

## **BAB III**

### **ANALISIS DAN PERANCANGAN SISTEM**

#### **3.1. Spesifikasi Kebutuhan Sistem**

Spesifikasi Kebutuhan Sistem ini terdiri dari Kebutuhan Perangkat Keras dan Kebutuhan Perangkat Lunak.

##### **3.1.1. Kebutuhan Perangkat Keras**

Perangkat keras atau *hardware* yang digunakan pada pembuatan sistem ini antara lain :

- a) Processor Intel(R) Core(TM) i3-5005U
- b) RAM 4 GB
- c) SSD 512 GB
- d) Keyboard
- e) Mouse

##### **3.1.2. Kebutuhan Perangkat Lunak**

Perangkat lunak atau *Software* yang digunakan pada pembuatan sistem ini antara lain :

- a) Windows 10 Home
- b) Google Chrome
- c) Visual Studio Code
- d) PhpMyadmin versi 2.5.1
- e) Python versi 3.12.1
- f) Numpy versi 1.22.3
- g) Pandas versi 1.5.3
- h) Streamlit versi 1.29.0
- i) Scikit-learn versi 1.3.2
- j) Pip versi 23.2
- k) Matplotlib versi 3.8.2
- l) Mysql-connector-python versi 8.2.0
- m) Seaborn versi 0.13.1

### 3.2. Analisis Sistem

*Diabetes mellitus* adalah penyakit serius yang dapat menyebabkan komplikasi jangka panjang yang berkembang secara perlahan. Penyakit *diabetes mellitus* dapat memiliki dampak serius pada kesehatan seseorang jika tidak dideteksi secara dini. Dengan menggunakan *data mining klasifikasi*, kita dapat mengidentifikasi faktor-faktor risiko yang dapat membantu dalam mengklasifikasi penyakit ini dengan akurasi yang tinggi.

Berdasarkan hal tersebut, untuk meningkatkan upaya penurunan angka penyakit diabetes maka perlu dilakukan penelitian yang mengarah pada pembuatan sistem yang dapat mendeteksi timbulnya penyakit diabetes sehingga dapat dilakukan upaya pencegahan sejak dini. Dengan menggunakan sistem klasifikasi seperti yang diusulkan dalam penelitian ini, diharapkan dapat membantu individu dan profesional medis untuk mengambil tindakan pencegahan yang tepat dan lebih awal.

Analisis sistem ini dapat memberikan landasan untuk pengembangan sistem klasifikasi penyakit diabetes yang lebih efektif. Dengan memahami faktor-faktor risiko secara individual, peneliti dapat merancang rencana sistem klasifikasi penyakit diabetes yang lebih sesuai.

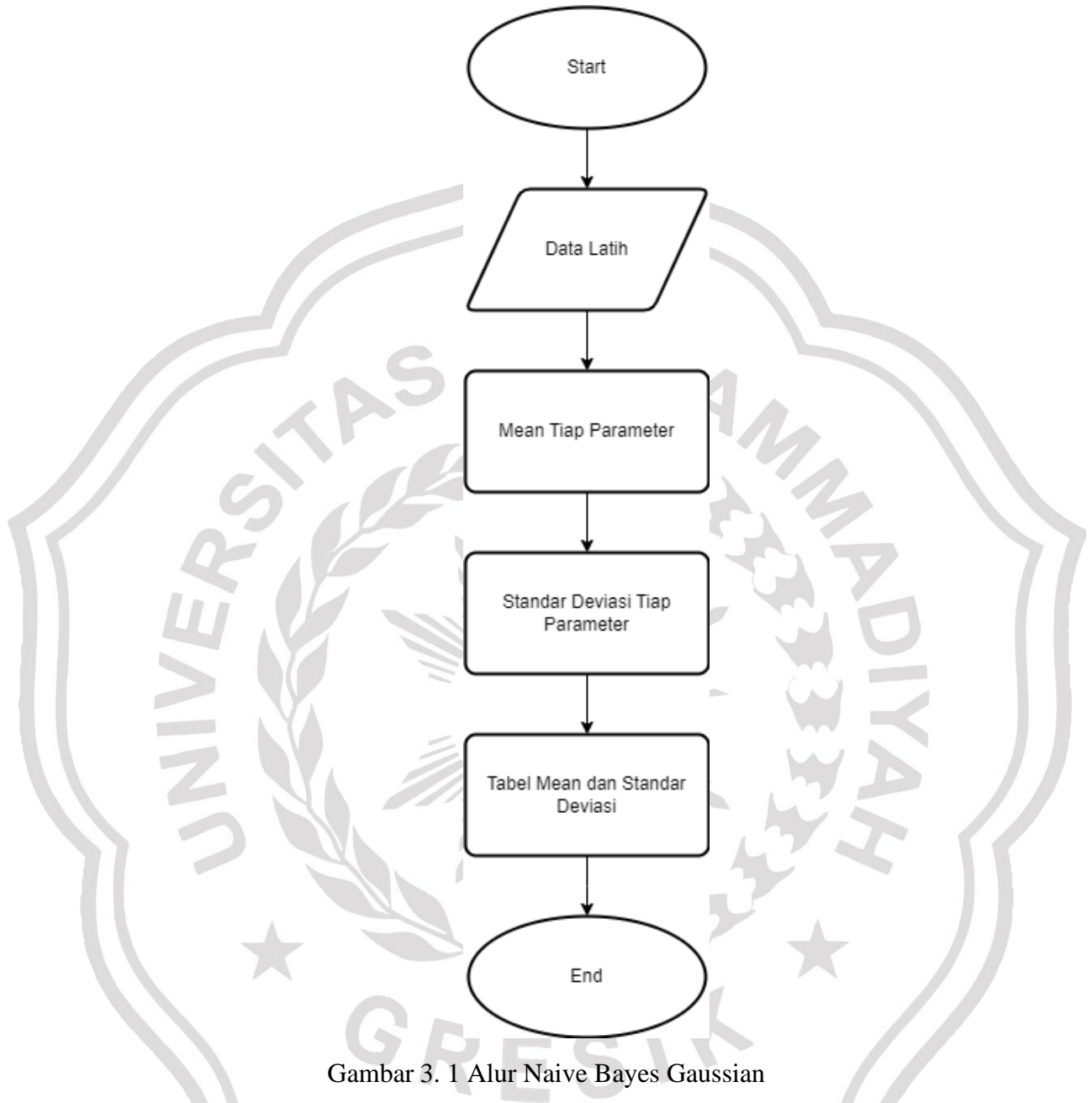
Penelitian ini menggunakan algoritme *Naïve Bayes* yang telah terbukti efektif dalam berbagai aplikasi *data mining* dan *klasifikasi*. Ini memungkinkan peneliti untuk memanfaatkan teknik yang kuat untuk melakukan analisis prediksi.

### 3.3. Perancangan Metode

Perancangan Metode ini terdiri dari Alur *Naïve Bayes Gaussian*, Alur *KNN*, Perhitungan *Naïve Bayes*, Perhitungan *KNN K=1* dan *K=3* dan Hasil Perbandingan Metode..

### 3.3.1. Alur Naïve Bayes Gaussian

Berikut adalah Alur *Naïve Bayes* data Numerik :



Gambar 3. 1 Alur Naive Bayes Gaussian

Adapun keterangan dari gambar diatas adalah sebagai berikut :

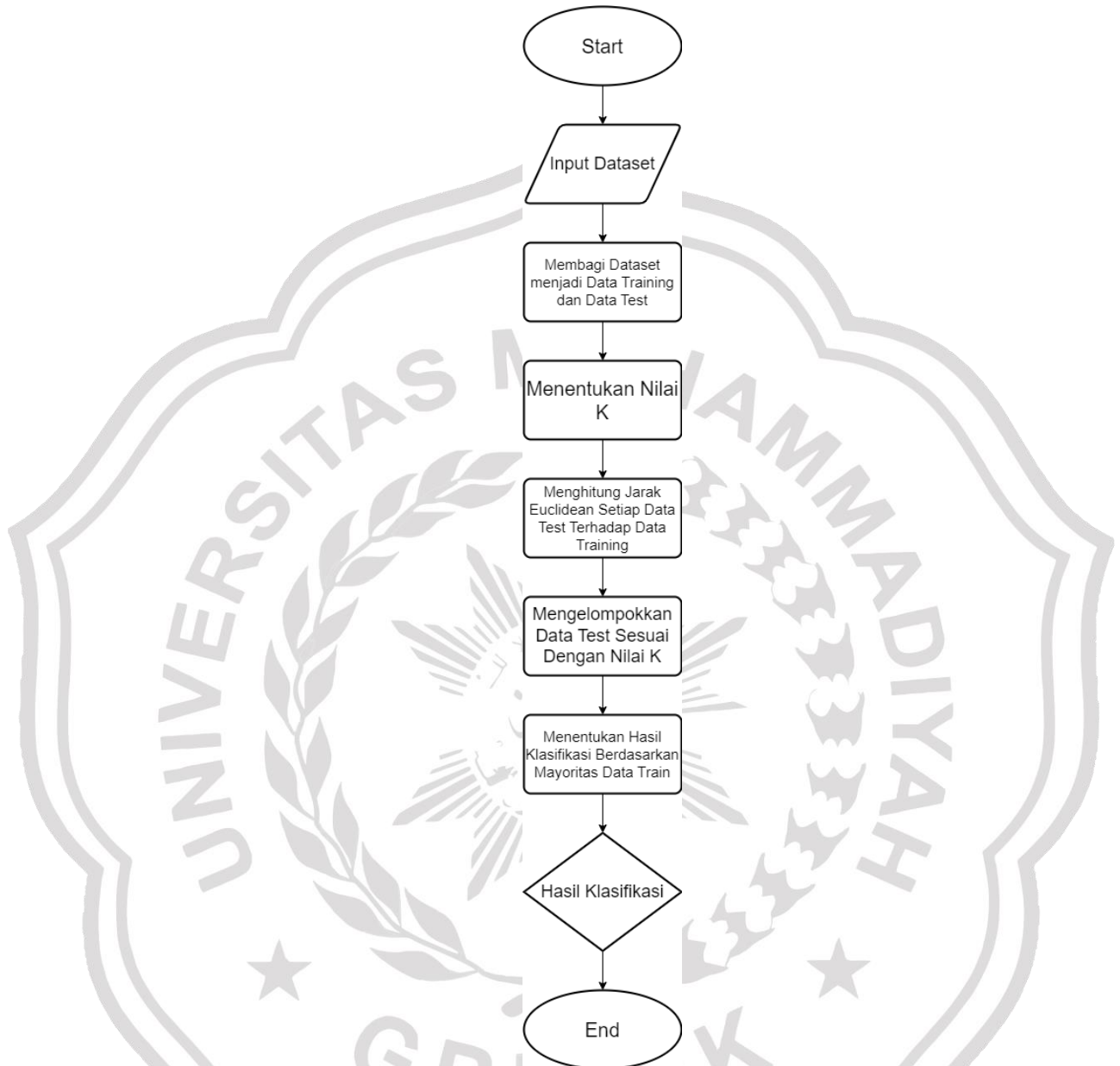
- 1) Start : pada tahap ini, akan dilakukan dengan pengumpulan data latih yang mencakup berbagai contoh dan kelas yang akan diprediksi.

- 2) Input data latih : Naïve bayes gaussian bekerja dengan memeriksa kemungkinan kelas yang mungkin terjadi berdasarkan atribut atau parameter yang ada dalam set data latih
- 3) Perhitungan mean tiap parameter : pada proses ini akan dilakukan perhitungan rata – rata (mean) dari setiap parameter dalam data latih
- 4) Perhitungan standar deviasi tiap parameter : pada proses ini akan dilakukan perhitungan standar deviasi yang berguna untuk mengukur sejauh mana nilai – nilai setiap parameter tersebar dari rata-ratanya.
- 5) Tabel mean dan standar deviasi : setelah menghitung nilai mean dan standar deviasi, hasil akan disusun dalam tabel yang menyajikan informasi tersebut untuk setiap parameter dan kela. Tabel ini nantinya akan digunakan untuk membuat prediksi pada data baru.
- 6) Selesai : dengan tabel yang sudah dibuat, model naïve bayes siap digunakan untuk memprediksi kelas dari data baru berdasarkan nilai probabilitas dan distribusi parameter yang telah dihitung.

Alur ini mencerminkan bagaimana Naïve Bayes digunakan dalam klasifikasi dengan data numeric, dimana perhitungan rata-rata (mean) dan standar deviasi digunakan untuk mengestimasi probabilitas kelas berdasarkan data numeric.

### 3.3.2. Alur KNN

Berikut adalah alur KNN



Gambar 3. 2 Alur KNN

Adapun keterangan dari gambar diatas adalah sebagai berikut :

- 1) Start : alur dimulai dengan memahami data yang digunakan.
- 2) Input dataset : data input adalah dataset yang telah dikumpulkan dan dipersiapn untuk analisis.
- 3) Pembagian dataset : data dibagi menjadi 2 yaitu data train dan data test.



- 4) Penentuan nilai K : pemilihan nilai K (jumlah tetangga terdekat) adalah Langkah penting dalam algoritme KNN. Nilai K yang tepat akan mempengaruhi akurasi.
- 5) Perhitungan jarak Euclidean : setiap data test diukur menggunakan jarak Euclidean terhadap setiap data train.
- 6) Pengelompokan data test : berdasarkan nilai K yang telah ditentukan, data testing dikelompokkan dengan cara menentukan nilai K dari data training terdekat.
- 7) Penentuan hasil klasifikasi : dengan menggunakan mayoritas kelas dari tetangga terdekat, model KNN akan menentukan hasil klasifikasi untuk setiap data testing
- 8) Hasil klasifikasi : setelah proses klasifikasi selesai, hasil klasifikasi akan muncul.
- 9) Selesai : dengan hasil klasifikasi yang diperoleh, alur KNN telah selesai.

Alur ini menunjukkan bahwa algoritme KNN adalah metode *klasifikasi* yang berbasis pada kesamaan dengan tetangga terdekat dalam ruang fitur. Proses ini memberikan pendekatan yang intuitif untuk memprediksi kelas data baru dengan mempertimbangkan mayoritas kelas dari tetangga terdekat.

### 3.3.3. Perhitungan Naïve Bayes

#### a. Persiapan Data

Tabel 3. 1 Data Train Perhitungan Manual

Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG	Diabetes
26	1	20,1	119	81	5,8	4,36	5,4	Negative
40	1	17,7	97	54	4,6	3,7	4,1	Negative
40	2	19,7	85	53	5,3	5,87	4,85	Negative
43	1	23,1	111	71	4,5	4,05	5,3	Negative
36	1	26,5	130	82	5,54	6,69	5,53	Negative
54	1	23,4	126	75	6,82	5	6,7	Positive
50	2	22,3	115	84	5,32	4,37	6,55	Positive
67	1	24,6	138	81	4,85	3,89	6,8	Positive
60	2	29	101	60	6,7	5,91	7	Positive
37	1	34,3	120	71	4,97	5,42	8,2	Positive

Tabel 3. 2 Tabel Penjelasan Parameter

	Nilai	Keterangan
Gender	1	Female
	2	Male
BMI (Body Mass Index)	$\leq 18,5$	Berat Badan Kurang
	18,5 – 24,9	Berat Badan Normal
	25 – 29,9	Berat Badan Berlebihan
	$\geq 30$	Obesitas
SBP (Systolic Blood Pressure)  Tekanan yang dicatat selama kontraksi jantung	$\leq 80$	Tekanan Darah Rendah
	80 – 120	Normal
	120 – 139	prahipertensi
	140 – 159	Tekanan Darah Tinggi (Hypertension Stadium 1)
	$\geq 160$	Tekanan Darah Tinggi (Hypertension Stadium 2)
	$\geq 180$	Krisis Tekanan Darah Tinggi
DBP (Diastolic Blood Pressure)  Tekanan yang dicatat selama relaksasi jantung	$\leq 60$	Tekanan Darah Rendah
	60 – 80	Normal
	80 – 89	prahipertensi
	90 – 99	Tekanan Darah Tinggi (Hypertension Stadium 1)
	$\geq 100$	Tekanan Darah Tinggi (Hypertension Stadium 2)
	$\geq 110$	Krisis Tekanan Darah Tinggi

FPG (Fasting Plasma Glucose) Kadar Glucose Puasa	$\leq 100$ mg/dL (5.55 mmol/L)	Normal
	100 mg/dL – 125 mg/dL (5.55 mmol/L – 6.9375 mmol/L)	Prediabetes
	$\geq 126$ mm/dL (7.003 mmol/L)	Diabetes
Cholesterol	$\leq 200$ mg/dL (5.17 mmol/L)	Normal
	200 – 239 mg/dL (5.17 – 6.18 mmol/L)	Borderline Tinggi
	$\geq 240$ mg/dL (6.21 mmol/L)	Tinggi
FFPG (Final Fasting Plasma Glucose) Kadar Glucose Dua Jam Setelah Makan	$\leq 100$ mg/dL (5.56 mmol/L)	Normal
	100 mg/dL – 125 mg/dL (5.56 mmol/L – 6.94 mmol/L)	Prediabetes
	$\geq 126$ mg/dL (7.00 mmol/L)	Diabetes

Pertama lakukan filter data berdasarkan kelasnya yaitu kelas Negative dan Positive. Dan buat tabel kelas Negative dan Positive untuk data yang sudah difilter.

Tabel 3. 3 Perhitungan Manual (Kelas Negative)

Jumlah Negative Diabetes							
Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG
26	1	20,1	119	81	5,8	4,36	5,4
40	1	17,7	97	54	4,6	3,7	4,1
40	2	19,7	85	53	5,3	5,87	4,85
43	1	23,1	111	71	4,5	4,05	5,3
36	1	26,5	130	82	5,54	6,69	5,53

Tabel 3. 4 Perhitungan Manual (Kelas Positive)

Jumlah Positive Diabetes							
Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG
54	1	23,4	126	75	6,82	5	6,7
50	2	22,3	115	84	5,32	4,37	6,55
67	1	24,6	138	81	4,85	3,89	6,8
60	2	29	101	60	6,7	5,91	7



37	1	34,3	120	71	4,97	5,42	8,2
----	---	------	-----	----	------	------	-----

b. Hitung Nilai Mean

Hitung nilai *mean* atau rata-rata di setiap atributnya berdasarkan kelasnya. Berikut adalah rumus mean :

$$\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$$

Dalam contoh ini kita akan menghitung nilai mean dari atribut Age dengan kondisi kelas Negative maka formulanya adalah

$$\text{Age} = \frac{26+40+40+43+36}{5} = \frac{185}{5} = 37.$$

Tabel 3. 5 Perhitungan Manual (Mean)

Mean								
Diabetes	Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG
Negative	37	1,2	21,42	108,4	68,2	5,148	4,934	5,036
Positive	53,6	1,4	26,72	120	74,2	5,732	4,918	7,05

c. Hitung Standar Deviasi

Hitung nilai *standar deviasi* dari masing-masing atribut berdasarkan kelasnya. Dengan rumus standar deviasi sebagai berikut :

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \mu)^2}{n}}$$

Hitung standar deviasi dari atribut Age dengan kondisi kelas Negative sebagai berikut:

1. Hitung selisih antara setiap data dengan rata-rata, kemudian kuadratkan selisih tersebut.

$$\text{Untuk } 26 = (26 - 37)^2 = 121$$

$$\text{Untuk } 40 = (40 - 37)^2 = 9$$

$$\text{Untuk } 40 = (40 - 37)^2 = 9$$

$$\text{Untuk } 43 = (43 - 37)^2 = 36$$

$$\text{Untuk } 36 = (36 - 37)^2 = 1$$

2. Hitung rata-rata dari selisih kuadrat yang dihitung di langkah 1

$$\text{Varians} = \frac{121 + 9 + 9 + 36 + 1}{5} = \frac{176}{5} = 35,2$$

3. Ambil akar kuadrat dari varians untuk mendapatkan standar deviasi ( $\sigma$ )

$$\text{Standar Deviasi } (\sigma) = \sqrt{35,2} = 5,932959$$

Tabel 3. 6 Perhitungan Manual (Standar Deviasi)

Standar Deviasi								
Diabetes	Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG
Negative	5,932 959	0,4	3,071 417	15,89 465	12,60 793	0,514 214	1,148 331	0,52094 5295
Positive	10,09 158	0,489 898	4,420 588	12,21 475	8,423 776	0,854 293	0,721 094	0,59329 5879

- d. Hitung Probabilitas Setiap Kelas

Probabilitas prior adalah probabilitas munculnya kelas tertentu tanpa memperhatikan atribut-atribut. Dalam hal ini, probabilitas prior untuk kelas Negative ada 5 dari total sampel yang ada, dan probabilitas prior untuk kelas Positive ada 5 dari total sampel yang ada. Berikut adalah perhitungan probabilitasnya :

- a) Probabilitas Kelas Negative

Jumlah data dengan kelas "Negative" = 5

Jumlah total data = 10

$$P(\text{Negative}) = \frac{5}{10} = 0,5$$

- b) Probabilitas Kelas Positive

Jumlah data dengan kelas "Positive" = 5

Jumlah total data = 10

$$P(\text{Positive}) = \frac{5}{10} = 0,5$$

Tabel 3. 7 Perhitungan Manual (Probabilitas)

Probabilitas Kelas	
Diabetes	Nilai
Negative	0,5
Positive	0,5

- e. Hitung Nilai Gaussiannya

Distribusi Gaussian adalah langkah akhir untuk mendapatkan hasil dari data latih atau model pengujian data dengan memperoleh nilai probabilitas dari data latih. Berikut adalah rumusnya :

$$P(X_i = x_i | Y = y_i) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

Berikut adalah contoh perhitungan dari data yang belum diketahui kelasnya:

Tabel 3. 8 Perhitungan Manual (Data Test)

Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG	Diabetes
36	1	23,4	113	72	5,13	4,14	4,9	?

Pertama hitung dulu **distribusi Gaussian** dari masing-masing kelas yaitu kelas **Negative** dan **Positive**. Dalam contoh ini akan menunjukkan cara menghitung nilai **distribusi Gaussian** dari atribut **Age Kelas Negative**. Berikut adalah rumusnya :

$$P(X) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$P(X) = \frac{1}{\sqrt{2 \times 3,14 \times 5,932959^2}} e^{-\frac{(36-37)^2}{2 \times 5,932959^2}}$$

$$P(X) = \frac{1}{\sqrt{37,25898252}} e^{-\frac{1}{70,4000049914}}$$

$$P(X) = \frac{1}{6,1040136402} e^{-0,0142045444}$$

$$P(X) = 0,1615159995$$

Terapkan rumus ini untuk setiap atribut. Gantilah distribusi Gaussian untuk kelas Positive dengan nilai kelas Positive.

Tabel 3. 9 Perhitungan Manual (Hasil)

	Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG	Diabetes
Diabetes	36	1	23,4	113	72	5,13	4,14	4,9	?
Negative	0,161516	0,5568053	0,184973	0,095986	0,107392	0,556137	0,293205	0,534348	7,4706
Positive	0,027451	0,4085099	0,143152	0,096886	0,132879	0,336813	0,262574	0,000729	6,663

Max adalah 7,4706

Setelah menghitung semua atribut dari **Age** sampai **FFPG**, maka langkah terakhir adalah mengalikan semua variabel dengan nilai probabilitas dari masing-masing kelas.

Berikut adalah rumusnya :  
 $0,161516 \times 0,5568053 \times 0,184973 \times 0,095986 \times 0,107392 \times 0,556137 \times 0,293205 \times 0,534348 \times 0,5 = 7,4706$ . Kemudian jika kelas **Positive** maka ganti dengan **nilai probabilitas kelas Positive**.

Kemudian cari nilai **Maximalnya** antara kelas Positive dan Negative yaitu (7,4706 dan 6,663) maka nilai terbesar adalah 7,4706. Sehingga, Pasien dengan Age = 36, Gender = 1, BMI = 23,4 , SBP = 113, DBP = 72, FPG = 5,13 , Chol = 4,14 dan FFPG = 4,9 diprediksi oleh algoritme Naïve Bayes dan hasil kelasnya adalah **Negative**.

Berikut adalah hasil dari 20 data testing :

Tabel 3. 10 Hasil 20 Data Testing

No	Id	Age	Gender	BMI	SBP	DBP	FPG	Cho l	FFPG	True Label	Predict Label
1	2811	36	1	23.40	113	72	5.13	4.14	4.90	0	0
2	1818	36	2	22.30	113	72	4.23	4.72	5.23	0	0
3	1808	27	1	19.30	112	68	4.03	3.21	4.70	0	0
4	271	53	1	24.20	133	85	3.76	5.02	4.90	0	0
5	474	33	1	22.40	110	62	5.41	3.08	5.00	0	0
6	2972	53	2	24.24	155	77	4.57	5.01	5.58	0	0
7	1607	36	1	24.50	116	68	4.61	3.59	4.40	0	0
8	3390	71	1	25.20	143	91	5.00	4.00	7.10	1	1
9	979	29	2	17.80	113	72	4.01	4.52	4.00	0	0
10	1493	38	1	32.60	121	72	5.60	4.94	5.49	0	0
11	3957	66	1	25.90	100	61	4.92	4.10	8.00	1	1
12	3699	57	1	21.80	108	65	5.40	4.79	5.00	1	0
13	4102	47	1	35.00	125	69	5.04	4.97	6.70	1	1
14	552	31	1	24.20	125	81	4.80	4.58	6.10	0	0
15	1814	35	2	23.90	106	63	4.78	4.72	4.90	0	0
16	1428	80	1	24.00	177	85	5.27	5.67	5.90	0	1
17	3260	62	1	24.50	122	76	6.90	5.27	6.40	1	1
18	2441	36	2	20.70	112	61	4.70	4.10	4.63	0	0
19	3701	49	1	25.00	132	93	6.09	4.26	7.30	1	1
20	2522	31	1	16.50	123	69	4.30	3.78	4.87	0	0

Terdapat 2 error yaitu pada nomor 12 dan 16, berikut adalah penjelasannya :

- Nomor 12 True Label menunjukkan hasil Positive, sedangkan hasil Predict Label menunjukkan hasil Negative.
- Nomor 16 True Label menunjukkan hasil Negative, sedangkan hasil Predict Label menunjukkan hasil Positive.

Tabel 3. 11 Confusion Matrix Hasil Perhitungan Naive Bayes

True Label	Negative	13	1
	Positive	1	5
		Negative	Positive
		Predict Label	

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{5+13}{5+13+1+1} = \frac{18}{20} = 0,9$$

$$\text{Laju error} = \frac{FP+FN}{FP+FN+TP+TN} = \frac{1+1}{1+1+5+13} = \frac{2}{20} = 0,1$$

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{5}{5+1} = \frac{5}{6} = 0,833$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{5}{5+1} = \frac{5}{6} = 0,833$$

### 3.3.4. Perhitungan KNN K1

- Persiapan Data

Tabel 3. 12 Data Train Perhitungan Manual KNN K1

No	Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG	Diabetes
1	2	1	20,1	119	81	5,8	4,36	5,4	Negative
2	2	1	17,7	97	54	4,6	3,7	4,1	Negative
3	2	2	19,7	85	53	5,3	5,87	4,85	Negative
4	2	1	23,1	111	71	4,5	4,05	5,3	Negative
5	2	1	26,5	130	82	5,54	6,69	5,53	Negative
6	2	1	23,4	126	75	6,82	5	6,7	Positive
7	2	2	22,3	115	84	5,32	4,37	6,55	Positive
8	3	1	24,6	138	81	4,85	3,89	6,8	Positive
9	2	2	29	101	60	6,7	5,91	7	Positive
10	2	1	34,3	120	71	4,97	5,42	8,2	Positive
11	2	1	23,4	113	72	5,13	4,14	4,9	?



Tabel 3. 13 Penjelasan Parameter

	Nilai	Keterangan
Age	1	0 – 14 tahun
	2	15 – 65 tahun
	3	≥ 65 tahun
Gender	1	Female
	2	Male
BMI (Body Mass Index)	≤ 18,5	Berat Badan Kurang
	18,5 – 24,9	Berat Badan Normal
	25 – 29,9	Berat Badan Berlebihan
	≥ 30	Obesitas
SBP (Systolic Blood Pressure) Tekanan yang dicatat selama kontraksi jantung	≤ 80	Tekanan Darah Rendah
	80 – 120	Normal
	120 – 139	prahipertensi
	140 – 159	Tekanan Darah Tinggi (Hypertension Stadium 1)
	≥ 160	Tekanan Darah Tinggi (Hypertension Stadium 2)
	≥ 180	Krisis Tekanan Darah Tinggi
DBP (Diastolic Blood Pressure) Tekanan yang dicatat selama relaksasi jantung	≤ 60	Tekanan Darah Rendah
	60 – 80	Normal
	80 – 89	prahipertensi
	90 – 99	Tekanan Darah Tinggi (Hypertension Stadium 1)
	≥ 100	Tekanan Darah Tinggi

		(Hypertension Stadium 2)
	$\geq 110$	Krisis Tekanan Darah Tinggi
FPG (Fasting Plasma Glucose) Kadar Glucose Puasa	$\leq 100$ mg/dL (5.55 mmol/L)	Normal
	100 mg/dL – 125 mg/dL (5.55 mmol/L – 6.9375 mmol/L)	Prediabetes
	$\geq 126$ mm/dL (7.003 mmol/L)	Diabetes
Cholesterol	$\leq 200$ mg/Dl (5.17 mmol/L)	Normal
	200 – 239 mg/dL (5.17 – 6.18 mmol/L)	Borderline Tinggi
	$\geq 240$ mg/Dl (6.21 mmol/L)	Tinggi
FFPG (Final Fasting Plasma Glucose) Kadar Glucose Dua Jam Setelah Makan	$\leq 100$ mg/dL (5.56 mmol/L)	Normal
	100 mg/dL – 125 mg/dL (5.56 mmol/L – 6.94 mmol/L)	Prediabetes
	$\geq 126$ mg/dL (7.00 mmol/L)	Diabetes

b. Normalisasi data

Langkah selanjutnya yaitu normalisasi dataset diatas menggunakan rumus min-max sebagai berikut :

$$X_{new} = \frac{X_{old} - X_{min}}{X_{max} - X_{min}}$$

Dalam contoh ini kita akan menghitung normalisasi data menggunakan min-max dari atribut SBP, maka formulanya adalah :

$$SBP = \frac{119 - 85}{138 - 85} = \frac{34}{53} = 0,641$$

Tabel 3. 14 Hasil Normalisasi Data K1

No	Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG	Diabetes
1	0,000	0	0,1446	0,6415	0,9032	0,5603	0,2207	0,3171	Negative
2	0,000	0	0,0000	0,2264	0,0323	0,0431	0,0000	0,0000	Negative

3	0,000	1	0,1205	0,0000	0,0000	0,3448	0,7258	0,1829	Negative
4	0,000	0	0,3253	0,4906	0,5806	0,0000	0,1171	0,2927	Negative
5	0,000	0	0,5301	0,8491	0,9355	0,4483	1,0000	0,3488	Negative
6	0,000	0	0,3434	0,7736	0,7097	1,0000	0,4348	0,6341	Positive
7	0,000	1	0,2771	0,5660	1,0000	0,3534	0,2241	0,5976	Positive
8	1,000	0	0,4157	1,0000	0,9032	0,1509	0,0635	0,6585	Positive
9	0,000	1	0,6807	0,3019	0,2258	0,9483	0,7391	0,7073	Positive
10	0,000	0	1,0000	0,6604	0,5806	0,2026	0,5753	1,0000	Positive
11	0,000	0	0,3434	0,5283	0,6129	0,2716	0,1472	0,1951	?

c. Tentukan Nilai K

Penentuan nilai K ini tidak ada rumus pastinya. Namun apabila kelas berjumlah genap maka sebaiknya nilai K-nya ganjil, sebaliknya jika kelas berjumlah ganjil maka sebaiknya nilai K-nya genap. Seperti contoh di atas ada dua kelas genap yaitu (Negative dan Positive), maka untuk nilai K nya akan tentukan jadi 1.

d. Hitung jarak antara data test dan data train menggunakan rumus Euclidean sebagai berikut :

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$= \sqrt{((0-0)^2 + (0-0)^2 + (0,343 - 0,1446)^2 + (0,5283 - 0,6415)^2 + (0,6129 - 0,9032)^2 + (0,2716 - 0,5603)^2 + (0,1472 - 0,2207)^2 + (0,1951 - 0,3171)^2)}$$

$$= \sqrt{0,4902}$$

Tabel 3. 15 Hasil Perhitungan Jarak Euclidean K1

No	Kelas asli	Jarak data test ke data train	Urutan Nilai Terkecil
1	Negative	0,4902	2
2	Negative	0,8112	3
3	Negative	1,4299	9
4	Negative	0,2949	1
5	Negative	1,0119	5
6	Positive	0,9358	4
7	Positive	1,1534	7

8	Positive	1,2443	8
9	Positive	1,5446	10
10	Positive	1,1338	6

Setelah itu pilih satu data dengan nilai urutan terkecil, yaitu pada data dengan nomor 4.

Tabel 3. 16 Hasil Penentuan Data Menggunakan K1

No	Kelas asli	Jarak data test ke data train	Urutan Nilai Terkecil
4	Negative	0,2949	1

Dari hasil diatas diketahui nilai  $K = 1$  berada pada nomor 4 dengan nilai jarak data test ke data train sebesar 0,2949 dan kelas asli "Negative".

Sehingga, Pasien dengan Age = 2, Gender = 1, BMI = 23,4 , SBP = 113, DBP = 72, FPG = 5,13 , Chol = 4,14 dan FFPG = 4,9 diklasifikasikan oleh algoritme KNN dan hasil kelasnya adalah **Negative**.

Tabel 3. 17 Hasil 20 Data Testing KNN K1

No	Id	Age	Gender	BMI	SBP	DBP	FPG	Cho l	FFPG	True Label	Predict Label
1	2811	36	1	23.40	113	72	5.13	4.14	4.90	0	0
2	1818	36	2	22.30	113	72	4.23	4.72	5.23	0	0
3	1808	27	1	19.30	112	68	4.03	3.21	4.70	0	0
4	271	53	1	24.20	133	85	3.76	5.02	4.90	0	0
5	474	33	1	22.40	110	62	5.41	3.08	5.00	0	0
6	2972	53	2	24.24	155	77	4.57	5.01	5.58	0	0
7	1607	36	1	24.50	116	68	4.61	3.59	4.40	0	0
8	3390	71	1	25.20	143	91	5.00	4.00	7.10	1	1
9	979	29	2	17.80	113	72	4.01	4.52	4.00	0	0
10	1493	38	1	32.60	121	72	5.60	4.94	5.49	0	0
11	3957	66	1	25.90	100	61	4.92	4.10	8.00	1	1
12	3699	57	1	21.80	108	65	5.40	4.79	5.00	1	0
13	4102	47	1	35.00	125	69	5.04	4.97	6.70	1	0
14	552	31	1	24.20	125	81	4.80	4.58	6.10	0	0
15	1814	35	2	23.90	106	63	4.78	4.72	4.90	0	0
16	1428	80	1	24.00	177	85	5.27	5.67	5.90	0	0
17	3260	62	1	24.50	122	76	6.90	5.27	6.40	1	1
18	2441	36	2	20.70	112	61	4.70	4.10	4.63	0	0
19	3701	49	1	25.00	132	93	6.09	4.26	7.30	1	0
20	2522	31	1	16.50	123	69	4.30	3.78	4.87	0	0

Terdapat 3 error yaitu pada nomor 12, 13 dan 19, berikut adalah penjelasannya :

- Nomor 12 True Label menunjukkan hasil Positive, sedangkan hasil Predict Label menunjukkan hasil Negative.
- Nomor 13 True Label menunjukkan hasil Positive, sedangkan hasil Predict Label menunjukkan hasil Negative.
- Nomor 19 True Label menunjukkan hasil Positive, sedangkan hasil Predict Label menunjukkan hasil Negative.

Tabel 3. 18 Confusion Matrix Hasil Perhitungan KNN K1

True Label	Negative	14	0
	Positive	3	3
		Negative	Positive
		Predict Label	

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{3+14}{3+14+3+0} = \frac{17}{20} = 0,85$$

$$\text{Laju error} = \frac{FP+FN}{FP+FN+TP+TN} = \frac{3+0}{3+0+3+14} = \frac{3}{20} = 0,15$$

$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{3}{3+3} = \frac{3}{6} = 0,5$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{3}{3+0} = \frac{3}{3} = 1$$

### 3.3.5 Perhitungan KNN K3

- Persiapan Data

Tabel 3. 19 Data Train Perhitungan Manual KNN K3

No	Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG	Diabetes
1	2	1	20,1	119	81	5,8	4,36	5,4	Negative
2	2	1	17,7	97	54	4,6	3,7	4,1	Negative
3	2	2	19,7	85	53	5,3	5,87	4,85	Negative
4	2	1	23,1	111	71	4,5	4,05	5,3	Negative
5	2	1	26,5	130	82	5,54	6,69	5,53	Negative
6	2	1	23,4	126	75	6,82	5	6,7	Positive
7	2	2	22,3	115	84	5,32	4,37	6,55	Positive
8	3	1	24,6	138	81	4,85	3,89	6,8	Positive
9	2	2	29	101	60	6,7	5,91	7	Positive
10	2	1	34,3	120	71	4,97	5,42	8,2	Positive



11	2	1	23,4	113	72	5,13	4,14	4,9	?
----	---	---	------	-----	----	------	------	-----	---

Tabel 3. 20 Penjelasan Parameter

	Nilai	Keterangan
Age	1	0 – 14 tahun
	2	15 – 65 tahun
	3	≥ 65 tahun
Gender	1	Female
	2	Male
BMI (Body Mass Index)	≤ 18,5	Berat Badan Kurang
	18,5 – 24,9	Berat Badan Normal
	25 – 29,9	Berat Badan Berlebihan
	≥ 30	Obesitas
SBP (Systolic Blood Pressure) Tekanan yang dicatat selama kontraksi jantung	≤ 80	Tekanan Darah Rendah
	80 – 120	Normal
	120 – 139	prahipertensi
	140 – 159	Tekanan Darah Tinggi (Hypertension Stadium 1)
	≥ 160	Tekanan Darah Tinggi (Hypertension Stadium 2)
DBP (Diastolic Blood Pressure) Tekanan yang dicatat selama relaksasi jantung	≤ 60	Tekanan Darah Rendah
	60 – 80	Normal
	80 – 89	prahipertensi
	90 – 99	Tekanan Darah Tinggi (Hypertension Stadium 1)

	$\geq 100$	Tekanan Darah Tinggi (Hypertension Stadium 2)
	$\geq 110$	Krisis Tekanan Darah Tinggi
FPG (Fasting Plasma Glucose) Kadar Glucose Puasa	$\leq 100$ mg/dL (5.55 mmol/L)	Normal
	100 mg/dL – 125 mg/dL (5.55 mmol/L – 6.9375 mmol/L)	Prediabetes
	$\geq 126$ mm/dL (7.003 mmol/L)	Diabetes
Cholesterol	$\leq 200$ mg/dL (5.17 mmol/L)	Normal
	200 – 239 mg/dL (5.17 – 6.18 mmol/L)	Borderline Tinggi
	$\geq 240$ mg/dL (6.21 mmol/L)	Tinggi
FFPG (Final Fasting Plasma Glucose) Kadar Glucose Dua Jam Setelah Makan	$\leq 100$ mg/dL (5.56 mmol/L)	Normal
	100 mg/dL – 125 mg/dL (5.56 mmol/L – 6.94 mmol/L)	Prediabetes
	$\geq 126$ mg/dL (7.00 mmol/L)	Diabetes

b. Normalisasi data

Langkah selanjutnya yaitu normalisasi dataset diatas menggunakan rumus min-max sebagai berikut :

$$X_{new} = \frac{X_{old} - X_{min}}{X_{max} - X_{min}}$$

Dalam contoh ini kita akan menghitung normalisasi data menggunakan min-max dari atribut SBP, maka formulanya adalah :

$$SBP = \frac{119 - 85}{138 - 85} = \frac{34}{53} = 0,641$$

Tabel 3. 21 Hasil Normalisasi Data K3

No	Age	Gender	BMI	SBP	DBP	FPG	Chol	FFPG	Diabetes
1	0,0000	0,0000	0,1446	0,6415	0,9032	0,5603	0,2207	0,3171	Negative
2	0,0000	0,0000	0,0000	0,2264	0,0323	0,0431	0,0000	0,0000	Negative
3	0,0000	1,0000	0,1205	0,0000	0,0000	0,3448	0,7258	0,1829	Negative
4	0,0000	0,0000	0,3253	0,4906	0,5806	0,0000	0,1171	0,2927	Negative
5	0,0000	0,0000	0,5301	0,8491	0,9355	0,4483	1,0000	0,3488	Negative
6	0,0000	0,0000	0,3434	0,7736	0,7097	1,0000	0,4348	0,6341	Positive
7	0,0000	1,0000	0,2771	0,5660	1,0000	0,3534	0,2241	0,5976	Positive
8	1,0000	0,0000	0,4157	1,0000	0,9032	0,1509	0,0635	0,6585	Positive
9	0,0000	1,0000	0,6807	0,3019	0,2258	0,9483	0,7391	0,7073	Positive
10	0,0000	0,0000	1,0000	0,6604	0,5806	0,2026	0,5753	1,0000	Positive
11	0,0000	0,0000	0,3434	0,5283	0,6129	0,2716	0,1472	0,1951	?

## c. Tentukan Nilai K

Penentuan nilai K ini tidak ada rumus pastinya. Namun apabila kelas berjumlah genap maka sebaiknya nilai K-nya ganjil, sebaliknya jika kelas berjumlah ganjil maka sebaiknya nilai K-nya genap. Seperti contoh di atas ada dua kelas genap yaitu (Negative dan Positive), maka untuk nilai K nya akan ditentukan jadi 3.

## d. Hitung jarak antara data test dan data train menggunakan rumus Euclidean sebagai berikut :

$$d = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}$$

$$= \sqrt{((0-0)^2 + (0-0)^2 + (0,343 - 0,1446)^2 + (0,5283 - 0,6415)^2 + (0,6129 - 0,9032)^2 + (0,2716 - 0,5603)^2 + (0,1472 - 0,2207)^2 + (0,1951 - 0,3171)^2)}$$

$$= \sqrt{0,4902}$$

Tabel 3. 22 Hasil Perhitungan Jarak Euclidean K3

No	Kelas asli	Jarak data test ke data train	Urutan Nilai Terkecil
1	Negative	0,4902	2
2	Negative	0,8112	3

3	Negative	1,4299	9
4	Negative	0,2949	1
5	Negative	1,0119	5
6	Positive	0,9358	4
7	Positive	1,1534	7
8	Positive	1,2443	8
9	Positive	1,5446	10
10	Positive	1,1338	6

Setelah itu pilih 3 data dengan nilai urutan terkecil, yaitu pada data dengan nomor 4, 1 dan 2.

Tabel 3. 23 Hasil Penentuan Data Menggunakan K3

No	Kelas asli	Jarak data test ke data train	Urutan Nilai Terkecil
4	Negative	0,2949	1
1	Negative	0,4902	2
2	Negative	0,8112	3

Dari hasil diatas diketahui kelas mayoritas dari 3 data diatas adalah "Negaive". Sehingga, Pasien dengan Age = 2, Gender = 1, BMI = 23,4 , SBP = 113, DBP = 72, FPG = 5,13 , Chol = 4,14 dan FFPG = 4,9 diklasifikasikan oleh algoritme KNN dan hasil kelasnya adalah **Negative**.

Tabel 3. 24 Hasil 20 Data Testing KNN K3

No	Id	Age	Gender	BMI	SBP	DBP	FPG	Cho l	FFPG	True Label	Predict Label
1	2811	36	1	23.40	113	72	5.13	4.14	4.90	0	0
2	1818	36	2	22.30	113	72	4.23	4.72	5.23	0	0
3	1808	27	1	19.30	112	68	4.03	3.21	4.70	0	0
4	271	53	1	24.20	133	85	3.76	5.02	4.90	0	0
5	474	33	1	22.40	110	62	5.41	3.08	5.00	0	0
6	2972	53	2	24.24	155	77	4.57	5.01	5.58	0	0
7	1607	36	1	24.50	116	68	4.61	3.59	4.40	0	0
8	3390	71	1	25.20	143	91	5.00	4.00	7.10	1	0
9	979	29	2	17.80	113	72	4.01	4.52	4.00	0	0
10	1493	38	1	32.60	121	72	5.60	4.94	5.49	0	1
11	3957	66	1	25.90	100	61	4.92	4.10	8.00	1	1
12	3699	57	1	21.80	108	65	5.40	4.79	5.00	1	1

13	4102	47	1	35.00	125	69	5.04	4.97	6.70	1	0
14	552	31	1	24.20	125	81	4.80	4.58	6.10	0	0
15	1814	35	2	23.90	106	63	4.78	4.72	4.90	0	0
16	1428	80	1	24.00	177	85	5.27	5.67	5.90	0	1
17	3260	62	1	24.50	122	76	6.90	5.27	6.40	1	1
18	2441	36	2	20.70	112	61	4.70	4.10	4.63	0	0
19	3701	49	1	25.00	132	93	6.09	4.26	7.30	1	0
20	2522	31	1	16.50	123	69	4.30	3.78	4.87	0	0

Terdapat 5 error yaitu pada nomor 8, 10, 13, 16 dan 19, berikut adalah penjelasannya :

- Nomor 8 True Label menunjukkan hasil Positive, sedangkan hasil Predict Label menunjukkan hasil Negative.
- Nomor 10 True Label menunjukkan hasil Negative, sedangkan hasil Predict Label menunjukkan hasil Positive.
- Nomor 13 True Label menunjukkan hasil Positive, sedangkan hasil Predict Label menunjukkan hasil Negative.
- Nomor 16 True Label menunjukkan hasil Negative, sedangkan hasil Predict Label menunjukkan hasil Positive.
- Nomor 19 True Label menunjukkan hasil Positive, sedangkan hasil Predict Label menunjukkan hasil Negative.

Tabel 3. 25 Confusion Matrix Hasil Perhitungan KNN K3

True Label	Negative	12	2
	Positive	3	3
		Negative	Positive
		Predict Label	

$$\text{Akurasi} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{3+12}{3+12+3+2} = \frac{15}{20} = 0,75$$

$$\text{Laju error} = \frac{FP+FN}{FP+FN+TP+TN} = \frac{3+2}{3+2+3+14} = \frac{5}{20} = 0,25$$

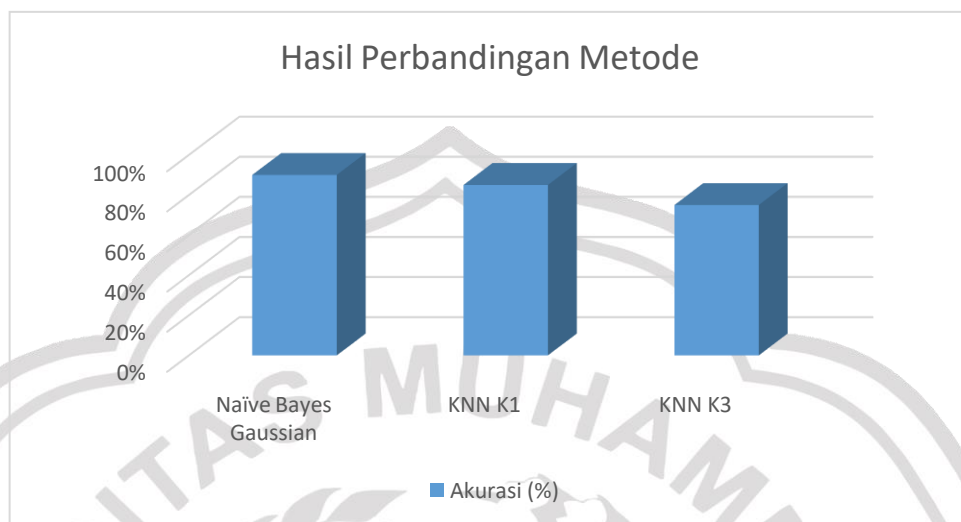
$$\text{Presisi} = \frac{TP}{TP+FP} = \frac{3}{3+3} = \frac{3}{6} = 0,5$$

$$\text{Recall} = \frac{TP}{TP+FN} = \frac{3}{3+2} = \frac{3}{5} = 0,6$$



### 3.3.6. Hasil Perbandingan Metode

Dibawah ini adalah hasil perbandingan 2 metode, yaitu metode *Naïve Bayes Gaussian* dan KNN dengan nilai K1 dan K3.



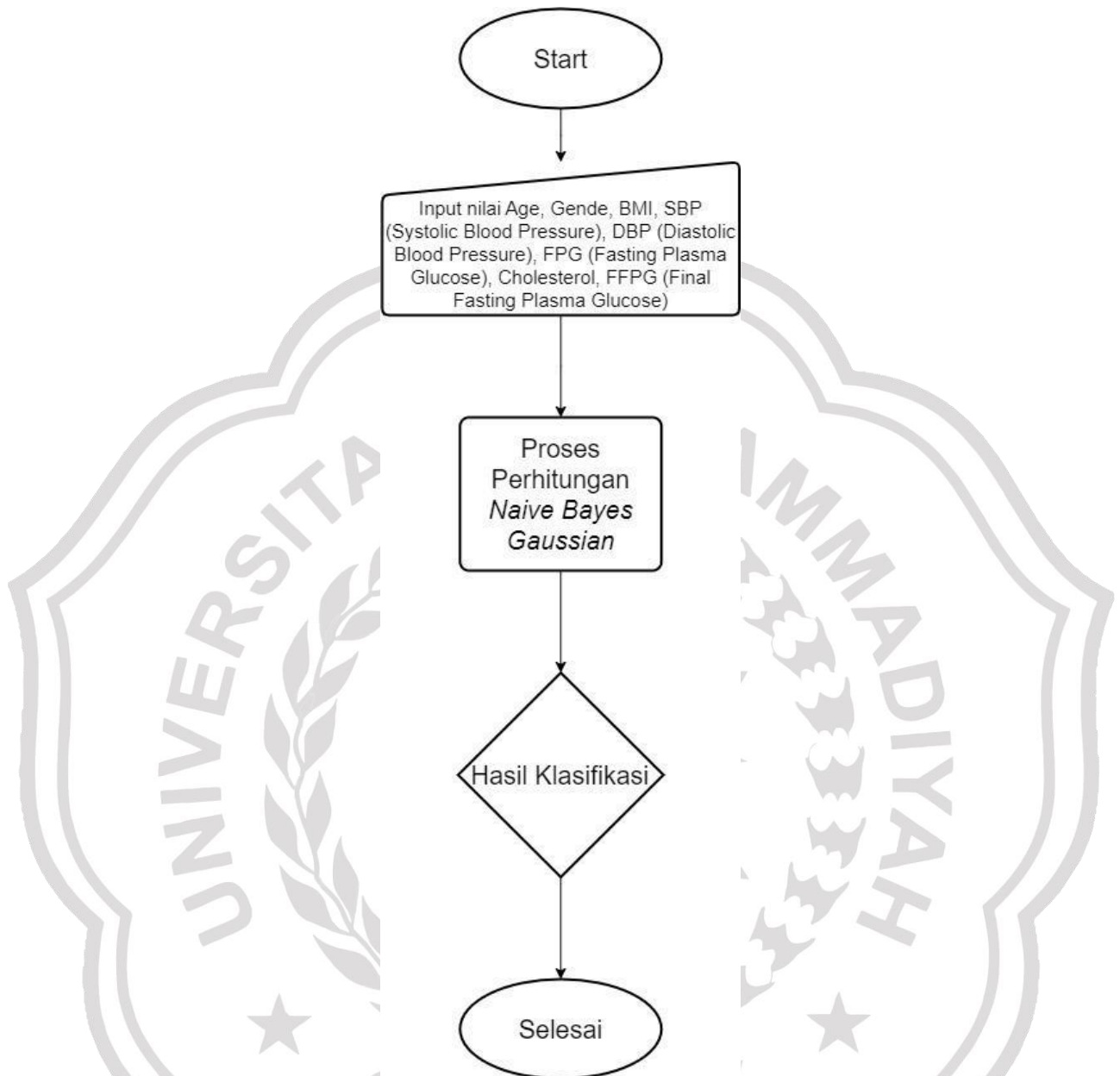
Gambar 3. 3 Hasil Perbandingan Metode

Dari gambar diatas, diperoleh nilai akurasi sebesar 90% untuk *Naïve Bayes Gaussian*, 85% untuk KNN K1 dan 75% untuk KNN K3. Berdasarkan, hasil akurasi diatas. Maka, pada penelitian ini peneliti menggunakan Algoritme *Naïve Bayes Gaussian* sebagai metode untuk pembuatan sistem “klasifikasi potensi penyakit diabetes mellitus tipe II pada pasien menggunakan algoritme naïve bayes gaussian”.

### 3.4. Perancangan Sistem

Perancangan sistem ini terdiri dari Flowchart Sistem, DFD, dan ERD.

### 3.4.1. Flowchart Sistem



Gambar 3. 4 Flowchart Sistem

Adapun keterangan dari gambar flowchart tersebut adalah sebagai berikut :

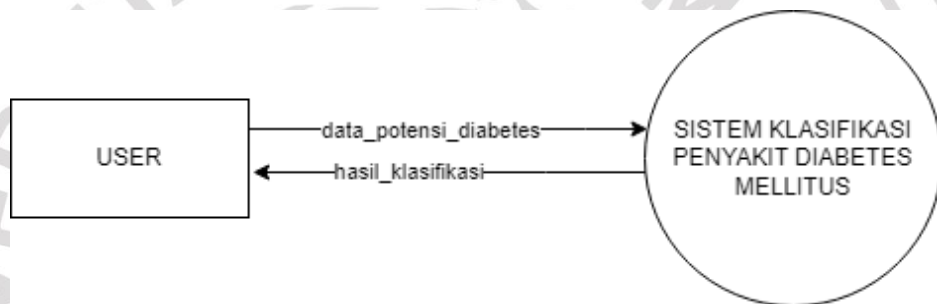
- 1) Mulai : Inialisasi langkah awal dalam sistem.
- 2) Input Nilai Age, Gender, BMI, SBP (*Systolic Blood Pressure*), DBP (*Diastolic Blood Pressure*), FPG (*Fasting Plasma Glucose*), Cholesterol, FFPG (*Final Fasting Plasma Glucose*) : User diminta untuk memasukkan nilai Age, Gender, BMI, SBP (*Systolic Blood Pressure*), DBP (*Diastolic*

*Blood Pressure*), FPG (*Fasting Plasma Glucose*), Cholesterol, FFBG (*Final Fasting Plasma Glucose*)

- 3) Proses perhitungan naïve bayes gaussian : sistem menerapkan algoritme naïve bayes gaussian untuk menghitung probabilitas kelas atau kelompok tertentu berdasarkan distribusi gaussian dari nilai – nilai input tersebut.
- 4) Hasil prediksi : setelah proses perhitungan selesai, sistem menghasilkan prediksi berupa kelas atau kelompok yang paling mungkin berdasarkan nilai – nilai input yang diberikan.
- 5) Selesai : Proses selesai dan output atau hasil prediksi akan ditampilkan.

### 3.4.2. Data Flow Diagram (DFD)

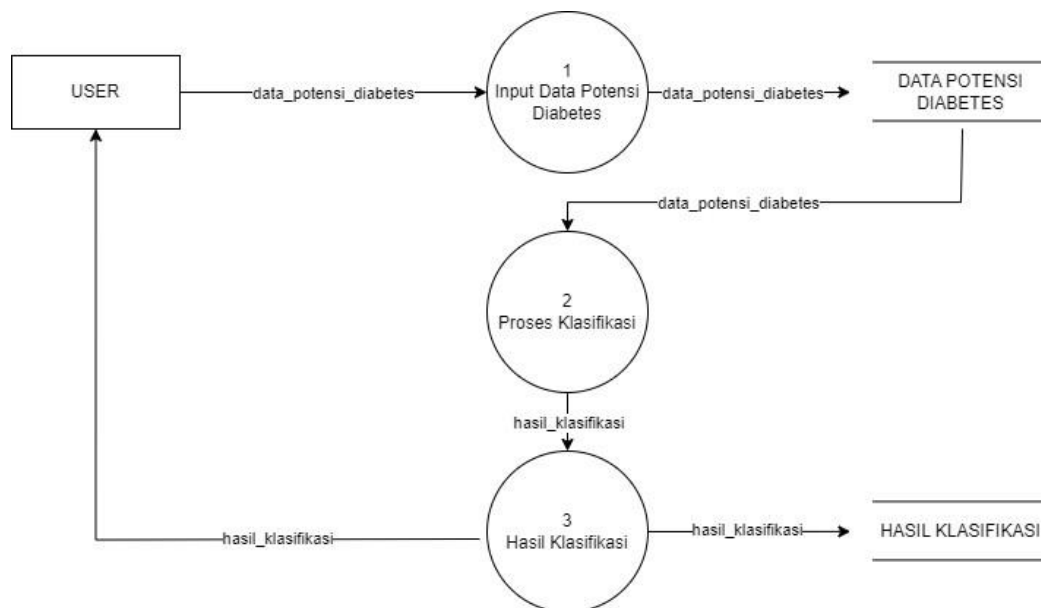
- a) Diagram Konteks (DFD Level 0)



Gambar 3. 5 Diagram Konteks (DFD Level 0)

Pada sistem ini , user melakukan input potensi diabetes. Jika di submit, user mendapatkan feedback berupa hasil klasifikasi apakah ciri - ciri tersebut sudah masuk ke dalam ciri - ciri penyakit diabetes mellitus.

- b) DFD Level 1

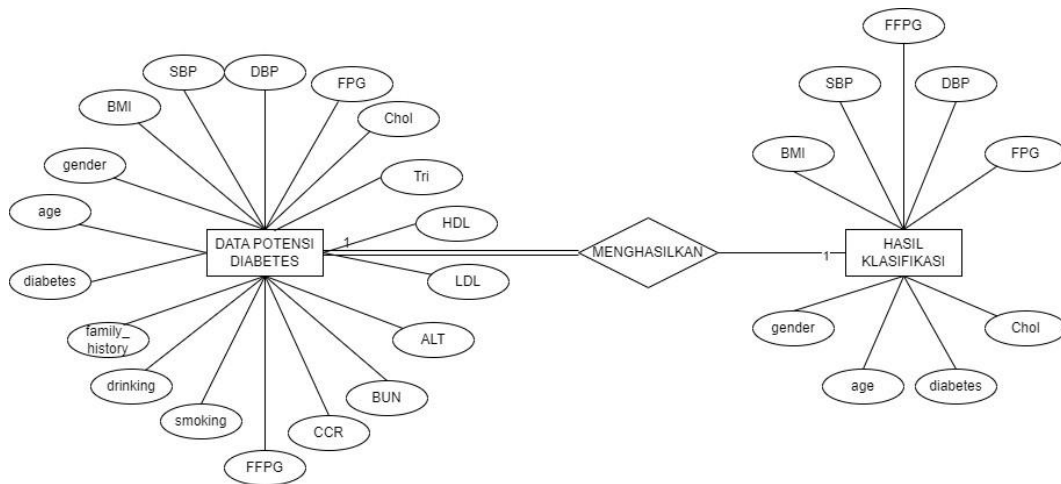


Gambar 3. 6 DFD Level 1

Pada DFD Level 1, proses dibagi ke dalam tiga bagian. Proses pertama, user melakukan input data potensi diabetes. Proses kedua adalah proses klasifikasi dan Proses ketiga adalah hasil klasifikasi. Pertama, user melakukan input data berupa data potensi diabetes yang akan tersimpan ke dalam database potensi diabetes. Lalu, proses kedua yaitu data potensi diabetes tersebut akan diproses sistem untuk dilakukan klasifikasi. Jika, proses selesai, maka akan dilanjutkan ke proses yang ketiga yaitu hasil klasifikasi. Hasil klasifikasi akan ditampilkan ke user dan akan tersimpan ke dalam database hasil klasifikasi.

### 3.4.3. Entity Relationship Diagram (ERD)

Dibawah ini adalah *Entity Relationship Diagram* pada klasifikasi potensi penyakit diabetes mellitus tipe II pada pasien.



Gambar 3. 7 Entity Relationship Diagram (ERD)

### 3.5. Perancangan Antarmuka Sistem

Perancangan Antarmuka Sistem ini terdiri dari Halaman Utama Sebelum Klasifikasi dan Halaman Utama Sesudah Klasifikasi.

#### 3.5.1. Halaman Utama Sebelum Klasifikasi

**KLASIFIKASI POTENSI PENYAKIT DIABETES MELLITUS TIPE II PADA PASIEN MENGGUNAKAN ALGORITMA NAÏVE BAUES GAUSSIAN**

<p>Input Nilai Age</p> <input style="width: 90%;" type="text"/>	<p>Input Nilai Gender (1 = Female, 2 = Male)</p> <input style="width: 90%;" type="text"/>
<p>Input Nilai BMI</p> <input style="width: 90%;" type="text"/>	<p>Input Nilai SBP (Systolic Blood Pressure)</p> <input style="width: 90%;" type="text"/>
<p>Input Nilai DBP (Diastolic Blood Pressure)</p> <input style="width: 90%;" type="text"/>	<p>Input Nilai FPG (Fasting Plasma Glucose)</p> <input style="width: 90%;" type="text"/>
<p>Input Nilai Cholesterol</p> <input style="width: 90%;" type="text"/>	<p>Input Nilai FFPG (Final Fasting Plasma Glucose)</p> <input style="width: 90%;" type="text"/>
<input style="width: 150px; height: 25px;" type="button" value="Klasifikasi Diabetes"/>	

Gambar 3. 8 Halaman Utama Sebelum Klasifikasi



### 3.5.2. Halaman Utama Sesudah Klasifikasi

**KLASIFIKASI POTENSI PENYAKIT DIABETES MELLITUS TIPE II  
PADA PASIEN MENGGUNAKAN ALGORITMA NAÏVE BAUES  
GAUSSIAN**

<p>Input Nilai Age</p> <input style="width: 90%;" type="text"/>	<p>Input Nilai Gender (1 = Female, 2 = Male)</p> <input style="width: 90%;" type="text"/>
<p>Input Nilai BMI</p> <input style="width: 90%;" type="text"/>	<p>Input Nilai SBP (Systolic Blood Pressure)</p> <input style="width: 90%;" type="text"/>
<p>Input Nilai DBP (Diastolic Blood Pressure)</p> <input style="width: 90%;" type="text"/>	<p>Input Nilai FPG (Fasting Plasma Glucose)</p> <input style="width: 90%;" type="text"/>
<p>Input Nilai Kolesterol</p> <input style="width: 90%;" type="text"/>	<p>Input Nilai FFPG (Final Fasting Plasma</p> <input style="width: 90%;" type="text"/>

Gambar 3. 9 Halaman Utama Sesudah Klasifikasi

### 3.6. Perancangan Pengujian Sistem

Perancangan pengujian sistem ini menggunakan black box testing

No.	Test Case	Input	Expected Result	Hasil	Status
1.	User memasukkan kriteria untuk proses klasifikasi.	Age, Gender, BMI, SBP (Systolic Blood Pressure), DBP (Diastolic Blood Pressure), FPG (Fasting Plasma Glucose), Kolesterol dan FFPG	Sistem dapat input data Age, Gender, BMI, SBP (Systolic Blood Pressure), DBP (Diastolic Blood Pressure), FPG (Fasting Plasma Glucose) ke	Sistem berhasil input data Age, Gender, BMI, SBP (Systolic Blood Pressure), DBP (Diastolic Blood Pressure), FPG (Fasting Plasma Glucose), Kolesterol dan FFPG (Final Fasting Plasma Glucose) ke	

		(Final Fasting Plasma Glucose).	database untuk proses klasifikasi.	database untuk proses klasifikasi.	
2.	User dapat submit tombol "klasifikasi diabetes".	Klik button "klasifikasi diabetes".	Sistem dapat menampilkan hasil klasifikasi menggunakan metode naïve bayes gaussian dan dapat menyimpan ke database.	Sistem berhasil menampilkan hasil klasifikasi menggunakan metode naïve bayes gaussian dan dapat menyimpan ke database.	

