

BAB 3

ANALISIS DAN PERANCANGAN SISTEM

3.1. ANALISIS SISTEM

Jantung adalah organ krusial yang bertanggung jawab memompa darah ke seluruh tubuh, memastikan suplai oksigen dan nutrisi yang cukup untuk fungsi organ-organ lainnya. Gangguan pada jantung dapat berdampak besar pada kesehatan secara keseluruhan.

Salah satu masalah utama adalah deteksi dini penyakit jantung, karena banyak pasien tidak menyadari data kesehatan awal atau memiliki informasi yang cukup mengenai risiko penyakit jantung. Deteksi dan diagnosis sering memerlukan kunjungan ke pengguna spesialis, pemeriksaan di pusat medis, dan biaya yang tinggi, membuat prosesnya tidak terjangkau bagi banyak orang. Selain itu, penyakit jantung sering kali berkembang tanpa data kesehatan yang jelas, sehingga sulit untuk mendeteksi sebelum kondisi memburuk. Penyumbatan pada pembuluh darah jantung, yang disebabkan oleh penumpukan lemak, kolesterol, dan plak di dinding arteri, adalah penyebab utama penyakit jantung.

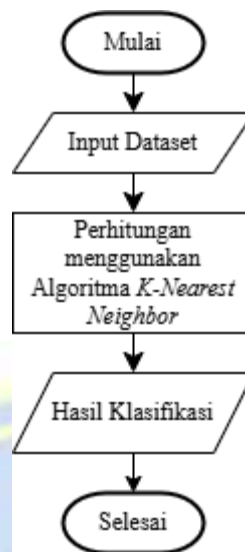
3.2. HASIL ANALISIS SISTEM

Analisis mengenai sistem prediksi risiko penyakit jantung menggunakan metode *K-Nearest Neighbor*. Penyakit jantung merupakan kondisi ketika jantung yang meliputi pembuluh jantung, selaput jantung, katup jantung, dan otot jantung mengalami gangguan. Penyakit jantung bisa disebabkan oleh berbagai hal, seperti sumbatan pembuluh darah jantung, peradangan, infeksi, atau kelainan bawaan. Setelah analisis sistem dilakukan, proses selanjutnya adalah perancangan sistem prediksi penyakit jantung dengan mengembangkan sistem yang dapat memprediksi penyakit jantung menggunakan data mining. Salah satu metode yang dapat digunakan adalah metode *K-Nearest Neighbor*. Dalam membuat penelitian ini peneliti mengambil data dari <https://www.kaggle.com/datasets/rcratos/heart-disease-data-compiled-from-uci?resource=download> untuk diolah, selanjutnya untuk mengetahui Accuracy yang baik akan dilakukan perhitungan nilai Accuracy.

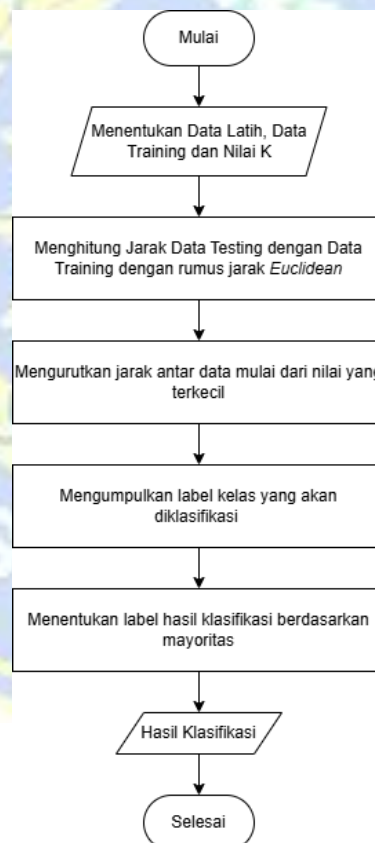
Diharapkan dari penelitian ini dapat memberikan informasi kepada pengguna mengenai prediksi penyakit jantung.

Pada penelitian ini, data yang digunakan adalah data dari <https://www.kaggle.com/datasets/rcratos/heart-disease-data-compiled-from-uci?resource=download>. data tersebut terdiri dari data kesehatan. Adapun variabel atau atribut masukan yang digunakan berjumlah 10 dan variabel keluaran berjumlah 2.

Metode data mining yang digunakan dalam penelitian ini adalah *metode K-Nearest Neighbor* (KNN). Algoritma KNN merupakan algoritma yang umum digunakan untuk klasifikasi, namun dapat juga digunakan untuk estimasi dan prediksi. Metode *K-Nearest Neighbors* merupakan model pembelajaran sederhana yang mempertahankan data pelatihan, memungkinkan klasifikasi data tidak terarah dibandingkan dengan informasi pelatihan yang disimpan. Nilai jarak inilah yang ditetapkan sebagai nilai kedekatan atau kemiripan antara data uji dengan data latih. Nilai K pada *K-Nearest Neighbor* (KNN) berarti K-data terdekat dari data uji (Argina 2020). Berdasarkan permasalahan yang ada, penulis akan membangun suatu sistem klasifikasi menggunakan algoritma *K-Nearest Neighbors* diharapkan dapat membantu pengguna memprediksi penyakit jantung melalui sistem ini dan dapat melihat serta mengetahui secara langsung tingkat risiko penyakit jantung yang didasarkan pada atribut yang telah ditetapkan pada sistem. *K-Nearest Neighbor* menawarkan fleksibilitas yang tinggi karena metode ini tidak memerlukan asumsi distribusi data yang ketat, sehingga dapat beradaptasi dengan baik terhadap berbagai jenis data, baik yang bersifat numerik atau kategorial, dan ini sangat menguntungkan dalam penelitian yang melibatkan dataset yang beragam dan kompleks, dimana karakteristik data tidak selalu memenuhi kriteria tertentu yang biasanya dibutuhkan oleh algoritma lain seperti C.45.



Gambar 3.1 Flowchart Sistem Klasifikasi Prediksi Risiko Penyakit Jantung



Gambar 3.2 Flowchart Algoritma *K-Nearest Neighbor*

Proses prediksi risiko penyakit jantung menggunakan algoritme *K-Nearest Neighbors* dilakukan dengan beberapa tahapan, diawali menghitung jarak dari data

testing terhadap data training menggunakan rumus jarak *Euclidean*. Proses selanjutnya melakukan pengurutan nilai jarak antar data dan mengelompokkan data yang akan diklasifikasikan berdasarkan nilai K, kemudian menentukan label hasil dari proses klasifikasi berdasarkan label mayoritas yang di dapat.

Pada penelitian ini, data yang digunakan adalah dataset *heart disease* yang terdiri dari 10 atribut data, dan target data yang akan digunakan adalah kolom *heart disease* yang berisi *presence* dan *absence*. *Presence* yang berarti seseorang yang diklasifikasikan berpotensi mengidap penyakit jantung, sedangkan *absence* adalah seseorang yang diklasifikasikan tidak berpotensi mengidap penyakit jantung, berikut adalah tabel dataset yang akan digunakan.

Tabel 3. 1 Atribut Data (Input)

No.	Name Field	Tipe Atribut	Keterangan
1.	Age	Numerik	Umur Pasien
2.	Sex	Numerik	Jenis Kelamin Pasien (1 = Laki-Laki, 0 = Perempuan).
3.	CPT	Numerik	Jenis Nyeri Dada (<i>typical angina, atypical angina, non-anginal pain, dan asymptomatic</i>). (0 = Data kesehatan umum nyeri dada dengan kemungkinan penyumbatan arteri koroner adalah <i>Typical Angina</i> . 1 = Data kesehatan tidak rinci, kemungkinann penyumbatan lebih rendah adalah <i>Atypical Angina</i> . 2 = Rasa menisik dan sakit dalam jangka waktu panjang atau pendek adalah <i>Non-anginal pain</i> . 3 = Tidak

			menunjukkan data kesehatan penyakit adalah <i>Asymptomatic</i>)
4.	Restbps	Numerik	Dalam satuan mm/Hg tekanan darah pasien ketika dalam kondisi istirahat.
5.	Chol	Numerik	Serum kolestrol dalam satuan mg/dl.
6.	Fbs	Numerik	(<i>Fasting Blood Sugar</i>) Besar gula darah dalam satuan mg/dl, lebih besar atau kurang dari 120mg/dl. (0 = kurang dari 120mg/dl, 1 = lebih dari 120mg/dl).
7.	Restecg	Numerik	Hasil elektrokardiografi istirahat. (0 = Normal, 1 = gelombang ST meningkat / menurun lebih dari 0,5 mV, 2 = Ventricular kiri mengalami hipertropi).
8.	Maxhr	Numerik	Detak jantung Maksimum.
9.	Exang	Numerik	Nyeri dada akibat olahraga (0 = tidak nyeri, 1 = nyeri).
10.	Oldpeak	Numerik	Besar segmen ST dari olahraga relatif terhadap kondisi istirahat.

Tabel 3. 2 Atribut Data (Output)

No.	Nama Field	Tipe Atribut	Keterangan
-----	------------	--------------	------------

1.	Kelas	Numerik	Risiko penyakit jantung. (0 = tidak terdapat risiko penyakit jantung, 1 = terdapat risiko penyakit jantung).
----	-------	---------	--

Jumlah data yang digunakan sebanyak 75 data training dan 25 data testing. Dimana setiap data training akan diuji terhadap seluruh data testing berdasarkan atribut yang nantinya mendapatkan hasil berupa nilai jarak kedekatan antar data untuk proses klasifikasi menggunakan jarak *Euclidean*.

3.3. REPRESENTASI MODEL

Penelitian ini menggunakan data kesehatan seperti *age*, *sex*, *CPT*, *restbps*, *chol*, *fbs*, *restecg*, *maxhr*, *exang*, dan *oldpeak* jumlah data yang digunakan berjumlah 100 data. Data tersebut diambil dari <https://www.kaggle.com/datasets/rcratos/heart-disease-data-compiled-from-uci?resource=download>. Dari data ke-100 data tersebut akan dibagi menjadi, data ke-1 sampai ke-75 akan dijadikan data latih, sedangkan data ke-76 sampai data ke-100 akan dijadikan data uji. Setelah pelatihan *K-Nearest Neighbor* selesai, selanjutnya dilakukan pengujian menggunakan perhitungan nilai Accuracy untuk mengukur seberapa baik sebuah sistem atau model dalam melakukan prediksi dengan membandingkan hasil yang dihasilkan, dengan hasil yang sebenarnya. Informasi mengenai data latih yang digunakan terdapat pada tabel 3.3 serta tabel 3.4 menggambarkan informasi data uji yang akan dilakukan.

Tabel 3. 3 Data Latih

No	Age	Sex	CPT	RestBps	Chol	Fbs	Restecg	Maxhr	Exang	Oldpeak	Kelas
1.	40	1	1	140	289	0	0	172	0	0	0
2.	49	0	2	160	180	0	0	156	0	1	1
3.	37	1	1	130	283	0	1	98	0	0	0

25.	49	1	2	140	187	0	0	172	0	0	0
-----	----	---	---	-----	-----	---	---	-----	---	---	---

Pada **Tabel 3.4** menjelaskan bahwa data uji yang digunakan untuk percobaan berdasarkan dari beberapa atribut atau variabel input yang digunakan seperti *age*, *sex*, *CPT*, *restbps*, *chol*, *fbs*, *restecg*, *maxhr*, *exang*, dan *oldpeak* dari data kesehatan yang digunakan sebagai target penelitian ini. Data untuk perhitungan data uji adalah data ke-76 hingga data ke-100. Berikut adalah proses perhitungan data latih dan data uji menggunakan metode *K-Nearest Neighbor*.

3.3.1. Proses Perhitungan Algoritma *K-Nearest Neighbor*

Berikut ini merupakan langkah-langkah perhitungan untuk memperoleh nilai jarak *Euclidean* label kelas pada data uji (tabel 3.6) terhadap data latih (tabel 3.5) :

Tabel 3. 5 Data Latih Perhitungan KNN

No	Age	Sex	CPT	RestBps	Chol	Fbs	Restecg	Maxhr	Exang	Oldpeak	Kelas
1.	40	1	1	140	289	0	0	172	0	0	0
2.	49	0	2	160	180	0	0	156	0	1	1
...	...	...	...	...	...	...	...	...	...	...	...
75.	55	1	3	140	268	0	0	128	1	1.5	1

Tabel 3. 6 Data Uji Perhitungan KNN

No	Age	Sex	CPT	RestBps	Chol	Fbs	Restecg	Maxhr	Exang	Oldpeak	Kelas
1.	54	1	3	130	603	1	0	125	1	1	1
2.	39	1	1	130	215	0	0	120	0	0	0
...	...	...	...	...	...	...	...	...	...	...	...
25.	49	1	2	140	187	0	0	172	0	0	0

Perhitungan data uji 1 terhadap data latih :

Menentukan nilai $K = 3$

Perhitungan data uji 1 terhadap data latih 1

$$d_{x1}d_{y1}$$

$$= \sqrt{(40 - 54)^2 + (1 - 1)^2 + (1 - 3)^2 + (140 - 130)^2 + (289 - 603)^2 + (0 - 1)^2 + (0 + 0)^2 + (172 - 125)^2 + (0 - 1)^2 + (0 - 1)^2}$$

$$= 317.974$$

Perhitungan data latih 2 terhadap data uji 1

$$d_{x1}d_{y1}$$

$$= \sqrt{(49 - 54)^2 + (0 - 1)^2 + (2 - 3)^2 + (160 - 130)^2 + (180 - 603)^2 + (0 - 1)^2 + (0 + 0)^2 + (156 - 125)^2 + (0 - 1)^2 + (0 - 1)^2}$$

$$= 425.228$$

Tabel 3. 7 Hasil Jarak *Euclidean* Data Uji 1 Terhadap data Latih 1-75

No	Jarak	Kelas	Peringkat
1	317.9748	0	13
2	425.2282	1	70
3	321.5991	0	17
4	389.5026	1	50
...	...	...	...
75	335.166	1	21

Pada tabel 3.7 dilakukan proses perhitungan dari data uji terhadap data latih untuk mendapatkan nilai jarak antar data. Selanjutnya dilakukan pengurutan peringkat data berdasarkan nilai jarak *Euclidean*.

Tabel 3. 8 Pengurutan Peringkat Nilai Jarak *Euclidean* Data Uji 1 Terhadap Jarak Data latih 1-75

No	Jarak	Kelas	Peringkat
31	86.4869	1	1
29	136.113	0	2
70	197.507	0	3

59	239.014	0	4
...	...	...	...
12	439.836	1	75

Pada **Tabel 3.8** dilakukan pengurutan peringkat data berdasarkan nilai jarak *euclidean*. Selanjutnya menentukan kelas dari hasil pengurutan nilai terkecil diperoleh dari **tabel 3.8**, ditunjukkan pada nomor 31, 29, dan 70. Pada hasil $K=3$ tersebut kelas asli yang menunjukkan negatif berada pada nomor 29 dan 70, sedangkan kelas asli yang menunjukkan positif berada pada nomor 31. Maka nilai mayoritas kelas terbanyak yang mendekati tetangga adalah “Negatif” hasil keseluruhan penentuan kelas dari 25 data *uji* yang diuji di paparkan dalam **Tabel 3.9** :

Tabel 3. 9 Klasifikasi KNN $K=3$

No	Age	Sex	CP T	Rest Bps	Chol	Fbs	Rest ecg	Max hr	Exang	Old peak	Kelas Asli	Hasil Prediksi
1.	54	1	3	130	603	1	0	125	1	1	1	0
2.	39	1	1	130	215	0	0	120	0	0	0	0
3.	48	1	1	100	159	0	0	100	0	0	0	1
4.	40	1	1	130	275	0	0	150	0	0	0	1
5.	55	1	3	120	270	0	0	140	0	0	0	1
6.	41	1	1	120	291	0	1	160	0	0	0	0
7.	56	1	3	155	342	1	0	150	1	3	1	0
8.	38	1	3	110	190	0	0	150	1	1	1	0
9.	49	1	3	140	185	0	0	130	0	0	0	1
10.	44	1	3	130	290	0	0	100	1	2	1	0
11.	54	1	1	160	195	0	1	130	0	1	0	0
12.	59	1	3	140	264	1	2	119	1	0	1	1
13.	49	1	3	128	212	0	0	96	1	0	1	1
14.	47	1	1	160	263	0	0	174	0	0	0	0
15.	42	1	1	120	196	0	0	150	0	0	0	0

16.	52	0	1	140	225	0	0	140	0	0	0	0
17.	46	1	0	140	272	1	0	175	0	2	1	0
18.	50	1	3	140	231	0	1	140	1	5	1	0
19.	48	1	1	140	238	0	0	118	0	0	0	0
20.	58	1	3	135	222	0	0	100	0	0	0	1
21.	58	1	2	140	179	0	0	160	0	0	0	0
22.	29	1	1	120	243	0	0	160	0	0	0	0
23.	40	1	2	140	235	0	0	188	0	0	0	0
24.	53	1	1	140	320	0	0	162	0	0	0	0
25.	49	1	2	140	187	0	0	172	0	0	0	0

Hasil perhitungan data *uji* keseluruhan terhadap data *latih* menggunakan algoritme *K-Nearest Neighbor* $K=3$ pada **Tabel 3.9** terdapat 7 pasien terdiagnosis positif penyakit jantung dan 18 pasien terdiagnosis negatif penyakit jantung. Hasil dari proses prediksi tersebut menunjukkan 14 data *uji* sesuai dengan hasil prediksi dan 11 data *uji* tidak sesuai dengan hasil prediksi (kolom berwarna merah).

Tabel 3. 10 Klasifikasi KNN $K=5$

No	Age	Sex	CP T	Rest Bps	Chol	Fbs	Rest ecg	Max hr	Exang	Old peak	Kelas Asli	Hasil Prediksi
1.	54	1	3	130	603	1	0	125	1	1	1	0
2.	39	1	1	130	215	0	0	120	0	0	0	0
3.	48	1	1	100	159	0	0	100	0	0	0	1
4.	40	1	1	130	275	0	0	150	0	0	0	1
5.	55	1	3	120	270	0	0	140	0	0	0	1
6.	41	1	1	120	291	0	1	160	0	0	0	0
7.	56	1	3	155	342	1	0	150	1	3	1	0
8.	38	1	3	110	190	0	0	150	1	1	1	0
9.	49	1	3	140	185	0	0	130	0	0	0	0

10.	44	1	3	130	290	0	0	100	1	2	1	0
11.	54	1	1	160	195	0	1	130	0	1	0	0
12.	59	1	3	140	264	1	2	119	1	0	1	1
13.	49	1	3	128	212	0	0	96	1	0	1	1
14.	47	1	1	160	263	0	0	174	0	0	0	0
15.	42	1	1	120	196	0	0	150	0	0	0	0
16.	52	0	1	140	225	0	0	140	0	0	0	0
17.	46	1	0	140	272	1	0	175	0	2	1	0
18.	50	1	3	140	231	0	1	140	1	5	1	0
19.	48	1	1	140	238	0	0	118	0	0	0	0
20.	58	1	3	135	222	0	0	100	0	0	0	1
21.	58	1	2	140	179	0	0	160	0	0	0	0
22.	29	1	1	120	243	0	0	160	0	0	0	0
23.	40	1	2	140	235	0	0	188	0	0	0	0
24.	53	1	1	140	320	0	0	162	0	0	0	0
25.	49	1	2	140	187	0	0	172	0	0	0	0

Hasil perhitungan data *uji* keseluruhan terhadap data *latih* menggunakan algoritme *K-Nearest Neighbor* $K=5$ pada **Tabel 3.11** terdapat 6 pasien terdiagnosis positif penyakit jantung dan 19 pasien terdiagnosis negatif penyakit jantung. Hasil dari proses prediksi tersebut menunjukkan 15 data *uji* sesuai dengan hasil prediksi dan 10 data *uji* tidak sesuai dengan hasil prediksi (kolom berwarna merah).

Tabel 3. 11 Klasifikasi KNN $K=7$

No	A ge	Se x	CP T	Rest Bps	Ch ol	Fb s	Rest ecg	Max hr	Exa ng	Old peak	Kelas Asli	Hasil Prediksi
1.	54	1	3	130	603	1	0	125	1	1	1	0
2.	39	1	1	130	215	0	0	120	0	0	0	0
3.	48	1	1	100	159	0	0	100	0	0	0	1

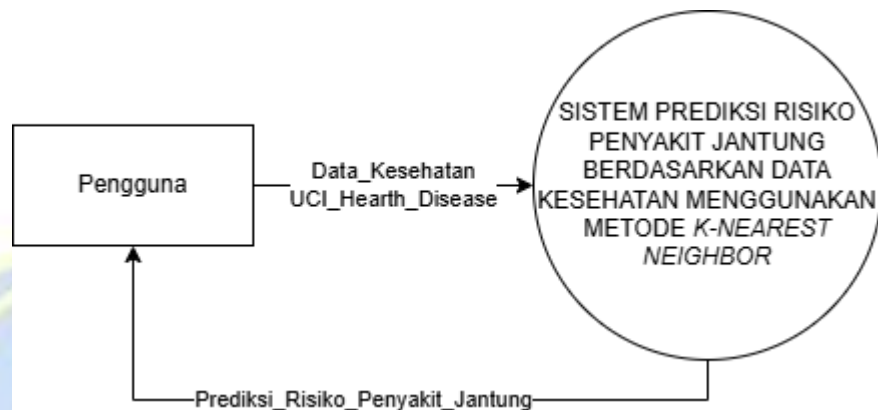
4.	40	1	1	130	275	0	0	150	0	0	0	1
5.	55	1	3	120	270	0	0	140	0	0	0	0
6.	41	1	1	120	291	0	1	160	0	0	0	1
7.	56	1	3	155	342	1	0	150	1	3	1	0
8.	38	1	3	110	190	0	0	150	1	1	1	0
9.	49	1	3	140	185	0	0	130	0	0	0	0
10.	44	1	3	130	290	0	0	100	1	2	1	0
11.	54	1	1	160	195	0	1	130	0	1	0	0
12.	59	1	3	140	264	1	2	119	1	0	1	0
13.	49	1	3	128	212	0	0	96	1	0	1	1
14.	47	1	1	160	263	0	0	174	0	0	0	0
15.	42	1	1	120	196	0	0	150	0	0	0	0
16.	52	0	1	140	225	0	0	140	0	0	0	0
17.	46	1	0	140	272	1	0	175	0	2	1	0
18.	50	1	3	140	231	0	1	140	1	5	1	0
19.	48	1	1	140	238	0	0	118	0	0	0	1
20.	58	1	3	135	222	0	0	100	0	0	0	1
21.	58	1	2	140	179	0	0	160	0	0	0	0
22.	29	1	1	120	243	0	0	160	0	0	0	0
23.	40	1	2	140	235	0	0	188	0	0	0	0
24.	53	1	1	140	320	0	0	162	0	0	0	0
25.	49	1	2	140	187	0	0	172	0	0	0	0

Hasil perhitungan data *uji* keseluruhan terhadap data *latih* menggunakan algoritme *K-Nearest Neighbor* $K=7$ pada **Tabel 3.11** terdapat 6 pasien terdiagnosis positif penyakit jantung dan 19 pasien terdiagnosis negatif penyakit jantung. Hasil dari proses prediksi tersebut menunjukkan 18 data *uji* sesuai dengan hasil prediksi dan 12 data *uji* tidak sesuai dengan hasil prediksi (kolom berwarna merah).

3.4. PERANCANGAN SISTEM

3.4.1. Diagram Konteks Sistem

Diagram ini menunjukkan entitas di luar atau terlibat dalam sistem, termasuk aktivitas dari data yang menghubungkan entitas sistem.



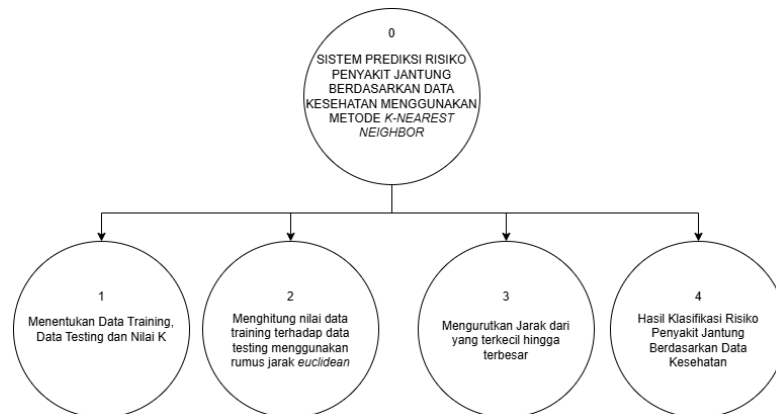
Gambar 3. 3 Diagram Konteks Sistem Prediksi Penyakit Jantung Berdasarkan Data Kesehatan

Penjelasan diagram konteks diatas adalah sebagai berikut:

1. Pengguna disini akan menginput data pasien dengan Data Kesehatan
2. Selanjutnya menginput data nilai diagnose pasien
3. Dari sistem akan di proses dan akan menghasilkan output hasil prediksi pasien Jantung dan Laporan proses perhitungan dengan menggunakan metode K-NN

3.4.2. *Hierarchy Process Diagram*

Hierarchy Process Diagram didefinisikan sebagai suatu representasi dari sebuah permasalahan yang kompleks. Dalam suatu struktur multi level dimana level pertama adalah tujuan, yang diikuti level faktor, kriteria, sub kriteria, dan seterusnya hingga level terakhir dari alternatif.

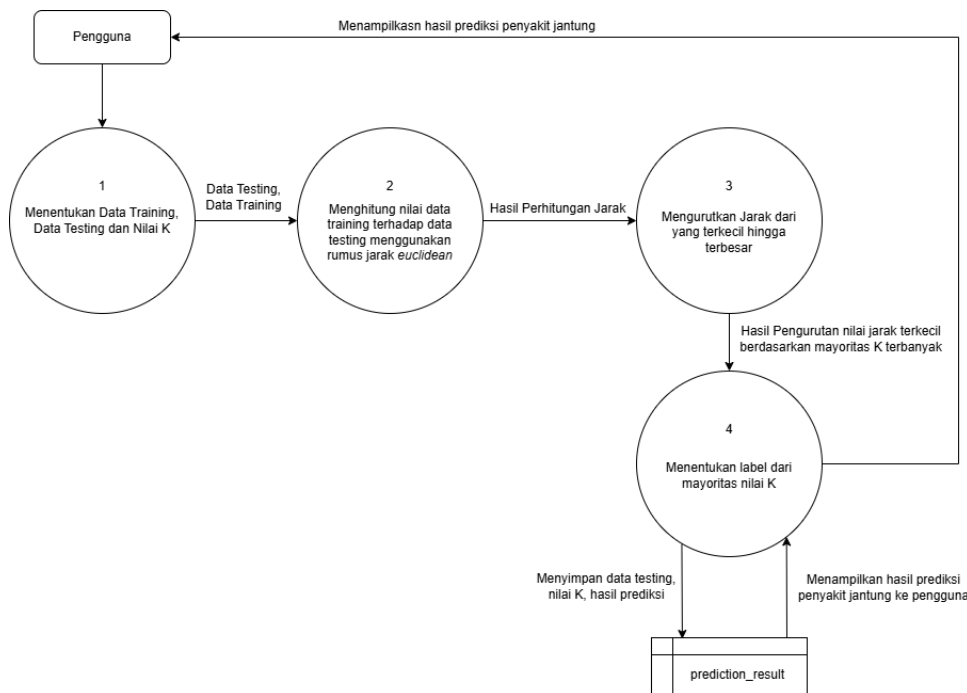


Gambar 3. 4 *Hierarchy Process Diagram* Sistem Prediksi

Pejelasan dari *Hierarchy Process Diagram* adalah sebagai berikut:

1. Level 0 adalah tingkatan paling atas yaitu sistem prediksi penyakit jantung berdasarkan data kesehatan menggunakan metode *K-Nearest Neighbor*.
2. Pada level 1 terdapat proses Menentukan Data Training, Data Testing, dan Nilai K.
3. Pada level 2 terdapat proses Menghitung nilai data training terhadap data testing menggunakan rumus jarak *euclidean*.
4. Pada level 3 terdapat proses mengurutkan dari nilai jarak yang terkecil hingga nilai jarak yang terbesar.
5. Pada level 4 terdapat proses hasil klasifikasi risiko penyakit jantung berdasarkan data kesehatan.

3.4.3. *Data Flow Diagram (DFD)* Sistem



Gambar 3. 5 Data Flow Diagram (DFD) Level 1 Sistem Prediksi Penyakit Jantung Berdasarkan Data Kesehatan

Pada **gambar 3.5** merupakan aktivitas dari DFD level 1 dengan tahapan proses dalam algoritme KNN yang terdapat pada sistem. Berawal dari pengguna menginputkan data kesehatan pasien berupa data *testing* beserta nilai K dan data *training*, kemudian ke proses menentukan nilai K, dan menghitung jarak dengan rumus *Euclidean Distance*, proses selanjutnya mengurutkan jarak dari nilai terkecil, dalam proses ini sebagai pengumpulan nilai mayoritas K terdekat dari nilai K yang ditentukan, proses tersebut disimpan dalam *database* *prediction_results* dan ditampilkan ke pengguna dalam berupa data kesehatan penyakit jantung pada pasien.

3.5. PERANCANGAN BASIS DATA

Perancangan basis data merupakan proses perancangan struktur penyimpanan data dari sistem yang akan digunakan dalam proses pengujian data pada algoritme *K-Nearest Neighbor*. Pada Sistem Prediksi Penyakit Jantung

penggunaan *database* terbagi dari tiga tabel yaitu tabel *users*, tabel *data_training*, dan tabel *prediction-result*. Berikut merupakan rincian dari tiga tabel sistem :

3.5.1. Tabel User

Tabel *users* pada *database* sistem klasifikasi digunakan untuk menyimpan data *users* untuk membuka laman klasifikasi sistem dengan memasukkan data pada halaman *login*. Berikut merupakan tabel dari *user* pada **Tabel 3.12** :

Tabel 3. 12 Tabel Pengguna Sistem Prediksi Penyakit Jantung

Name	Type	Deskripsi
_id	String	PK, Unique
username	String	
password	String	

3.5.2. Tabel Data_Training

Tabel *data_training* pada *database* sistem uji data kesehatan penyakit jantung digunakan untuk menyimpan data training yang digunakan perhitungan jarak menggunakan rumus *Euclidean Distance*. Berikut merupakan tabel dari *training* pada **Tabel 3.13** :

Tabel 3. 13 Tabel *data_training* Sistem Prediksi Penyakit Jantung

Name	Type	Deskripsi
_id	String	PK, Unique
feature1	INT	
feature2	INT	
feature3	INT	
feature4	INT	
feature5	INT	
feature6	INT	
feature7	INT	
feature8	INT	
feature9	INT	
feature10	INT	

label	String	
-------	--------	--

Sistem ini memiliki database yang menyimpan data-data yang diperlukan saat pengujian data menggunakan algoritme KNN. Tabel pengguna berhubungan dengan tabel data_training dan tabel prediction_results, tipe relasi dari tabel pengguna ke tabel data_training dan tabel prediction_results *adalah one to many*. Berikut merupakan gambaran hubungan setiap tabel yang tesaji dalam ERD pada **Gambar 3.6** :

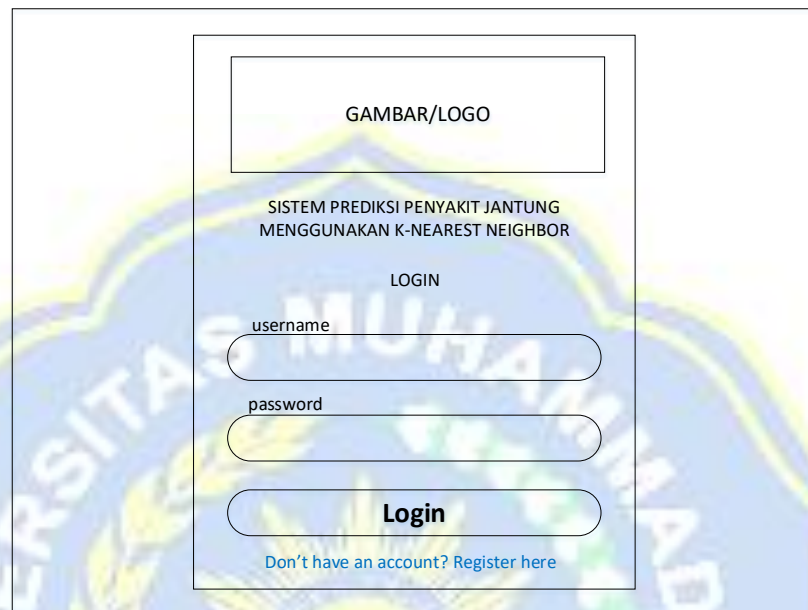


Gambar 3. 6 ERD (*Entity Relationship Diagram*)

3.6. PERANCANGAN ANTARMUKA SISTEM

3.6.1. Halaman Login

Halaman login merupakan tampilan awal dari sebuah sistem. Users pada sistem harus memasukkan username dan password yang sesuai untuk dapat masuk kedalam sistem. Tampilan rancangan halaman login terdapat pada **Gambar 3.7**.



The image shows a login form design overlaid on a large, faded watermark of the Universitas Muhammadiyah Gresik logo. The form is contained within a rectangular box and includes the following elements:

- A placeholder box at the top labeled "GAMBAR/LOGO".
- The system title: "SISTEM PREDIKSI PENYAKIT JANTUNG MENGGUNAKAN K-NEAREST NEIGHBOR".
- The section header: "LOGIN".
- A label "username" above a rounded rectangular input field.
- A label "password" above a rounded rectangular input field.
- A rounded rectangular button labeled "Login".
- A link at the bottom: "Don't have an account? Register here".

Gambar 3. 7 Halaman login

3.6.2. Halaman Register User

Dalam halaman *register user* menampilkan form input untuk mendaftarkan sebuah username dan password untuk dapat login dan menampilkan halaman klasifikasi. Berikut rancangan tampilan halaman Register **Gambar 3.8**.

REGISTER

username

password

Login

[Already have an account? Login here](#)

Gambar 3.8 Halaman Register

3.6.3. Halaman Data Klasifikasi

Halaman data klasifikasi memamparkan hasil data yang telah diklasifikasikan oleh algoritme KNN. Halaman ini menampilkan data berupa tabel yang berisikan data gejala yang telah di *input* oleh dokter beserta hasil yang diperoleh dari proses klasifikasi ditampilkan dalam halaman. Berikut rancangan tampilan halaman data klasifikasi pada gamar **Gambar 3.9**.

Home											Data Klasifikasi			Logout	
Data Preidiksi Penyakit Jantung															
_Id	usia asli	Usia (dinor)	sex	cpt	restbps	chol	fbs	restecg	maxhr	exang	oldpeak	prediksi	K	Akurasi	date

Gambar 3.9 Halaman Data Klasifikasi

3.6.4. Halaman Klasifikasi

Halaman Klasifikasi merupakan rancangan tampilan dari dokter menginputkan sebuah gejala untuk dilakukan proses klasifikasi dan pada halaman tersebut, sistem menampilkan hasil klasifikasi kepada pengguna. Berikut rancangan tampilan dari halaman klasifikasi pada **Gambar 3.10**.

		Home	Data Klasifikasi	Logout
Preidiksi Penyakit Jantung				
Usia		Jenis Kelamin	☐ Laki-Laki ☐ Perempuan	
Cpt		Exang :		
Restbps		Oldpeak :		
Chol		Nilai K	☐ 3 ☐ 5 ☐ 7 ☐ 9	
Fbs				
Restecg				
Maxhr				

Gambar 3.10 Halaman Hasil Klasifikasi

3.7. PERENCANAAN PENGUJIAN SISTEM

3.7.1. Evaluasi Performa *Confusion Matrix*

Confusion Matrix merupakan metode evaluasi yang digunakan dalam menghitung kinerja atau tingkat kebenaran dari proses klasifikasi. *Confusion Matrix* merupakan tabel dengan 4 kombinasi dari nilai prediksi dan nilai aktual yang berbeda. Dalam pengujian ini menggunakan data testing sejumlah 75 data dengan rentang nilai dari $K = 5$ dari Algoritme *K-Nearest Neighbor*.

Tabel 3.15 Evaluasi Performa *Confusion Matrix*

Three-Class Prediction		Predicted Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

$$Accuracy = \frac{P11+P22}{P11+P12+P21+P22} \quad (3.1)$$

$$Precision = \frac{TP1}{TP1+FP1} \quad (3.2)$$

$$Recall = \frac{TP1}{TP1+FN1} \quad (3.3)$$

