Kardiovaskular, KNN

# **BAB I**

## **PENDAHULUAN**

### **1.1. Latar Belakang**

Penyakit kardiovaskular merupakan salah satu masalah kesehatan utama di negara maju maupun berkembang. Penyakit ini menjadi penyebab nomor satu kematian di dunia setiap tahunnya. Sejak tahun 2008 diperkirakan sebanyak 17,3 juta kematian disebabkan oleh penyakit kardiovaskular (Juslim & Herawati, 2018). Lebih dari 3 juta kematian tersebut terjadi sebelum usia 60 tahun. Terjadinya kematian dini yang disebabkan oleh penyakit kardiovaskular berkisar sebesar 4% di negara berpenghasilan tinggi, dan 42% terjadi di negara berpenghasilan rendah. Kematian yang disebabkan oleh penyakit kardiovaskular diperkirakan akan terus meningkat mencapai 23,3 juta kematian pada tahun 2030 (Benjamin, 2012). Di Indonesia penyakit kardiovaskular ini terus meningkat dan akan memberikan beban kesakitan, kecacatan dan beban sosial ekonomi bagi keluarga penderita, masyarakat, dan negara (Fadlilah dkk, 2019).

Penyakit kardiovaskular adalah penyakit yang disebabkan gangguan fungsi jantung dan pembuluh darah. Penyebab utama penyakit kardiovaskular adalah konsumsi tembakau, aktivitas fisik yang kurang, diet yang tidak sehat dan penggunaan berbahaya dari alkohol (Zhao, 2021). Faktor resiko penyakit kardiovaskular terdiri dari faktor resiko yang tidak dapat dimodifikasi dan yang dapat dimodifikasi. Faktor yang tidak dapat dimodifikasi yaitu riwayat keluarga, umur, jenis kelamin, dan obesitas, sedangkan faktor yang dapat dimodifikasi seperti hipertensi, diabetes melitus, dislipidemia, kurang aktivitas fisik, diet tidak sehat, dan stres (Zhao, 2021).

Semua orang memiliki risiko terkena penyakit kardiovaskular. Tingginya faktor kematian akibat penyakit kardiovaskular dapat dicegah dan ditekan faktor resikonya jika masyarakat memiliki pengetahuan tentang gejala penyakit kardiovaskular. Banyaknya faktor-faktor yang dikumpulkan guna mendeteksi seseorang apakah menderita penyakit kardiovaskular atau tidak membutuhkan sistem olah data yang cukup besar. Oleh karena itu perlunya sebuah penggalian

informasi dari faktor-faktor merupakan langkah yang strategis. Data mining merupakan disiplin ilmu yang mempelajari metode untuk mengesktrak pengetahuan atau menemukan pola dari suatu data yang besar. Data mining terbukti merupakan langkah yang terbaik dalam menangani jumlah data yang besar dan menghasilkan sejumlah data dengan cepat dengan akurasi yang tinggi serta mampu mengolah dari berbagai sumber data (Hermawati, 2016).

Metode data mining yang digunakan pada penelitian ini adalah metode *K-Nearest Neighbor* (KNN). Algoritma KNN adalah algoritma yang sering digunakan untuk klasifikasi, meskipun juga dapat digunakan untuk estimasi dan prediksi. Tetangga terdekat K adalah contoh dasar pembelajaran di mana pelatihan data disimpan sehingga klasifikasi untuk data yang tidak terarah dapat ditemukan jika dibandingkan dengan catatan pelatihan data (Kusrini & Luthfi, 2015).

Penelitian oleh (Isman, et al., 2021) melakukan pebandingan antara metode KNN dengan metode *Local Binary Pattern Histogram* (LBPH) dengan hasil metode KNN memiliki akurasi yang lebih baik yaitu 97,5 % dibandingkan dengan metode LBPH dengan akurasi sebesar 94 % dalam menentukan klasifikasi daun herbal. Berdasarkan permasalahan diatas, penelitian ini mengambil judul “Klasifikasi Penyakit Kardiovaskular Dengan Algoritma *K-Nearest Neighbor*”.

### **1.2. Perumusan Masalah**

Berdasarkan uraian latar belakang yang ada maka dirumuskan pokok permasalahannya adalah bagaimana merancang sistem yang dapat digunakan untuk melakukan klasifikasi penyakit kardiovaskular ke dalam klasifikasi positif atau negatif menggunakan algoritma KNN.

### **1.3. Pembatasan Masalah**

Pembatasan masalah dilakukan agar pembahasan yang akan dibahas dalam penelitian ini lebih terarah, maka penulis membatasi masalah meliputi:

1. Data yang digunakan berjumlah 1.000 data yang didapatkan dari *website* Kaggle Data Set yang terdiri dari atribut *age, height, weight, gender, systolic, diastolic, cholesterol, glucose, smooking, alcohol, physical activity*.

2. Penelitian yang dilakukan akan mengubah nilai parameter K dari metode KNN menjadi 3 varian nilai saja yaitu 1,3 dan 4.

#### **1.4. Tujuan**

Tujuan penelitian ini adalah membuat sistem yang dapat digunakan untuk melakukan klasifikasi penyakit kardiovaskular ke dalam klasifikasi positif atau negatif menggunakan algoritma KNN.

#### **1.5. Manfaat**

Manfaat yang diperoleh dari penelitian klasifikasi penyakit kardiovaskular menggunakan metode KNN adalah

1. Membantu memberikan informasi deteksi awal dalam menentukan penyakit kardiovaskular sehingga dapat dilakukan penanganan lebih lanjut.
2. Sebagai alternatif alat bantu untuk melakukan klasifikasi penyakit kardiovaskular.

#### **1.6. Sistematika Penulisan**

Untuk memudahkan pemahaman tentang penulisan Tugas Akhir ini, maka pembahasan akan dibagi dalam beberapa bab sesuai dengan pokok permasalahannya, yaitu:

#### **BAB I PENDAHULUAN**

Berisi latar belakang, rumusan masalah, batasan masalah, tujuan, manfaat, dan sistem penulisan.

#### **BAB II TINJAUAN PUSTAKA DAN DASAR TEORI**

Berisi penjelasan atau pembahasan dari teori-teori penelitian terkait topik yang dipilih dan akan menjadi dasar pada penulisan tugas akhir. Teori-teori ini telah dikutip dalam banyak artikel.

#### **BAB III METODE PENELITIAN**

Berisi penerapan dalam rancangan teori dasar sebagai pendekatan untuk mendapatkan jalan keluar masalah dalam pembuatan sistem klasifikasi penyakit kardiovaskular menggunakan metode KNN.

#### **BAB IV HASIL DAN ANALISIS PENELITIAN**

Meliputi hasil pengujian dari pembuatan klasifikasi penyakit kardiovaskular menggunakan metode KNN, menguji perangkat, dan menemukan penyebab jika ternyata tidak konsisten atau berbeda dari pemikiran atau hasil pengujian yang ada.

#### **BAB V KESIMPULAN DAN SARAN**

Bab ini memberikan kesimpulan dan saran dari klasifikasi penyakit kardiovaskular menggunakan metode KNN.



## BAB II

### TINJAUAN PUSTAKA DAN DASAR TEORI

#### 2.1. Tinjauan Pustaka

Kanker payudara adalah penyakit non kulit yang berasal dari sel kelenjar, saluran kelenjar, dan jaringan penunjang payudara. Penelitian tersebut bertujuan untuk melakukan klasifikasi penyakit kanker payudara dengan metode KNN. Dataset kanker payudara yang diambil dari sklearn dengan 30 atribut dan 569 baris data. Metode KNN dapat melakukan klasifikasi penyakit kanker payudara sebesar 93% (Atthalla dkk, 2018).

*Diabetes Mellitus* merupakan salah satu penyakit kronis yang mematikan yang diakibatkan kadar glukosa di dalam darah terlalu tinggi. Menurut WHO saat ini lebih dari 246 juta jiwa menderita diabetes dan diperkirakan akan meningkat menjadi 380 juta jiwa pada tahun 2025 apabila tidak dilakukan penanganan yang serius. Diabetes menyebabkan penyakit lain / komplikasi yang setiap tahunnya mengakibatkan kematian hingga 3,8 juta jiwa. Dari hasil penelitian yang telah dilakukan disimpulkan bahwa penggunaan algoritma seleksi fitur *gain ratio* dapat meningkatkan akurasi dari klasifikasi penyakit diabetes mellitus dengan menggunakan algoritma KNN. Kenaikan akurasi tertinggi didapatkan pada nilai treshold 0,152 dengan hanya mempertahankan 4 atribut dari keseluruhan 8 atribut data (Indrayanti dkk, 2017).

*Tweet* yang berasal dari calon karyawan akan di proses dan setelah itu menghasilkan karakter-karakter sebagai salah satu acuan dalam penempatan calon karyawan. Dengan memasukkan nilai K yaitu 4 sebagai nilai yang diuji didapatkan hasil akurasi sistem yang diperoleh dari klasifikasi karakter calon karyawan berdasarkan *tweet* adalah 66%. Hasil tersebut merupakan hasil dimana ada 53 hasil data uji yang benar dan 27 hasil data uji yang salah dalam proses pengujian (Claudy dkk, 2018).

Penelitian selanjutnya akan dilakukan penelitian tentang klasifikasi penyakit kardiovaskular menggunakan metode KNN. Data yang digunakan

berjumlah 1.000 data yang didapatkan dari *website* Kaggle Data Set yang terdiri dari atribut *age, height, weight, gender, systolic, diastolic, cholesterol, glucose, smooking, alcohol, physical activity*. Penelitian yang dilakukan akan mengubah nilai parameter K dari metode KNN menjadi 3 varian nilai saja yaitu 1,3 dan 4.

## 2.2 Dasar Teori

### 2.2.1 Data Mining

Data mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstrasi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai database besar (Santosa, 2016). Istilah data mining memiliki hakikat sebagai disiplin ilmu yang tujuan utamanya adalah untuk menemukan, menggali, atau menambang pengetahuan dari data atau informasi yang dimiliki. Data mining atau *Knowledge Discovery in Database* (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data, historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar.

#### 1) Metode Pelatihan

Secara garis besar metode pelatihan yang digunakan dalam teknik data mining dibedakan ke dalam dua pendekatan yaitu (Santosa, 2016):

##### a. *Unsupervised Learning*

Metode ini diterapkan tanpa adanya latihan (*training*) dan tanpa ada guru (*teacher*). Guru di sini adalah label dari data.

##### b. *Supervised Learning*

Metode belajar dengan adanya latihan dan pelatih. Dalam pendekatan ini untuk menemukan fungsi keputusan, fungsi pemisah atau fungsi regresi.

#### 2) Pengelompokan Data Mining

Ada beberapa teknik yang dimiliki *data mining* berdasarkan tugas yang bisa dilakukan, yaitu (Santosa, 2016):

a. Deskripsi

Para peneliti biasanya mencoba menemukan cara untuk mendiskripsikan pola dan tren yang tersembunyi dalam data.

b. Estimasi

Estimasi mirip dengan klasifikasi, kecuali variabel tujuan yang lebih kearah numeric dari pada kategori.

c. Prediksi

Prediksi memiliki kemiripan dengan estimasi dan klasifikasi. Hanya saja, prediksi hasilnya menunjukkan sesuatu yang belum terjadi (mungkin terjadi di masa depan).

d. Klasifikasi

Dalam klasifikasi variabel, tujuan bersifat kategorik. Misalnya, Pengklasifikasian pendapatan dalam tiga kelas, yaitu pendapatan tinggi, pendapatan sedang, dan pendapatan rendah.

e. *Clustering*

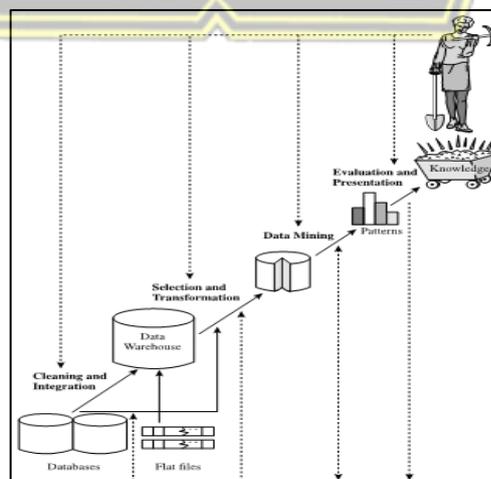
*Clustering* lebih kearah pengelompokan *record*, pengamatan, atau kasus dalam kelas yang memiliki kemiripan.

f. Asosiasi

Mengidentifikasi hubungan antara berbagai peristiwa yang terjadi pada satu waktu.

3) Tahap-tahap Data Mining

Sebagai suatu rangkaian proses, data mining dapat dibagi menjadi beberapa tahap proses yang diilustrasikan pada gambar 2.1.



Gambar 2.1. Tahap-Tahap Data Mining

Tahap-tahap data mining adalah sebagai berikut (Santosa, 2016):

a. Pembersihan data (*data cleaning*)

Pembersihan data merupakan proses menghasilkan *noise* dan data yang tidak konsisten atau data tidak relevan.

b. Integrasi data (*data integration*)

Integrasi data merupakan penggabungan data dari berbagai *database* ke dalam suatu *database* baru.

c. Seleksi data (*data selection*)

Data yang ada pada *database* sering kali tidak semuanya dipakai, oleh karena itu hanya data yang sesuai untuk dianalisis yang akan diambil dari *database*.

d. Transformasi data (*data transformasi*)

Data diubah atau digabung ke dalam format yang sesuai untuk diproses dalam *data mining*.

e. Proses *mining*

Merupakan suatu proses utama saat metode diterapkan untuk menemukan pengetahuan berharga dan tersembunyi dari data.

f. Evaluasi pola (*pattern evaluation*)

Untuk mengidentifikasi pola-pola menarik ke dalam *knowledge based* yang ditemukan.

g. Presentasi pengetahuan (*knowledge presentation*)

Merupakan visualisasi dan penyajian pengetahuan mengenai metode yang digunakan untuk memperoleh pengetahuan.

### 2.2.2 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek (Han & Kamber, 2016).

Dalam pengklasifikasian data terdapat 2 proses yang dilakukan yaitu:

1) Proses *training*

Pada proses *training* digunakan *training set* yang telah diketahui label-labelnya untuk membangun model atau fungsi.

## 2) Proses *testing*

Untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses training, maka digunakan data yang disebut dengan testing set untuk memprediksi label-labelnya.

Proses klasifikasi didasarkan pada empat komponen yaitu:

- a) Kelas : Variabel dependen berupa kategorikal yang merepresentasikan “label” yang terdapat pada objek. Contohnya: resiko penyakit jantung, resiko kredit, customer loyalty, jenis gempa.
- b) *Predictor* : Variabel independen yang direpresentasikan oleh karakteristik (atribut) data. Contohnya: merokok, minum alkohol, tekanan darah, tabungan, aset, gaji.
- c) *Training dataset* : Satu set data yang berisi nilai dari kedua komponen di atas yang digunakan untuk menentukan kelas yang cocok berdasarkan *predictor*.
- d) *Testing dataset* : Berisi data baru yang akan diklasifikasikan oleh model yang telah dibuat.

### 2.2.3 KNN

Algoritma *K-Nearest Neighbor* (KNN) adalah algoritma yang digunakan untuk mengklasifikasikan suatu objek, berdasarkan kereta data pelatihan yang lebih dekat dengan objek. Persyaratan  $K$  seharusnya tidak lebih besar dari jumlah data pelatihan, dan nilai  $K$  harus aneh dan lebih dari satu (Rivki & Bachtiar, 2017).

Algoritma *K-Nearest Neighbor* menggunakan klasifikasi lapangan kerja sebagai nilai prediksi konsultasi konsultasi baru. Algoritma ini sederhana, berfungsi berdasarkan jarak terpendek dari kueri instance ke pelatihan sampling untuk menentukan ketetanggan.

*K-Nearest Neighbor* (K-NN) memiliki kelebihan yang dapat menghasilkan data yang kuat atau jelas dan efektif jika digunakan dalam data besar. Tetangga terdekat juga memiliki kerugian, yang membutuhkan nilai  $K$ , jarak dari data uji tidak jelas dengan jarak yang digunakan, untuk mendapatkan hasil terbaik, Anda harus menggunakan semua atribut atau hanya atribut yang ditentukan. Metode *K-Nearest Neighbor* (KNN) cukup sederhana, tidak ada asumsi sehubungan dengan

distribusi data dan mudah dipublikasikan. Pemilihan nilai K (nomor data / tetangga) ditentukan oleh peneliti. Tahapan menghitung metode KNN sebagai berikut:

1. Memastikan parameter K ( jumlah tetangga terdekat )
2. Menghitung kuadrat *euclidean distance* (*query instance*) masing- masing objek terhadap data sampel.
3. Kemudian mengurutkan objek- objek tersebut kedalam kelompok yang mempunyai jarak *euclidean* terkecil.
4. Menggabungkan kategori y ( klasifikasi *Nearest Neighbor*)

Dengan menggunakan kategori *Nearest Neighbor* yang paling mayoritas, maka dapat diprediksi nilai *query instance* yang telah dihitung.

#### 2.2.4 UML (*Unified Modelling Language*)

*Unified Modelling Language* (UML) adalah sebuah bahasa yang telah menjadi standar dalam industri untuk visualisasi, merancang dan mendokumentasikan sistem piranti lunak. UML menawarkan sebuah standar untuk merancang model sebuah sistem. Dengan menggunakan UML kita dapat membuat model untuk semua jenis aplikasi piranti lunak, dimana aplikasi tersebut dapat berjalan pada piranti keras, sistem operasi dan jaringan apapun, serta ditulis dalam bahasa pemrograman apapun. Tetapi karena UML juga menggunakan *class* dan *operation* dalam konsep dasarnya, maka ia lebih cocok untuk penulisan piranti lunak dalam bahasa-bahasa berorientasi objek seperti C++, Java, C# atau VB.NET. Walaupun demikian, UML tetap dapat digunakan untuk modeling aplikasi prosedural dalam VB atau C (Munawar, 2018).

Seperti bahasa-bahasa lainnya, UML mendefinisikan notasi dan *syntax* atau semantik. Notasi UML merupakan sekumpulan bentuk khusus untuk menggambarkan berbagai diagram piranti lunak. Setiap bentuk memiliki makna tertentu, dan UML *syntax* mendefinisikan bagaimana bentuk-bentuk tersebut dapat dikombinasikan. Notasi UML terutama diturunkan dari 3 notasi yang telah ada sebelumnya: Grady Booch OOD (*Object-Oriented Design*), Jim Rumbaugh OMT (*Object Modeling Technique*), dan Ivar Jacobson OOSE (*Object-Oriented Software Engineering*). Sejarah UML sendiri cukup panjang. Sampai era tahun 1990 seperti

kita ketahui puluhan metodologi pemodelan berorientasi objek telah bermunculan di dunia.. Masa itu terkenal dengan masa perang metodologi (*method war*) dalam pendesainan berorientasi objek. Masing-masing metodologi membawa notasi sendiri-sendiri, yang mengakibatkan timbul masalah baru apabila kita bekerjasama dengan group atau perusahaan lain yang menggunakan metodologi yang berlainan (Muslihudin, 2017)

### 1) *Use Case Diagram*

*Use case diagram* menggambarkan fungsionalitas yang diharapkan dari sebuah sistem. Yang ditekankan adalah “apa” yang diperbuat sistem, dan bukan “bagaimana”. Sebuah *use case* merepresentasikan sebuah interaksi antara aktor dengan sistem. *Use case* merupakan sebuah pekerjaan tertentu, misalnya *Login* ke sistem, meng-*create* sebuah daftar belanja, dan sebagainya. Seorang/sebuah aktor adalah sebuah entitas manusia atau mesin yang berinteraksi dengan sistem untuk melakukan pekerjaan-pekerjaan tertentu (Munawar, 2018).

*Use case diagram* dapat sangat membantu bila sedang menyusun *requirement*, mengkomunikasikan rancangan dengan klien, dan merancang *test case* untuk semua *feature*. Sebuah *use case* dapat meng-*include* fungsionalitas *use case* lain sebagai bagian dari proses dalam dirinya. Secara umum diasumsikan bahwa *use case* yang di-*include* akan dipanggil setiap kali *use case* yang meng-*include* dieksekusi secara normal. Sebuah *use case* dapat di-*include* oleh lebih dari satu *use case* lain, sehingga duplikasi fungsionalitas dapat dihindari dengan cara menarik keluar fungsionalitas yang *common*. Sebuah *use case* juga dapat meng-*extend use case* lain dengan *behavior*-nya sendiri. Sementara hubungan generalisasi antar *use case* menunjukkan bahwa *use case* yang satu merupakan spesialisasi dari yang lain.

Tabel 2. 1. Simbol *Use Case*

SIMBOL	NAMA	KETERANGAN
	<i>Actor</i>	Menspesifikasikan himpunan peran yang pengguna mainkan ketika berinteraksi dengan <i>use case</i> .

SIMBOL	NAMA	KETERANGAN
	<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri ( <i>independent</i> ) akan mempengaruhi elemen yang bergantung padanya elemen yang tidak mandiri ( <i>independent</i> ).
	<i>Generalization</i>	Hubungan dimana objek anak ( <i>descendent</i> ) berbagi perilaku dan struktur data dari objek yang ada di atasnya objek induk ( <i>ancestor</i> ).
	<i>Include</i>	Menspesifikasikan bahwa <i>use case</i> sumber secara eksplisit.
	<i>Extend</i>	Menspesifikasikan bahwa <i>use case</i> target memperluas perilaku dari <i>use case</i> sumber pada suatu titik yang diberikan.
	<i>Association</i>	Apa yang menghubungkan antara objek satu dengan objek lainnya.
	<i>System</i>	Menspesifikasikan paket yang menampilkan sistem secara terbatas.
	<i>Use Case</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor
	<i>Collaboration</i>	Interaksi aturan-aturan dan elemen lain yang bekerja sama untuk menyediakan perilaku yang lebih besar dari jumlah dan elemen-elemennya ( <i>sinergi</i> ).
	<i>Note</i>	Elemen fisik yang eksis saat aplikasi dijalankan dan mencerminkan suatu sumber daya komputasi

Tabel 2.1 merupakan simbol yang dipakai dalam pembuatan *use case* yang terdiri dari 10 simbol.

## 2) *Class Diagram*

*Class* adalah sebuah spesifikasi yang jika diinstansiasi akan menghasilkan sebuah objek dan merupakan inti dari pengembangan dan desain berorientasi objek. *Class* menggambarkan keadaan (atribut atau properti) suatu sistem, sekaligus menawarkan layanan untuk memanipulasi keadaan tersebut (metoda/fungsi). *Class diagram* menggambarkan struktur dan deskripsi *class*, *package* dan objek beserta hubungan satu sama lain seperti *containment*, pewarisan, asosiasi, dan lain-lain (Munawar, 2018).

*Class* dapat merupakan implementasi dari sebuah *interface*, yaitu *class* abstrak yang hanya memiliki metoda. *Interface* tidak dapat langsung diinstansiasikan, tetapi harus diimplementasikan dahulu menjadi sebuah *class*. Dengan demikian *interface* mendukung resolusi metoda pada saat *run-time*. Hubungan antar *class* yaitu (Munawar, 2018) :

### a. Asosiasi

Hubungan statis antar *class*. Umumnya menggambarkan *class* yang memiliki atribut berupa *class* lain, atau *class* yang harus mengetahui eksistensi *class* lain. Panah *navigability* menunjukkan arah query antar *class*.

### b. Agregasi

Hubungan yang menyatakan bagian (“terdiri atas..”).

### c. Pewarisan

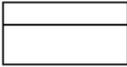
Hubungan hirarkis antar *class*. *Class* dapat diturunkan dari *class* lain dan mewarisi semua atribut dan metoda *class* asalnya dan menambahkan fungsionalitas baru, sehingga ia disebut anak dari *class* yang diwarisinya. Kebalikan dari pewarisan adalah generalisasi.

### d. Hubungan Dinamis

Rangkaian pesan (message) yang di-passing dari satu *class* kepada *class* lain.

Tabel 2.2. Simbol *Class Diagram*

SIMBOL	NAMA	KETERANGAN
	<i>Generalization</i>	Hubungan dimana objek anak ( <i>descendent</i> ) berbagi

		perilaku dan struktur data dari objek yang ada di atasnya objek induk ( <i>ancestor</i> ).
	<i>Nary Association</i>	Upaya untuk menghindari asosiasi dengan lebih dari 2 objek.
	<i>Class</i>	Himpunan dari objek-objek yang berbagi atribut serta operasi yang sama.
	<i>Collaboration</i>	Deskripsi dari urutan aksi-aksi yang ditampilkan sistem yang menghasilkan suatu hasil yang terukur bagi suatu aktor
	<i>Realization</i>	Operasi yang benar-benar dilakukan oleh suatu objek.
	<i>Dependency</i>	Hubungan dimana perubahan yang terjadi pada suatu elemen mandiri ( <i>independent</i> ) akan memengaruhi elemen yang bergantung padanya elemen yang tidak mandiri
	<i>Association</i>	Apa yang menghubungkan antara objek satu dengan objek lainnya

### 3) Activity Diagram

*Activity diagram* menggambarkan berbagai alir aktivitas dalam sistem yang sedang dirancang, bagaimana masing-masing alir berawal, *decision* yang mungkin terjadi, dan bagaimana mereka berakhir. *Activity diagram* juga dapat menggambarkan proses paralel yang mungkin terjadi pada beberapa eksekusi (Munawar, 2018).

*Activity diagram* merupakan *state diagram* khusus, di mana sebagian besar *state* adalah *action* dan sebagian besar transisi di-*trigger* oleh selesainya *state* sebelumnya (*internal processing*). Oleh karena itu *activity diagram* tidak menggambarkan behaviour internal sebuah sistem (dan interaksi antar subsistem)

secara eksak, tetapi lebih menggambarkan proses-proses dan jalur-jalur aktivitas dari level atas secara umum.

Sebuah aktivitas dapat direalisasikan oleh satu *use case* atau lebih. Aktivitas menggambarkan proses yang berjalan, sementara *use case* menggambarkan bagaimana aktor menggunakan sistem untuk melakukan aktivitas. Sama seperti *state*, standar UML menggunakan segiempat dengan sudut membulat untuk menggambarkan aktivitas. *Decision* digunakan untuk menggambarkan behaviour pada kondisi tertentu. Untuk mengilustrasikan proses-proses paralel (*fork* dan *join*) digunakan titik sinkronisasi yang dapat berupa titik, garis horizontal atau vertikal. *Activity diagram* dapat dibagi menjadi beberapa *object swimlane* untuk menggambarkan objek mana yang bertanggung jawab untuk aktivitas tertentu.

Tabel 2. 3. Simbol *Activity Diagram*

SIMBOL	NAMA	KETERANGAN
	State	Kondisi yang mungkin dialami oleh suatu obyek.
	Note	Note digunakan untuk memberikan keterangan atau komentar
	Aktivitas	Perilaku obyek yang dilakukan saat obyek berada dalam <i>state</i> tertentu.
	Start State	<i>Start state</i> digunakan untuk memulai diagram <i>statechart</i> .
	End State	End start digunakan untuk mengakhiri diagram.
	Decision	<i>Decision</i> digunakan sebagai pilihan untuk pengambilan keputusan.
	Penggabungan / Join	digunakan untuk split dan join.pada saat diagram akanmembagi 2, bar ini akan ditambahkan. dan

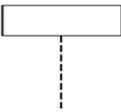
SIMBOL	NAMA	KETERANGAN
		sebelum diagram digabung menjadi satu, sebagai join.
—	Asosiasi ( <i>association</i> )	Apa yang menghubungkan antara objek satu dengan objek yang lainnya.

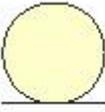
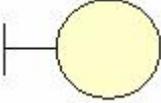
#### 4) Sequence Diagram

*Sequence diagram* menggambarkan interaksi antar objek di dalam dan di sekitar sistem (termasuk pengguna, *display*, dan sebagainya) berupa *message* yang digambarkan terhadap waktu. *Sequence diagram* terdiri atas dimensi vertikal (waktu) dan dimensi horizontal (objek-objek yang terkait). *Sequence diagram* biasa digunakan untuk menggambarkan skenario atau rangkaian langkah-langkah yang dilakukan sebagai respons dari sebuah *event* untuk menghasilkan *output* tertentu. Diawali dari apa yang men-*trigger* aktivitas tersebut, proses dan perubahan apa saja yang terjadi secara internal dan *output* apa yang dihasilkan (Munawar, 2018).

Masing-masing objek, termasuk aktor, memiliki *lifeline* vertikal. *Message* digambarkan sebagai garis berpanah dari satu objek ke objek lainnya. Pada fase desain berikutnya, *message* akan dipetakan menjadi operasi/metoda dari *class*. *Activation bar* menunjukkan lamanya eksekusi sebuah proses, biasanya diawali dengan diterimanya sebuah *message*. Tabel *sequence diagram* diperlihatkan seperti pada Tabel 2.4.

Tabel 2. 4. Simbol *Sequence Diagram*

SIMBOL	NAMA	KETERANGAN
	<i>LifeLine</i>	Menggambarkan objek <i>entity</i> , antarmuka yang saling berinteraksi.
	<i>Message</i>	Menggambarkan pengiriman pesan

SIMBOL	NAMA	KETERANGAN
	<i>Actor</i>	Menggambarkan orang yang sedang berinteraksi dengan sistem
	<i>Entity Class</i>	Menggambarkan hubungan kegiatan yang akan dilakukan
	<i>Boundary Class</i>	Menggambarkan sebuah penggambaran dari form
	<i>Control Class</i>	Menggambarkan penghubung antara <i>boundary</i> dengan tabel

### 2.2.5 PHP

PHP/FI merupakan akronim dari *Personal Home Page/Forms Interpreter*. Pada awal penyusunan, PHP/FI hanya mempunyai fungsi dasar dari PHP yang ada sekarang ini karena ketika pertama kali dibuat dengan menggunakan Perl maka PHP/FI juga mempunyai susunan dan karakter pemrograman yang sama dengan Perl. PHP/FI merupakan akronim dari *Personal Home Page/Forms Interpreter* (Abdulloh, 2018). Pada awal penyusunan, PHP/FI hanya mempunyai fungsi dasar dari PHP yang ada sekarang ini karena ketika pertama kali dibuat dengan menggunakan Perl maka PHP/FI juga mempunyai susunan dan karakter pemrograman yang sama dengan Perl. Berapa kelebihan PHP dari bahasa pemrograman *web*, antara lain (Setiawan, 2018):

1. Bahasa pemrograman PHP adalah sebuah bahasa *script* yang tidak melakukan sebuah kompilasi dalam penggunaannya.
2. *Web Server* yang mendukung PHP dapat ditemukan dimana-mana, mulai dari apache, IIS, Lighttpd, hingga Xitami dengan konfigurasi yang relatif mudah.
3. Dalam sisi pengembangan lebih mudah, karena banyaknya milis-milis dan developer yang siap membantu dalam pengembangan.

4. Dalam sisi pemahaman, PHP adalah bahasa *scripting* yang paling mudah karena memiliki referensi yang banyak.
5. PHP adalah bahasa *open source* yang dapat digunakan di berbagai mesin (Linux, Unix, Macintosh, Windows) dan dapat dijalankan secara *runtime* melalui *console* serta juga dapat menjalankan perintah-perintah sistem.

### 2.2.6 MySQL

*MySQL* adalah sebuah program database *server* yang mampu menerima dan mengirimkan datanya dengan sangat cepat, *multi user* serta menggunakan perintah dasar SQL (*Structured Query Language*). *MySQL* merupakan sebuah database yang gratis artinya bebas untuk digunakan untuk keperluan pribadi atau usaha tanpa membeli atau membayar lisensi. *MySQL* pertama kali dirintis oleh seorang programmer database bernama Michael Widenius. Selain database server, *MySQL* juga merupakan program yang dapat mengakses suatu database *MySQL* yang berposisi sebagai *server*, jadi *MySQL* adalah sebuah database yang dapat digunakan baik sebagai *client* maupun *server*. Selain itu *MySQL* mempunyai beberapa kelebihan dibanding database lain, yaitu (Nugroho, 2015):

1. *MySQL* sebagai *Database Management Systems (DBMS)*.
2. *MySQL* sebagai *Relation Database Management Systems (RDBMS)*.
3. *MySQL* adalah sebuah software database yang *open source* yang artinya program ini bersifat gratis atau bebas digunakan oleh siapa saja tanpa harus membeli dan membayar lisensi kepada pembuatnya.
4. *MySQL* merupakan sebuah database *server* sehingga dapat dihubungkan ke media internet yang dapat diakses dari jauh.
5. *MySQL* merupakan sebuah database *client*. *MySQL* dapat melakukan *query* yang mengakses database pada *server*.
6. *MySQL* mampu menerima *query* yang bertumpuk dalam satu permintaan atau yang disebut *Multi Threading*.
7. *MySQL* merupakan database yang mampu menyimoan data berkapasitas sangat besar hingga berukuran *gigabyte* sekalipun.
8. *MySQL* didukung oleh driver ODBC sehingga dapat diakses menggunakan aplikasi apa saja termasuk visual seperti Delphi maupun Visual Basic.

9. MySQL adalah database menggunakan enkripsi *password*.
10. MySQL merupakan *server* database yang *multi user* yang artinya tidak hanya digunakan oleh satu orang saja akan tetapi dapat digunakan oleh banyak orang.
11. MySQL dapat menciptakan lebih dari 16 kunci per tabel dan dalam satu kunci memungkinkan berisi belasan *field* (kolom).
12. MySQL mendukung *field* yang dijadikan sebagai kunci primer dan kunci unik (*unique*).
13. MySQL didukung oleh sebuah komponen C atau perl API sehingga dapat diakses melalui sebuah program aplikasi yang berada dibawah protokol internet berupa web.
14. MySQL memiliki kecepatan dalam pembuatan tabel maupun pengupdaten tabel.
15. MySQL menggunakan bahasa permintaan yang standar yaitu SQL yaitu sebuah bahasa permintaan yang distandarkan pada beberapa database *server* seperti Oracle, PostGreSQL dan lain-lain.

### 2.2.7 Confusion Matrix

*Confusion matrix* merupakan suatu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. *Confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem untuk diukur keakuratannya. *Confusion matrix* juga merupakan salah satu cara dalam melakukan visualisasi terhadap hasil pembelajaran sistem, visualisasi yang ditampilkan memuat dua kategori atau lebih (Ting, 2017). Tabel 2.5 merupakan contoh hasil *confusion matrix* klasifikasi dua kelas.

Tabel 2. 5. *Confusion Matrix*

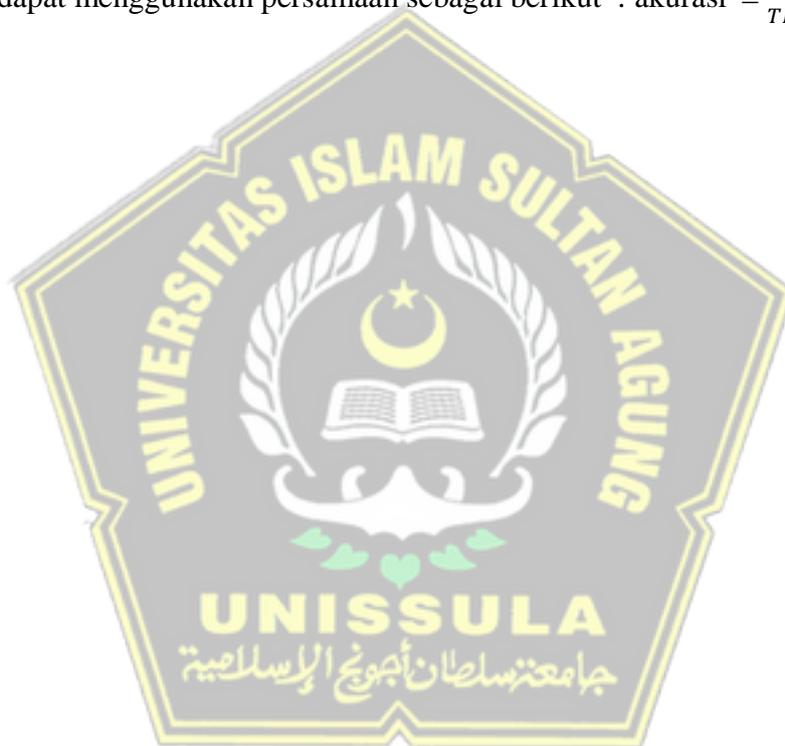
		Kelas Sebenarnya	
		1	2
Kelas Klasifikasi	1	<i>True Positive</i>	<i>False Negative</i>
	2	<i>False Positive</i>	<i>True Negative</i>

Keterangan:

- a. *True Positive* (TP) : merupakan jumlah data dengan kelas positif yang diklasifikasikan positif.

- b. *True Negative* (TN) : merupakan jumlah data dengan kelas negatif yang diklasifikasikan negatif
- c. *False Positive* (FP) : merupakan jumlah data dengan kelas positif yang diklasifikasikan negatif.
- d. *False Negative* (FN) : merupakan jumlah data dengan kelas negatif yang diklasifikasikan positif.

Perhitungan Akurasi yang dilakukan oleh *confusion matrix* berdasarkan tabel di atas dapat menggunakan persamaan sebagai berikut : akurasi =  $\frac{TP+TN}{TP+FN+FP+FN} \times 100 \%$ .



## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Metode Penelitian**

Metode penelitian merupakan suatu cara atau prosedur yang digunakan untuk mengumpulkan data, dengan perantara teknik tertentu. Dalam penelitian ini, akan menggunakan beberapa metode penelitian yaitu :

##### **3.1.1 Metode Pengumpulan Data**

Metode yang digunakan dalam proses pengumpulan data sebagai bahan pembuatan sistem adalah metode studi pustaka. Metode studi pustaka melakukan pengumpulan data dengan cara mempelajari literatur, laporan, maupun jurnal yang berhubungan dengan klasifikasi penyakit kardiovaskular dengan algoritma KNN

##### **3.1.2 Metode Pengembangan Sistem**

Metode yang digunakan untuk pengembangan sistem adalah *prototype*. Tahap-tahap pengembangannya adalah (Pressman, 2012):

###### 1. Komunikasi

Tahap ini melakukan identifikasi permasalahan-permasalahan serta informasi-informasi lain yang diperlukan dalam penerapan data mining untuk klasifikasi penyakit kardiovaskular menggunakan algoritma KNN.

###### 2. Perencanaan

Tahap ini mengidentifikasi kebutuhan sistem yaitu kebutuhan perangkat lunak (*software*), kebutuhan perangkat keras (*hardware*).

###### 3. Pemodelan

Tahap ini dilakukan perancangan sistem dengan menggunakan UML, perancangan database dan dibuat suatu desain antar muka sistem.

###### 4. Kontruksi

Tahap ini menerapkan data mining untuk klasifikasi penyakit kardiovaskular menggunakan algoritma KNN dengan PHP dan MySQL.

###### 5. Penyerahan

Tahapan ini dibutuhkan untuk mendapatkan *feedback* dari pengguna, sebagai hasil evaluasi dari penerapan data mining untuk klasifikasi penyakit kardiovaskular menggunakan algoritma KNN.

### 3.2 Identifikasi Masalah

Permasalahan yang dihadapi dalam melakukan klasifikasi penyakit kardiovaskular yaitu:

1. Penyakit kardiovaskular merupakan pembunuh nomor satu di dunia.
2. Tingginya faktor kematian akibat penyakit kardiovaskular dapat dicegah dan ditekan faktor resikonya jika masyarakat memiliki pengetahuan tentang gejala penyakit kardiovaskular.
3. Belum adanya sistem yang dapat digunakan untuk mendeteksi seseorang apakah menderita kardiovaskular atau tidak

Berdasarkan permasalahan di atas, penulis melakukan suatu penelitian yang akan mempermudah dalam menentukan klasifikasi mengenai klasifikasi kardiovaskular dengan menggunakan metode KNN.

### 3.3 Identifikasi Perangkat Keras (*Hardware*)

Perangkat keras yang digunakan dalam pembuatan klasifikasi penyakit kardiovaskular menggunakan algoritma KNN yaitu laptop dengan spesifikasi sebagai berikut:

- 1) Processor Intel (R) Pentium IV Core I3
- 2) Memori 4 GB DDR3
- 3) HDD 500 Gb
- 4) LCD 14"

### 3.4 Identifikasi Kebutuhan Perangkat Lunak (*Software*)

Perangkat lunak yang digunakan dalam pengembangan klasifikasi penyakit kardiovaskular menggunakan algoritma KNN yaitu :

- 1) Sistem Operasi : Microsoft Window 10
- 2) Editor Web : Dreamweaver CS5
- 3) HTTP Server : Apache + PHP
- 4) Database : MySQL

5) Browser : Mozilla Firefox

### 3.5 Analisis Kebutuhan Data

Data yang digunakan dalam penelitian ini yaitu data sekunder. Data sekunder dari penyakit kardiovaskular didapat dari website kaggle data set sebanyak 1.000 data.

Tabel 3.1. Dataset Penyakit Kardiovaskular

<i>age</i>	<i>Gender</i>	<i>Height</i>	<i>weight</i>	<i>Systolic</i>	<i>diastolic</i>	<i>Cholesterol</i>	<i>gluc</i>	<i>smoke</i>	<i>alco</i>	<i>active</i>	<i>cardio</i>
62	1	155	69	130	80	2	2	0	0	1	0
40	1	163	71	110	70	1	1	0	0	1	1
60	1	165	70	120	80	1	1	0	0	1	0
40	2	165	85	120	80	1	1	1	1	1	0
64	1	155	62	120	80	1	1	0	0	1	0
58	1	164	61	100	70	1	1	0	0	1	0
50	1	162	72	100	70	1	1	0	0	1	0
62	1	161	93	140	90	2	1	0	0	1	1
58	1	157	68	110	80	1	1	0	0	1	0
61	1	158	65	140	90	3	1	0	0	1	1
46	1	153	57	120	80	3	2	0	0	1	1
56	2	169	82	130	90	1	1	0	0	1	1
63	2	150	56	130	80	2	1	0	0	1	1
62	1	176	72	130	90	1	3	0	0	1	1
56	1	154	85	160	100	3	1	0	0	1	1
65	1	159	60	120	80	1	1	0	0	1	1
55	2	146	45	120	80	1	1	0	0	1	0
54	1	165	90	140	90	1	1	0	0	1	1
55	2	167	65	120	80	1	1	0	0	0	0
52	1	154	70	110	70	1	1	0	0	1	0
64	2	170	74	140	1000	1	1	0	0	1	1
63	1	150	56	130	80	1	1	0	0	1	1
56	1	162	88	160	90	2	1	0	0	1	1
60	2	167	74	110	70	1	1	0	0	0	0
54	2	167	95	110	70	1	1	0	0	1	0
65	1	163	103	140	100	1	1	0	0	1	1
48	1	162	80	120	80	1	1	0	0	1	0
64	1	151	63	120	80	2	2	0	0	0	1
44	1	161	55	100	60	1	1	0	0	0	0
40	2	183	91	130	80	2	1	0	0	1	1

41	2	160	62	120	80	1	1	0	0	1	0
58	1	164	77	110	80	1	2	0	1	1	0

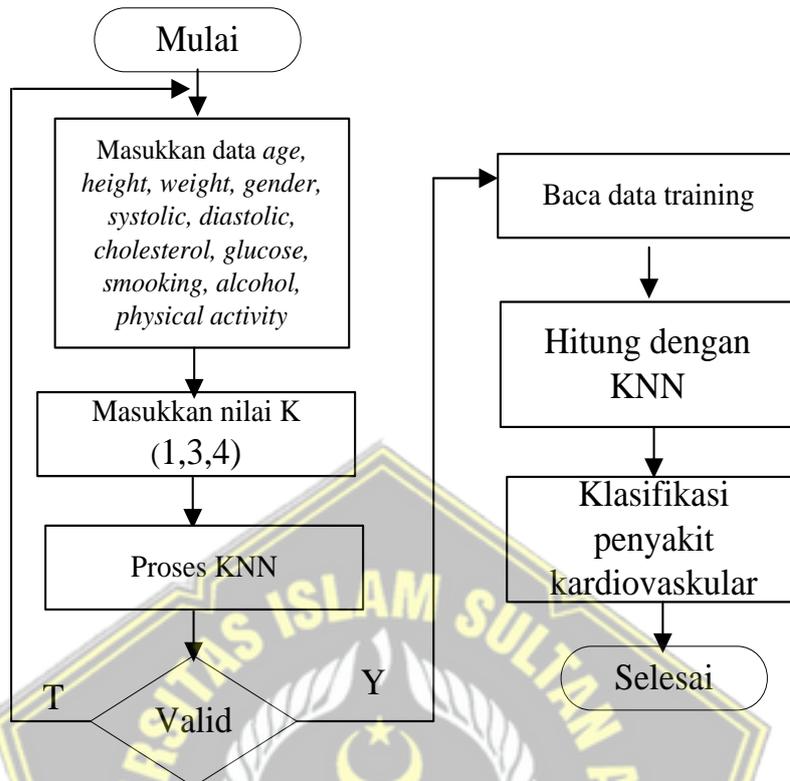
Dataset penyakit kardiovaskular terdapat 11 atribut dan 1 label dengan tipe data integer dan binary. Dari total 1.000 data penyakit kardiovaskular, dinyatakan 523 orang tidak terkena penyakit kardiovaskular dan 477 orang lainnya terkena penyakit kardiovaskular. Penjelasan mengenai atribut dataset penyakit kardiovaskular dijabarkan pada tabel 3.2.

Tabel 3.2. Atribut Dataset Penyakit Kardiovaskular

No	Atribut	Tipe Data	Keterangan
1.	<i>Age</i>	Integer	Usia yang dinyatakan dalam tahun
2.	<i>Gender</i>	Binary	Jenis kelamin (1 : perempuan / 0 : laki-laki.
3.	<i>Height</i>	Integer	Tinggi/panjang badan yang dinyatakan dalam Cm
3.	<i>Weight</i>	Integer	Berat badan yang dinyatakan dalam hitungan Kg.
4.	<i>Systolic</i>	Integer	Tekanan darah atas (sistolik) dalam satuan mmHg.
5.	<i>Diastolic</i>	Integer	Tekanan darah bawah (diastolik) dalam satuan mmHg.
6.	<i>Cholesterol</i>	Integer	Kadar kolesterol dimana jika normal diberi nilai 1, diatas normal diberi nilai 2 sedangkan jika jauh diatas normal diberi nilai 3
7.	<i>Glucose</i>	Integer	Kadar glukosa dimana jika normal diberi nilai 1, diatas normal diberi nilai 2 sedangkan jika jauh diatas normal diberi nilai 3
8	<i>Smoking</i>	Binary	Kondisi merokok atau tidak, jika tidak diberi nilai 0 sedangkan jika ya diberi nilai 1
9	<i>Alcohol Intake</i>	Binary	Konsumsi alkohol atau tidak, jika tidak diberi nilai 0 sedangkan jika ya diberi nilai 1
10	<i>Physical activity</i>	Binary	Melakukan olahraga atau tidak, jika tidak diberi nilai 0 sedangkan jika ya diberi nilai 1
11	<i>Cardiovascular</i>	Binary	Terkena penyakit kardio atau tidak, jika tidak diberi nilai 0 sedangkan jika ya diberi nilai 1

### 3.6 Flowchart Sistem

*Flowchart* penerapan data mining untuk klasifikasi penyakit kardiovaskular menggunakan algoritma KNN diperlihatkan seperti gambar 3.1.



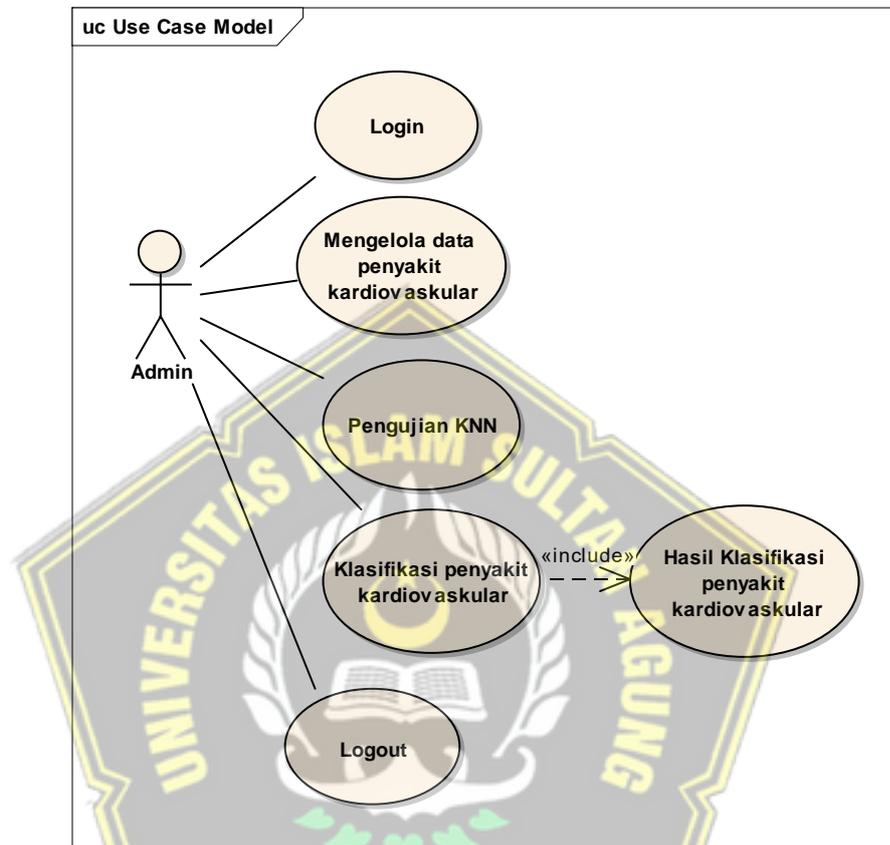
Gambar 3.1. Flowchart Sistem

Proses klasifikasi penyakit kardiovaskular menggunakan KNN yaitu

1. Penentuan nilai K adalah langkah paling awal dan paling krusial dalam algoritma KNN melakukan proses klasifikasi penyakit kardiovaskular. Dimana misalkan nilai K diinisialisasi dengan nilai 1,3 dan 4, maka perhitungan jarak dengan algoritma tertentu pada KNN akan dari nilai n terkecil.
2. Menentukan variabel yang digunakan dalam klasifikasi penyakit kardiovaskular menggunakan KNN yaitu *age, height, weight, gender, systolic, diastolic, cholesterol, glucose, smoking, alcohol, physical activity*.
3. Baca data *training* sebanyak 975 data.
4. Pengujian data penyakit kardiovaskular dengan memasukkan nilai K.
5. Hitung klasifikasi penyakit kardiovaskular menggunakan KNN dengan rumus rumus *euclidean distance* :  $D = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$
6. Tampilkan hasil klasifikasi penyakit kardiovaskular

### 3.7 Use Case Diagram

Use case diagram klasifikasi penyakit kardiovaskular menggunakan algoritma KNN diperlihatkan seperti gambar 3.2.



Gambar 3.2. Use Case Diagram

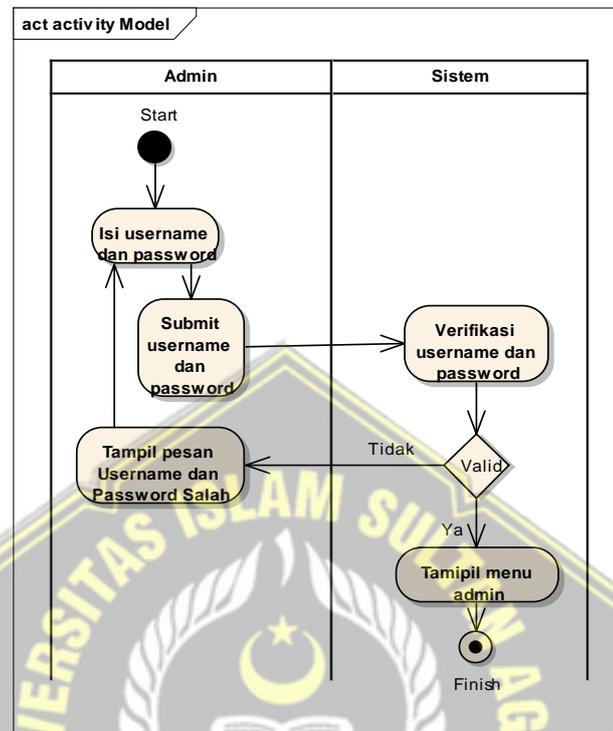
Gambar 3.2 menjelaskan *Admin* melakukan *Login* ke sistem dengan memasukkan *username* dan *password* kemudian mengelola data penyakit kardiovaskular, melakukan pengujian penyakit kardiovaskular dengan metode KNN, melakukan klasifikasi penyakit kardiovaskular dan mendapatkan hasil klasifikasi penyakit kardiovaskular menggunakan metode KNN kemudian *logout*.

### 3.8 Activity Diagram

#### 1. Login

*Login* dimulai dengan *Admin* mengisi *username* dan *password* kemudian sistem melakukan verifikasi *username* dan *password*, jika *username* dan *password* akan tampil menu *Admin* sedangkan jika pengisian *username* dan *password* tidak

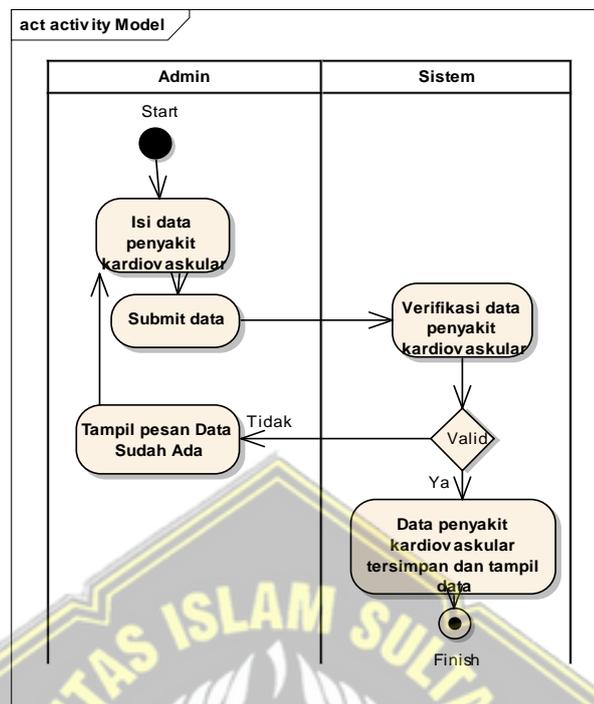
*valid* akan ditampilkan pesan *username* dan *password* salah dan *Admin* melakukan pengisian *username* dan *password* lagi.



Gambar 3.3. Activity Diagram Login

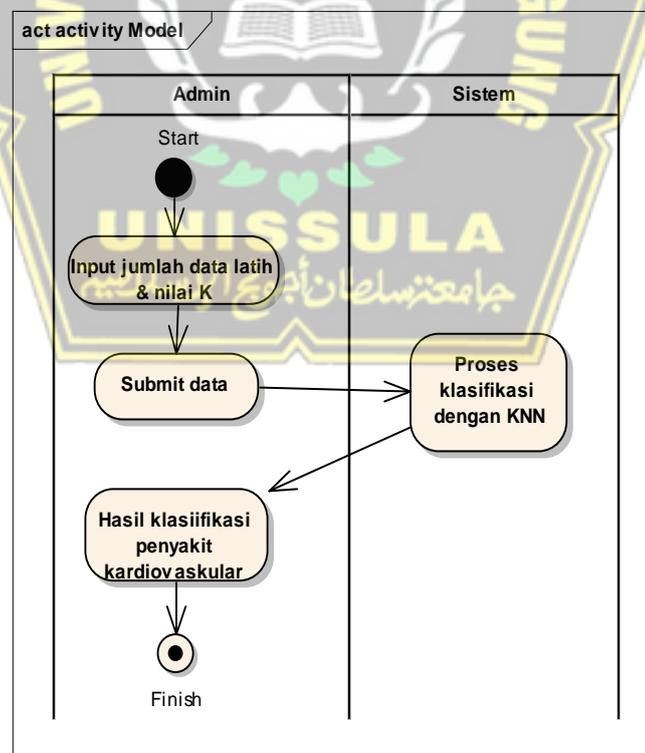
## 2. Mengelola Data Penyakit Kardiovaskular

Mengelola data penyakit kardiovaskular dimulai dengan *Admin* mengisi data penyakit kardiovaskular dan sistem akan melakukan verifikasi dari data yang dimasukkan, jika data penyakit kardiovaskular yang dimasukkan *valid*, maka data akan disimpan di tabel data.



Gambar 3. 4. Activity Diagram Mengelola Kardiovaskular

### 3. Pengujian KNN

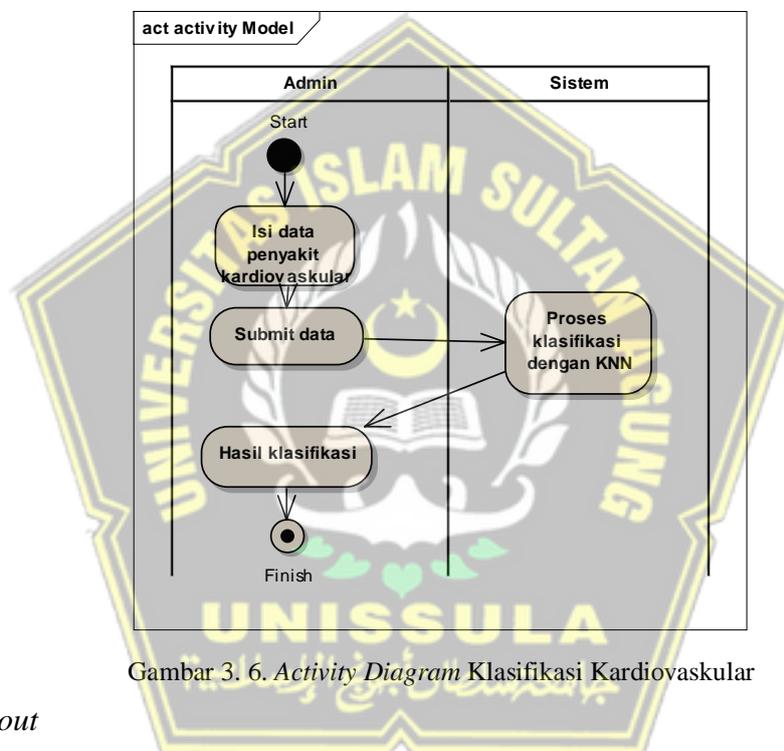


Gambar 3. 5. Activity Diagram Pengujian KNN

Pengujian KNN dimulai dengan *Admin* mengisi jumlah data latih dan nilai *K* yang diinginkan, sistem akan melakukan proses klasifikasi penyakit kardiovaskular kemudian akan ditampilkan hasil klasifikasi dengan metode KNN.

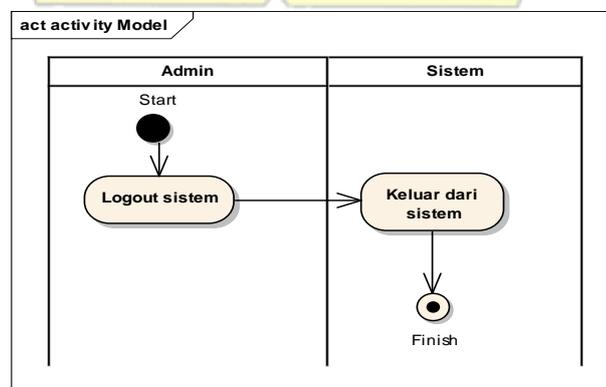
#### 4. Klasifikasi Penyakit Kardiovaskular

Klasifikasi penyakit kardiovaskular dimulai dengan *Admin* mengisi data Penyakit Kardiovaskular, sistem akan melakukan proses klasifikasi penyakit kardiovaskular kemudian akan ditampilkan hasil klasifikasi penyakit kardiovaskular.



Gambar 3. 6. *Activity Diagram* Klasifikasi Kardiovaskular

#### 5. Logout



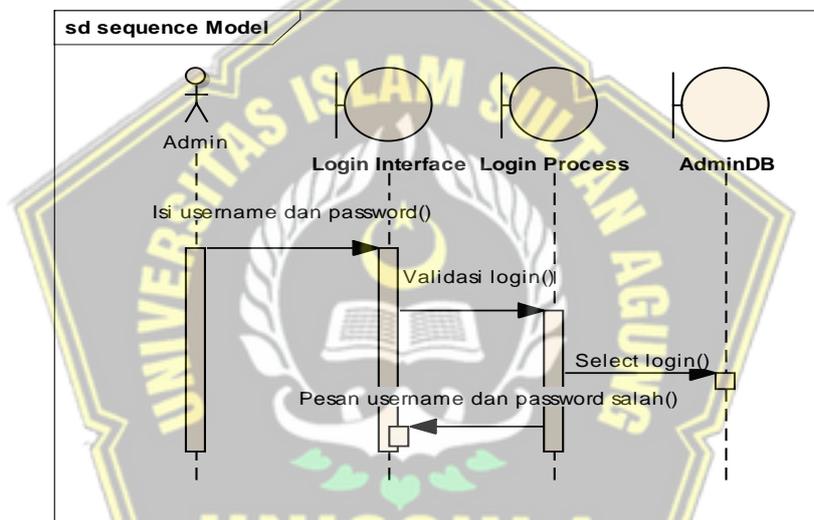
Gambar 3.7. *Activity Diagram* Logout

*Logout* sistem menjelaskan tentang *Admin* memilih *Logout* dan keluar dari sistem

### 3.9 Sequence Diagram

#### 1. Login

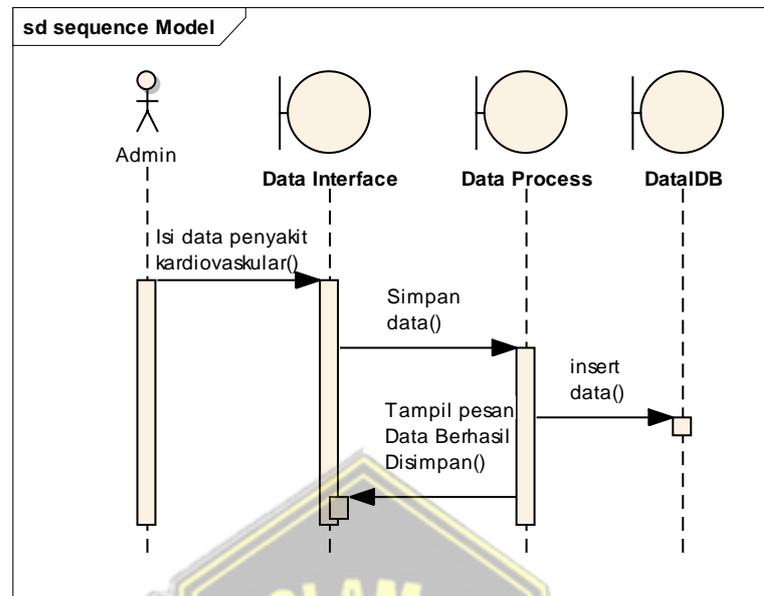
*Login* dimulai dengan *Admin* mengisi *username* dan *password* kemudian sistem melakukan verifikasi *username* dan *password*, jika *username* dan *password* akan tampil menu *Admin* sedangkan jika pengisian *username* dan *password* tidak *valid* akan ditampilkan pesan *username* dan *password* salah dan *Admin* melakukan pengisian *username* dan *password* lagi.



Gambar 3. 8. Sequence Diagram *Login*

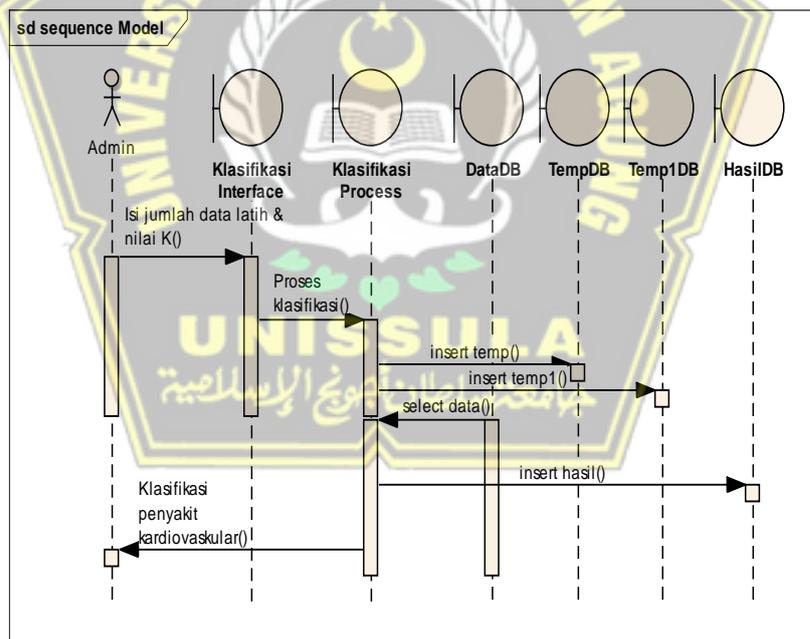
#### 2. Mengelola Data Penyakit Kardiovaskular

Mengelola data penyakit kardiovaskular dimulai dengan *Admin* mengisi data penyakit kardiovaskular dan sistem akan melakukan verifikasi dari data yang dimasukkan, jika data yang dimasukkan *valid*, maka data akan disimpan di tabel data.



Gambar 3.9. *Sequence Diagram* Mengelola Kardiovaskular

### 3. Pengujian KNN

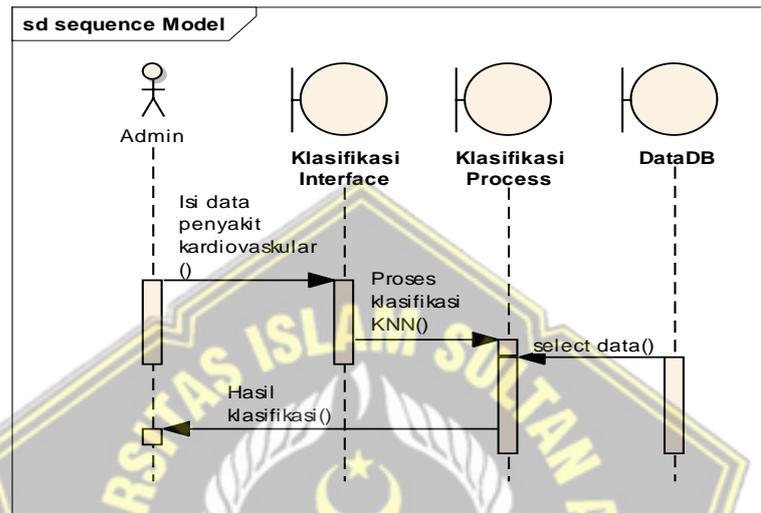


Gambar 3.10. *Sequence Diagram* Pengujian KNN

Pengujian KNN dimulai dengan *Admin* memasukkan jumlah data latih dan nilai *K*, sistem akan melakukan proses klasifikasi penyakit kardiovaskular kemudian akan ditampilkan hasil klasifikasi dengan metode KNN.

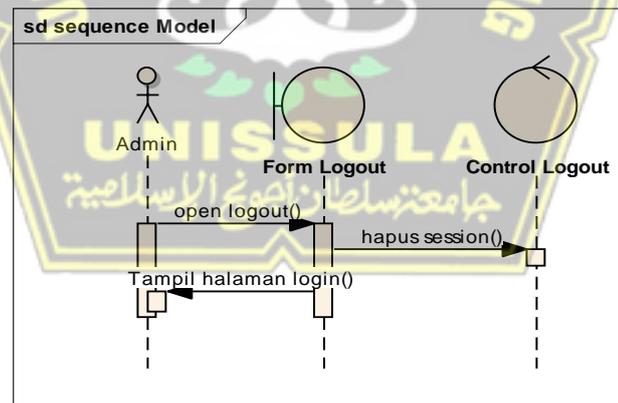
#### 4. Klasifikasi Penyakit Kardiovaskular

Klasifikasi penyakit kardiovaskular dimulai dengan *Admin* mengisi data penyakit kardiovaskular, sistem akan melakukan proses klasifikasi penyakit kardiovaskular kemudian akan ditampilkan hasil klasifikasi penyakit kardiovaskular.



Gambar 3.11. *Sequence Diagram* Klasifikasi Kardiovaskular

#### 5. Logout

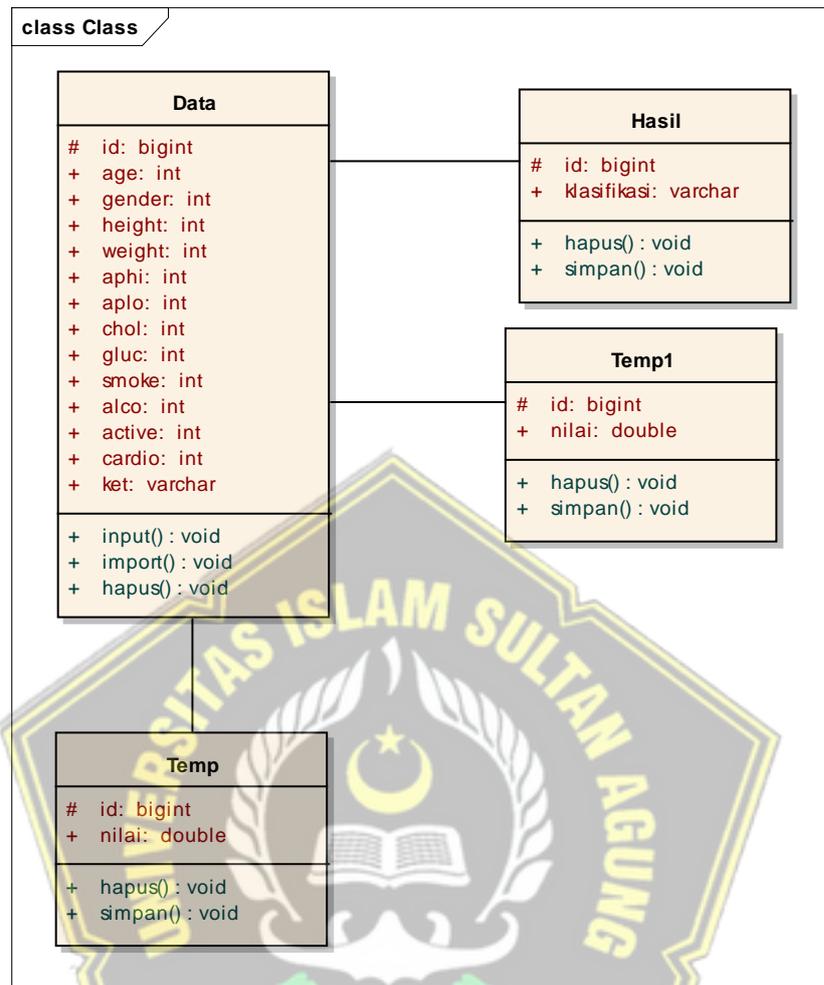


Gambar 3.12. *Sequence Diagram* Logout

*Logout* sistem menjelaskan tentang *Admin* memilih *Logout* dan keluar dari sistem.

### 3.10 *Class Diagram*

*Class diagram* sistem klasifikasi penyakit kardiovaskular menggunakan algoritma KNN diperlihatkan seperti gambar 3.13.



Gambar 3. 13. Class Diagram

Gambar 3.13 menjelaskan *class* data dengan *primary key* id berelasi dengan *class* hasil dengan *primary key* id. *Class* data dengan *primary key* id berelasi dengan *class* temp dengan *primary key* id. *Class* data dengan *primary key* id berelasi dengan *class* temp1 dengan *primary key* id.

### 3.11 Rancangan Database

Database penerapan data mining untuk klasifikasi penyakit kardiovaskular menggunakan algoritma KNN terdiri dari 4 tabel yaitu tabel data, tabel hasil, tabel temp dan tabel temp1.

#### 1. Tabel Data

Tabel data digunakan untuk menyimpan data penyakit kardiovaskular. Tabel data diperlihatkan seperti pada tabel 3.3.

Tabel 3.3. Tabel Data

No	Field Name	Size	Type	Keterangan
1	<i>ID</i>	20	<i>Bigint</i>	ID
2	<i>Age</i>	3	<i>Int</i>	Usia
3	<i>Gender</i>	1	<i>Int</i>	Jenis Kelamin
4	<i>Height</i>	3	<i>Int</i>	Tinggi Badan
5	<i>Weight</i>	3	<i>Int</i>	Berat Badan
6	<i>Aphi</i>	3	<i>Int</i>	Tekanan Darah Sistolik
7	<i>Aplo</i>	3	<i>Int</i>	Tekanan Darah Diastolik
8	<i>Chol</i>	1	<i>Int</i>	Kolesterol
9	<i>Gluc</i>	1	<i>Int</i>	Kadar Gula dalam Darah
10	<i>Smoke</i>	1	<i>Int</i>	Merokok
11	<i>Alco</i>	1	<i>Int</i>	Alkohol
12	<i>Active</i>	1	<i>Int</i>	<i>Physical Activity</i>
13	<i>Cardio</i>	1	<i>Int</i>	Kardiovaskular
14	<i>Ket</i>	5	<i>Varchar</i>	Keterangan

## 2. Tabel Hasil

Tabel hasil digunakan untuk menyimpan hasil klasifikasi penyakit kardiovaskular dengan algoritma KNN. Tabel hasil diperlihatkan seperti pada tabel 3.4.

Tabel 3.4. Tabel Hasil

No	Field Name	Size	Type	Keterangan
1	<i>ID</i>	20	<i>Bigint</i>	ID
2	<i>Klasifikasi</i>	1	<i>Int</i>	Hasil Klasifikasi

## 3. Tabel Temp

Tabel temp digunakan untuk menyimpan hasil perhitungan jarak *euclidean* dengan algoritma KNN. Tabel temp diperlihatkan seperti pada tabel 3.5.

Tabel 3.5. Tabel Temp

No	Field Name	Size	Type	Keterangan
1	<i>ID</i>	20	<i>Bigint</i>	ID
2	<i>Nilai</i>		<i>Double</i>	Nilai Jarak <i>Euclidean</i>

## 4. Tabel Temp1

Tabel temp1 digunakan untuk menyimpan hasil perhitungan algoritma KNN. Tabel temp1 diperlihatkan seperti pada tabel 3.6.

Tabel 3.6. Tabel Temp1

No	Field Name	Size	Type	Keterangan
1	ID	20	Bigint	ID
2	Nilai		Double	Nilai KNN

### 3.12 Rancangan Interface

#### 1. Login

Gambar 3.14 menjelaskan tentang halaman *Login* yang digunakan *Admin* untuk masuk ke sistem klasifikasi penyakit kardiovaskular dengan algoritma KNN. Isi *username* dan *password* kemudian klik *Login* untuk masuk ke sistem klasifikasi penyakit kardiovaskular dengan algoritma KNN, jika data *valid* akan ditampilkan halaman sistem klasifikasi penyakit kardiovaskular dengan algoritma KNN.

Gambar 3. 14. *Login*

#### 2. Input Data Kardiovaskular

Gambar 3.15 menjelaskan tentang halaman *input* data yang digunakan untuk mengelola data dari penyakit kardiovaskular.

Data Pengujian Akurasi Klasifikasi Admin Logout

**Data Penyakit Kardiovaskular**

Age

Gender

Height

Weight

Systolic

Diastolic

Cholesterol

Glucose

Smoking

Alcohol Intake

Physical Activity

Cadiovaskular

Tampilkan  Item Masukkan Kata Kunci

No	Age	Height	Weight	Gender	Systolic	Ddiastolic	Cholesterol	Proses

Gambar 3.15. Input Data Kardiovaskular

### 3. Pengujian

Gambar 3.16 menjelaskan tentang halaman pengujian yang digunakan untuk memproses data penyakit kardiovaskular menggunakan metode KNN.

Home Pengujian Akurasi Klasifikasi Admin Logout

Jumlah Data

Data Latih

Nilai K

Gambar 3. 16. Pengujian

### 4. Akurasi

Gambar 3.17 menjelaskan tentang halaman akurasi yang digunakan untuk melihat hasil akurasi klasifikasi penyakit kardiovaskular dengan algoritma KNN.

Home	Pengujian	Akurasi	Klasifikasi	Admin	Logout				
<h2>Hasil Akurasi KNN</h2>									
Tampilkan <input type="text"/> Item			Masukkan Kata Kunci <input type="text"/>						
ID	Age	Height	Weight	Gender	Systolic	Ddiastolic	Cholesterol	Cardio	Hasil

Gambar 3. 17. Hasil

## 5. Klasifikasi

Home	Pengujian	Akurasi	Klasifikasi	Admin	Logout
<h3>Klasifikasi Penyakit Kardiovaskular</h3>					
Age	<input type="text"/>				
Gender	<input type="text"/>				
Height	<input type="text"/>				
Weight	<input type="text"/>				
Systolic	<input type="text"/>				
Diastolic	<input type="text"/>				
Cholesterol	<input type="text"/>				
Glukose	<input type="text"/>				
Smoking	<input type="text"/>				
Alcohol	<input type="text"/>				
Physical Activity	<input type="text"/>				
Nilai K	<input type="text"/>				
		<input type="button" value="Proses"/>	<input type="button" value="Batal"/>		
Hasil Klasifikasi		<input type="text"/>			

Gambar 3. 18. Klasifikasi

Gambar 3.18 menjelaskan tentang halaman klasifikasi yang digunakan untuk melihat hasil penyakit kardiovaskular dengan algoritma KNN.

## 6. Admin

Gambar 3.19 menjelaskan tentang halaman *Admin* yang digunakan *Admin* untuk mengelola data *Admin* sistem klasifikasi penyakit kardiovaskular dengan algoritma KNN.

Home	Pengujian	Akurasi	Klasifikasi	Admin	Logout
<b>Data Admin</b>					
User Name	<input type="text"/>				
Password	<input type="text"/>				
	<input type="button" value="Simpan"/>		<input type="button" value="Batal"/>		
Tampilkan	<input type="text"/>	Item	Masukkan Kata Kunci	<input type="text"/>	
No	Username	Password	Proses		

Gambar 3.19. Admin

### 3.13 Metode Pengujian Sistem

Tahap ini adalah melakukan pengujian dari metode KNN yang akan diuji menggunakan diagram *confusion matrix* untuk mengetahui nilai akurasi yang diperoleh. *Confusion matrix* merupakan salah satu metode yang dapat digunakan untuk mengukur kinerja suatu metode klasifikasi. Pada dasarnya *confusion matrix* mengandung informasi yang membandingkan hasil klasifikasi yang dilakukan oleh sistem dengan hasil klasifikasi yang seharusnya. Pengukuran klasifikasi menggunakan *confusion matrix*, terdapat beberapa langkah yang digunakan yaitu

a. Hitung *True Positive* (TP)

Merupakan data positif yang diprediksi benar. Contohnya, pasien menderita penyakit kardiovaskular dan dari model yang dibuat memprediksi pasien tersebut menderita penyakit kardiovaskular.

b. Hitung *True Negative* (TN)

Merupakan data negatif yang diprediksi benar. Contohnya, pasien tidak menderita penyakit kardiovaskular dan dari model yang dibuat memprediksi pasien tersebut tidak menderita penyakit kardiovaskular.

c. Hitung *False Postive* (FP)

Merupakan data negatif namun diprediksi sebagai data positif. Contohnya, pasien tidak menderita penyakit kardiovaskular tetapi dari model yang telah memprediksi pasien tersebut menderita penyakit kardiovaskular.

d. *Hitung False Negative (FN)*

Merupakan data positif namun diprediksi sebagai data negatif. Contohnya, pasien menderita penyakit kardiovaskular tetapi dari model yang dibuat memprediksi pasien tersebut tidak menderita penyakit kardiovaskular.

e. *Hitung Akurasi*

Setelah diketahui TP, TN, FP dan FN proses selanjutnya adalah menghitung akurasi atau rasio prediksi benar (positif dan negatif) dari metode KNN dalam melakukan klasifikasi penyakit kardiovaskular yang dihitung dengan menggunakan

$$\text{rumus akurasi} = \frac{TP+TN}{TP+TN+FP+FN} \times 100 \%$$



## BAB IV HASIL DAN ANALISIS PENELITIAN

### 4.1 *Login*

Gambar 4.1 menjelaskan tentang halaman *Login* yang digunakan *Admin* untuk masuk ke sistem klasifikasi penyakit kardiovaskular dengan algoritma KNN.



Gambar 4.1. *Login*

Isi *username* dan *password* kemudian klik *Login* untuk masuk ke sistem, jika data *valid* akan ditampilkan halaman sistem klasifikasi penyakit kardiovaskular dengan algoritma KNN sedangkan jika tidak *valid* akan ditampilkan pesan *username* atau *password* salah.

### 4.2 *Home*

Gambar 4.2 menjelaskan tentang halaman home yang digunakan *Admin* untuk mengelola data penyakit kardiovaskular.

Age :

Gender :

Height :  Cm

Weight :  Kg

Systolic :  mmHg

Diastolic :  mmHg

Cholesterol :

Glucose :

Smoking :

Alcohol Intake :

Physical Activity :

Cardiovaskular :

Tampilkan  item

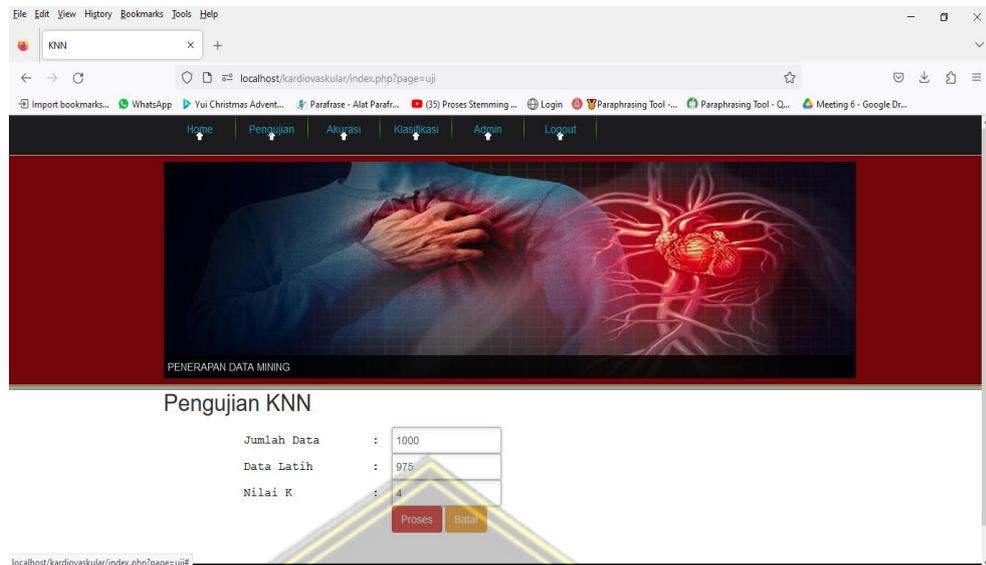
No.	Age	Height	Weight	Gender	Smoking	Systolic	Diastolic	Cholesterol	Proses
1.	62	155	69	1	0	130	80	2	<input type="button" value="edit"/> <input type="button" value="hapus"/>
2.	40	163	71	1	0	110	70	1	<input type="button" value="edit"/> <input type="button" value="hapus"/>

Gambar 4.2. Home

Isi *age*, *height*, *weight*, *gender*, *systolic*, *diastolic*, *cholesterol*, *glucose*, *smooking*, *alcohol intake*, *physical activity*, *cardiovaskular* kemudian tekan tombol simpan untuk menyimpan data penyakit kardiovaskular ke tabel data. Klik edit kemudian isi *age*, *height*, *weight*, *gender*, *systolic*, *diastolic*, *cholesterol*, *glucose*, *smooking*, *alcohol*, *physical activity*, *cardiovaskular* dan tekan tombol simpan untuk mengubah data penyakit kardiovaskular dari tabel data. Klik hapus kemudian pilih oke untuk menghapus data penyakit kardiovaskular dari tabel data.

### 4.3 Pengujian

Gambar 4.3 menjelaskan tentang halaman pengujian yang digunakan *Admin* untuk menguji metode KNN dalam melakukan klasifikasi penyakit kardiovaskular . Isi data latih dan nilai K yang diinginkan untuk menguji metode KNN kemudian klik tombol proses untuk menampilkan hasil klasifikasi dan akurasi penyakit kardiovaskular menggunakan metode KNN.



Gambar 4.3. Pengujian

Pada gambar 4.3 merupakan tampilan pengujian sistem, yang dimana user dapat melakukan pengujian dengan meng-input-kan seeperti Jumlah data, data latih, dan nilai K.

#### 4.4 Akurasi

Pada tampilan halaman hasil akurasi seperti yang ditunjukkan gambar 4.4 merupakan data hasil dari perhitungan dengan metode KNN. Dimana pada nilai data yang terdekat dengan akurasi akan direkomendasikan kepada *user*.

Hasil Akurasi KNN = 88%

ID	Age	Height	Weight	Gender	Smoking	Systolic	Diastolic	Cholesterol	Cardio	Hasil
976	54	168	94	1	0	120	80	1	0	0
977	60	159	64	2	0	100	60	1	0	0
978	54	179	113	2	0	160	90	2	1	1
979	54	182	92	1	0	145	90	1	1	1
980	44	172	67	2	0	130	70	2	0	0
981	59	170	76	2	0	120	80	1	1	1
982	64	147	36	1	0	110	70	3	0	0
983	44	150	34	1	0	110	70	1	0	0
984	60	170	67	1	0	130	90	1	1	1
985	40	168	69	1	0	120	90	1	0	0

Gambar 4.4. Akurasi

Gambar 4.4 menjelaskan tentang halaman akurasi yang digunakan *Admin* untuk melihat hasil klasifikasi dan akurasi penyakit kardiovaskular menggunakan metode KNN. Hasil akan ditampilkan dalam bentuk tabel yang berisi *id, age, height, weight, gender, smooking, systolic, diastolic, cholesterol, cardio* dan hasil klasifikasi penyakit kardiovaskular dengan metode KNN. Pada atas tabel juga akan ditampilkan akurasi dari metode KNN menggunakan pengukuran *confusion matrix*.

#### 4.5 Klasifikasi

Gambar 4.5 menjelaskan tentang halaman klasifikasi yang digunakan *Admin* untuk melihat hasil klasifikasi penyakit kardiovaskular menggunakan metode KNN dengan hasil klasifikasi 1 atau 0.

Klasifikasi Penyakit Kardiovaskular

Age	: 45	Tahun
Gender	: Perempuan	
Height	: 145	Cm
Weight	: 80	Kg
Systolic	: 145	mmHg
Diastolic	: 90	mmHg
Cholesterol	: Above Normal	
Glucose	: Above Normal	
Smoking	: No	
Alcohol Intake	: Yes	
Physical Activity	: Yes	
Nilai K	: 4	
		Proses Batal
Hasil Klasifikasi		

Gambar 4.5. Klasifikasi

Isi *age, height, weight, gender, systolic, diastolic, cholesterol, glucose, smooking, alcohol, physical activity*, nilai K kemudian tekan tombol proses untuk memproses data penyakit kardiovaskular dengan menggunakan metode KNN dan hasil klasifikasi akan ditampilkan seperti gambar 4.6. jika hasil klasifikasi 1 maka berpotensi terkena penyakit kardiovaskular sedangkan jika 0 tidak berpotensi terkena penyakit kardiovaskular.

Klasifikasi Penyakit Kardiovaskular

Age : 45 Tahun

Gender : Perempuan

Height : 145 Cm

Weight : 80 Kg

Systolic : 145 mmHg

Diastolic : 90 mmHg

Cholesterol : Above Normal

Glucose : Above Normal

Smoking : No

Alcohol Intake : Yes

Physical Activity : Yes

Nilai K : 4

Hasil Klasifikasi : 1

Proses

Gambar 4.6. Hasil Klasifikasi

#### 4.6 Admin

PENERAPAN DATA MINING

Data Admin

Username :

Password :

Simpan

Tampilkan 10 Item

Masukkan Kata Kunci:

No.	Username	Password	Proses
1.	admin	123	<input type="checkbox"/> <input type="checkbox"/>

Tampilkan 1 - 1 dari 1 Item

Prev 1 Next

Gambar 4.7. Admin

Gambar 4.7 menjelaskan tentang halaman *Admin* yang digunakan *Admin* untuk mengelola data *Admin* sistem klasifikasi penyakit kardiovaskular dengan algoritma KNN. Isi *username*, *password* kemudian tekan tombol simpan untuk menyimpan data *Admin* ke tabel *Admin*. Klik edit kemudian isi *password* dan tekan tombol simpan untuk mengubah data *Admin* dari tabel *Admin*. Klik hapus kemudian pilih oke untuk menghapus data *Admin* dari tabel *Admin*.

#### 4.7 Perhitungan KNN

Diketahui data uji dari klasifikasi penyakit kardiovaskular yang diperlihatkan seperti tabel 4.1.

Tabel 4.1. Data Uji

age	gender	Height	Weig ht	ap_ hi	ap_l o	chol ester ol	glu c	Smo ke	alc o	activ e	cardio
54	1	168	94	120	80	1	1	0	0	1	0
60	2	159	64	100	60	1	1	0	0	1	0
54	2	179	115	160	90	2	1	0	1	0	1
54	1	182	92	145	90	1	1	0	1	1	1
44	2	172	67	130	70	2	1	0	0	1	0
59	2	170	76	120	80	1	1	0	0	1	1
64	1	147	56	110	70	3	3	0	0	1	0
44	1	150	54	110	70	1	1	0	0	1	0
60	1	170	67	130	90	1	1	0	0	1	1
40	1	168	69	120	90	1	1	0	0	1	0
58	1	155	63	140	90	1	1	0	0	1	0
62	1	160	88	120	80	2	1	0	0	0	1
40	2	170	81	120	80	1	1	0	0	1	0
60	1	153	71	140	90	1	1	0	0	1	1
58	1	163	79	130	80	1	2	0	0	1	0
53	1	162	94	120	80	2	1	0	0	1	0
50	1	162	58	110	70	1	1	0	0	1	0
54	1	171	45	110	70	1	1	0	0	1	0
50	1	165	69	130	80	1	2	0	0	1	0
50	1	157	73	130	60	1	1	0	0	1	0
52	1	157	80	110	70	1	1	0	0	1	0
60	1	155	69	110	80	1	2	0	0	1	0
56	1	155	58	100	70	1	1	0	0	0	0
50	2	165	80	120	80	1	2	0	0	1	0
44	1	150	64	120	80	1	1	0	0	1	0

### 1) Penentuan Nilai K

Penentuan nilai K adalah langkah paling awal dan paling krusial dalam algoritma KNN melakukan proses klasifikasi penyakit kardiovaskular . Dimana misalkan nilai K diinisialisasi dengan nilai 4, maka perhitungan jarak dengan algoritma tertentu pada KNN akan diambil 4 nilai terkecil.

### 2) Perhitungan Jarak dengan *Euclidean Distance*

*Euclidean distance* adalah salah satu perhitungan matematis untuk menghitung jarak dua titik dengan menggunakan rumus pythagoras. Dalam hal ini dua titik yang dimaksud akan memiliki atribut yang sama persis seperti yang ada pada data latih. Berikut adalah rumus *euclidean distance* :

$$D = \sqrt{(X_1 - X_2)^2 + (Y_1 - Y_2)^2}$$

### 3) Tahap KNN

Langkah pertama, lakukan perhitungan jarak antara data uji dengan data latih.

Data uji yang digunakan yaitu

Age	Gender	height	weig ht	ap_ hi	ap_l o	chol ester ol	glu c	smok e	Alco	activ e
54	1	168	94	120	80	1	1	0	0	1

$$D(976,1) = \sqrt{\begin{aligned} &(54 - 62)^2 + (1 - 1)^2 + (168 - 155)^2 \\ &+ (94 - 69)^2 + (120 - 130)^2 + (80 - 80)^2 \\ &+ (1 - 2)^2 + (1 - 2)^2 + (0 - 0)^2 \\ &+ (0 - 0)^2 + (1 - 1)^2 \end{aligned}}$$

$$= 21,82$$

$$D(976,2) = \sqrt{\begin{aligned} &(54 - 40)^2 + (1 - 1)^2 + (168 - 163)^2 \\ &+ (94 - 71)^2 + (120 - 110)^2 + (80 - 70)^2 \\ &+ (1 - 1)^2 + (1 - 1)^2 + (0 - 0)^2 \\ &+ (0 - 0)^2 + (1 - 1)^2 \end{aligned}}$$

$$= 20,83$$

$$\begin{aligned}
 D(976,3) &= \sqrt{(54 - 21901)^2 + (1 - 1)^2 + (168 - 165)^2} \\
 &\quad + (94 - 70)^2 + (120 - 120)^2 + (80 - 80)^2 \\
 &\quad + (1 - 1)^2 + (1 - 1)^2 + (0 - 0)^2 \\
 &\quad + (0 - 0)^2 + (1 - 1)^2 \\
 &= 22,74
 \end{aligned}$$

...

$$\begin{aligned}
 D(976,975) &= \sqrt{(54 - 62)^2 + (1 - 1)^2 + (168 - 158)^2} \\
 &\quad + (94 - 75)^2 + (120 - 120)^2 + (80 - 80)^2 \\
 &\quad + (1 - 1)^2 + (1 - 1)^2 + (0 - 0)^2 \\
 &\quad + (0 - 0)^2 + (1 - 1)^2 \\
 &= 22,56
 \end{aligned}$$

Setelah semua jarak didapat dengan rumus *euclidean distance*, urutkan sebanyak K nilai terkecil dari jarak-jarak yang sudah dihitung seperti tabel 4.2.

Tabel 4. 2. Hasil Perhitungan *Euclidean Distance*

No	Data Ke	Jarak Euclidean	Hasil	Keterangan
1	626	4,58	0	Tidak Terjangkit
2	644	6,56	0	Tidak Terjangkit
3	200	6,63	0	Tidak Terjangkit
4	212	7,28	0	Tidak Terjangkit

Setelah diurutkan dari 4 nilai terkecil, maka disimpulkan bahwa data yang baru berlabel 0 karena dari 4 nilai terkecil tersebut label paling banyak adalah 0. Hasil perhitungan keseluruhan dengan K=4 diperlihatkan seperti tabel 4.3.

Tabel 4. 3. Hasil Perhitungan K=4

No	Data Ke	Hasil	Data Ke	Hasil	Keterangan
1	976	0	989	1	Terjangkit
2	977	0	990	1	Terjangkit
3	978	1	991	0	Tidak Terjangkit
4	979	1	992	0	Tidak Terjangkit
5	980	0	993	0	Tidak Terjangkit
6	981	1	994	0	Tidak Terjangkit

No	Data Ke	Hasil	Data Ke	Hasil	Keterangan
1	982	0	995	0	Tidak Terjangkit
2	983	0	996	0	Tidak Terjangkit
3	984	1	997	0	Tidak Terjangkit
4	985	0	998	0	Tidak Terjangkit
5	986	1	999	0	Tidak Terjangkit
6	987	0	1000	0	Tidak Terjangkit
7	988	0		1	Terjangkit

Dari perhitungan dengan menggunakan seluruh data uji, selanjutnya dilakukan pengujian menggunakan diagram *confusion matrix* untuk mengetahui nilai akurasi yang diperoleh dari algoritma KNN yang ditunjukkan seperti tabel 4.4.

Tabel 4.4. Hasil *Confusion Matrix* Algoritma KNN

	True Positif	True Negatif
Pred. Positif	5	2
Pred. Negatif	1	17

Data uji yang berjumlah 25 menghasilkan hasil klasifikasi sebagai berikut:

1. *True positif* menghasilkan 5 data
2. *False positif* menghasilkan 2 data
3. *False negatif* menghasilkan 1 data
4. *True negatif* menghasilkan 17 data

Hasil dari tabel *confusion matrix* 5.3 didapatkan nilai akurasi sebesar 88 % dengan perhitungan sebagai berikut:

Akurasi =  $\frac{5+17}{5+2+1+17} \times 100 \% = 88 \%$ . Hasil akurasi untuk nilai K=4 adalah 88 % yang diperlihatkan seperti gambar 4.8.

File Edit View History Bookmarks Tools Help

KNN

localhost/kardiovaskular/index.php?page=akurasi

PENERAPAN DATA MINING

Hasil Akurasi KNN = 88%

Tampilkan 10 item Masukan Kata Kunci:

ID	Age	Height	Weight	Gender	Smoking	Systolic	Diastolic	Cholesterol	Cardio	Hasil
976	54	168	94	1	0	120	80	1	0	0
977	60	159	64	2	0	100	60	1	0	0
978	54	179	115	2	0	160	90	2	1	1
979	54	182	92	1	0	145	90	1	1	1
980	44	172	67	2	0	130	70	2	0	0
981	59	170	76	2	0	120	80	1	1	1
982	64	147	56	1	0	110	70	3	0	0
983	44	150	54	1	0	110	70	1	0	0
984	60	170	67	1	0	130	90	1	1	1
985	40	168	69	1	0	120	90	1	0	0

Gambar 4. 8. Hasil Akurasi K=4

Hasil akurasi untuk nilai K=3 adalah 84 % yang diperlihatkan seperti gambar 4.9.

File Edit View History Bookmarks Tools Help

KNN

localhost/kardiovaskular/index.php?page=akurasi

Hasil Akurasi KNN = 84%

Tampilkan 10 item Masukan Kata Kunci:

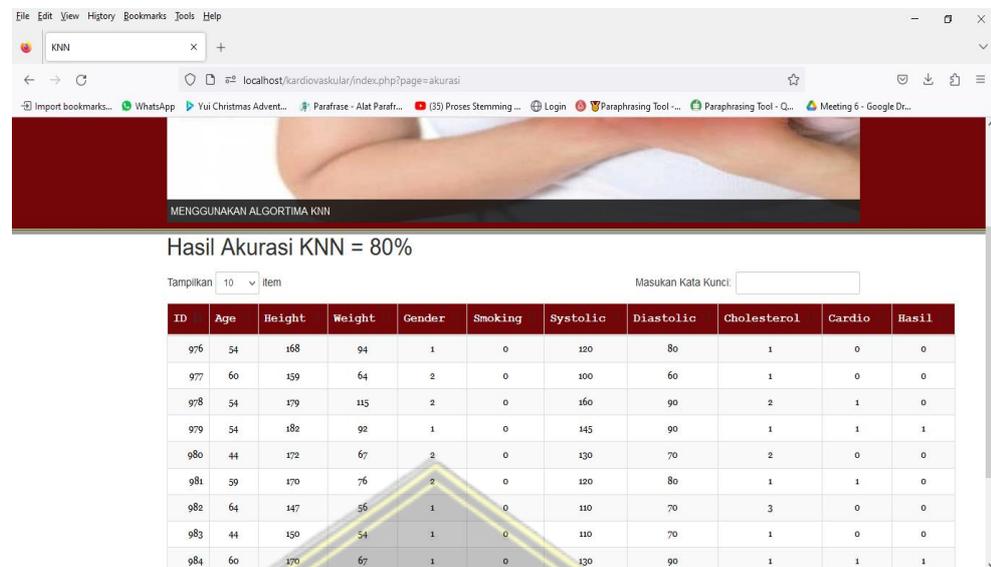
ID	Age	Height	Weight	Gender	Smoking	Systolic	Diastolic	Cholesterol	Cardio	Hasil
976	54	168	94	1	0	120	80	1	0	0
977	60	159	64	2	0	100	60	1	0	0
978	54	179	115	2	0	160	90	2	1	1
979	54	182	92	1	0	145	90	1	1	1
980	44	172	67	2	0	130	70	2	0	0
981	59	170	76	2	0	120	80	1	1	0
982	64	147	56	1	0	110	70	3	0	0
983	44	150	54	1	0	110	70	1	0	0
984	60	170	67	1	0	130	90	1	1	0
985	40	168	69	1	0	120	90	1	0	0

Tampilkan 1 - 10 dari 25 item

Prev 1 2 3 Next

Gambar 4. 9. Hasil Akurasi K=3

Hasil akurasi untuk nilai K=1 adalah 80 % yang diperlihatkan seperti gambar 4.10.



The screenshot shows a web browser window with the title 'KNN'. The address bar shows 'localhost/kardiovaskular/index.php?page=akurasi'. Below the browser, there is a banner image of a person's arm with the text 'MENGGUNAKAN ALGORITMA KNN'. Below the banner, the text 'Hasil Akurasi KNN = 80%' is displayed. There is a search bar labeled 'Masukan Kata Kunci:' and a dropdown menu for 'Tampilkan' set to '10' items. Below this is a table with 11 columns: ID, Age, Height, Weight, Gender, Smoking, Systolic, Diastolic, Cholesterol, Cardio, and Hasil. The table contains 10 rows of data.

ID	Age	Height	Weight	Gender	Smoking	Systolic	Diastolic	Cholesterol	Cardio	Hasil
976	54	168	94	1	0	120	80	1	0	0
977	60	159	64	2	0	100	60	1	0	0
978	54	179	115	2	0	160	90	2	1	0
979	54	182	92	1	0	145	90	1	1	1
980	44	172	67	2	0	130	70	2	0	0
981	59	170	76	2	0	120	80	1	1	0
982	64	147	56	1	0	110	70	3	0	0
983	44	150	54	1	0	110	70	1	0	0
984	60	170	67	1	0	130	90	1	1	1

Gambar 4. 10. Hasil Akurasi K=1

Nilai akurasi tertinggi dari perubahan nilai parameter K yang terdiri dari 1,3 dan 4 pada metode KNN didapatkan hasil akurasi tertinggi pada nilai parameter K = 4 dengan akurasi 88 %.



## **BAB V**

### **PENUTUP**

Berdasarkan hasil yang telah dicapai dalam penelitian ini penulis berharap dengan tercapainya penerapan data mining untuk klasifikasi penyakit kardiovaskular menggunakan algoritma KNN. Dalam penelitian ini dapat ditarik suatu kesimpulan dan saran yang tentunya tidak melupakan saran-saran dari pembaca sebagai bahan masukan bilamana pembaca tertarik untuk mengembangkan aplikasi ini lebih lanjut.

#### **5.1. Kesimpulan**

Kesimpulan yang diambil dari penelitian penerapan data mining untuk klasifikasi penyakit kardiovaskular menggunakan algoritma KNN sebagai berikut :

- 1) Penelitian ini dapat digunakan untuk melakukan klasifikasi penyakit kardiovaskular dengan menentukan status positif (1) dan negatif (0) menggunakan algoritma KNN.
- 2) Data yang digunakan berjumlah 1.000 data dibagi menjadi 2 data dimana 975 digunakan sebagai data latih dan 25 digunakan sebagai data uji yang didapatkan dari *website* Kaggle Data Set.
- 3) Nilai akurasi tertinggi dari perubahan nilai parameter K yang terdiri dari 1,3 dan 4 pada metode KNN didapatkan hasil akurasi tertinggi pada nilai parameter K = 4 dengan akurasi 88 %.

#### **5.2. Saran**

Berikut ini saran penulis terhadap penerapan data mining untuk klasifikasi penyakit kardiovaskular menggunakan algoritma KNN lebih lanjut yaitu penelitian ini dapat dikembangkan dengan metode optimasi seperti *PSO* atau *forward selection* yang berfungsi untuk meningkatkan akurasi.

## DAFTAR PUSTAKA

- Abdulloh, R., 2018. *7 in 1 Pemrograman Web untuk Pemula*. Jakarta: PT. Elex Media Komputindo.
- Atthalla, I. N., Jovandy, A. & Habibie, H., 2018. Klasifikasi Penyakit Kanker Payudara Menggunakan Metode K Nearest Neighbor. *Prosiding Annual Research Seminar*, IV(1), pp. 148-151.
- Benjamin, R. M., 2012. The Million Heart Initiative: Progress In Preventing Heart Attacks And Stroke. *Surgeon General's Perspectives*, pp. 558-560.
- Claudy, Y. I., Perdana, R. S. & Fauzi, M. A., 2018. Klasifikasi Dokumen Twitter Untuk Mengetahui Karakter Calon Karyawan Menggunakan Algoritma K-Nearest Neighbor (KNN). *Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer*, II(8), pp. 2761-2765.
- Fadlilah, S., Sucipto, A. & Amestiasih, T., 2019. Usia, Jenis Kelamin, Perilaku Merokok dan IMT Berhubungan Dengan Resiko Penyakit Jantung. *Jurnal Keperawatan*, XII(4), pp. 261-268.
- Han, J. & Kamber, M., 2016. *Data Mining: Concept and Techniques, Third Edition*. Waltham: Morgan Kaufmann.
- Hermawati, F. A., 2016. *Data Mining*. Yogyakarta: Andi.
- Indrayanti, Sugianti, D. & Karomi, A., 2017. Peningkatan Akurasi Algoritma KNN dengan Seleksi Fitur Gain Ratio untuk Klasifikasi Penyakit Diabetes Mellitus. *IC-Tech*, XII(2), pp. 1-6.
- Isman, Ahmad, A. & Latief, A., 2021. Perbandingan Metode KNN Dan LBPH Pada Klasifikasi Daun Herbal. *RESTI*, V(3), p. 557 – 564 .
- Juslim, R. R. & Herawati, F., 2018. *Penyakit Jantung Kardiovaskular*. Yogyakarta: Graha Ilmu.
- Kusrini & Luthfi, E. T., 2015. *Algoritma Data Mining*. Yogyakarta: Andi.
- Munawar, 2018. *Analisis Perancangan Sistem Berorientasi Objek dengan UML (Unified Modeling Language)*. Bandung: Informatika.
- Muslihudin, M., 2017. *Analisis Dan Perancangan Sistem Informasi Menggunakan Model Terstruktur Dan UML*. Yogyakarta: Andi.
- Nugroho, B., 2015. *Database Relasional Dengan MySQL*. Yogyakarta: Andi.

- Pressman, R., 2012. *Rekayasa Perangkat Lunak : Pendekatan Praktisi (Buku Satu)*. Yogyakarta: Andi.
- Rivki, M. & Bachtiar, A. M., 2017. Implementasi Algoritma K-Nearest Neighbor Dalam Pengklasifikasian Follower Twitter Yang Menggunakan Bahasa Indonesia. *Jurnal Sistem Informasi*, XIII(1), pp. 31-37.
- Santosa, B., 2016. *Data Mining: Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Setiawan, D., 2018. *Buku Sakti Pemrograman Web HTML, CSS, PHP, MYSQL & JAVASCRIPT*. Yogyakarta: Start Up.
- Ting, K. M., 2017. *Confusion Matrix*, in *Encyclopedia of Machine Learning and Data Mining*. Boston: Springer.
- Zhao, D., 2021. Epidemiology Features of Heart Disease in Asia. *JACC*, I(1), pp. 1-13.

