

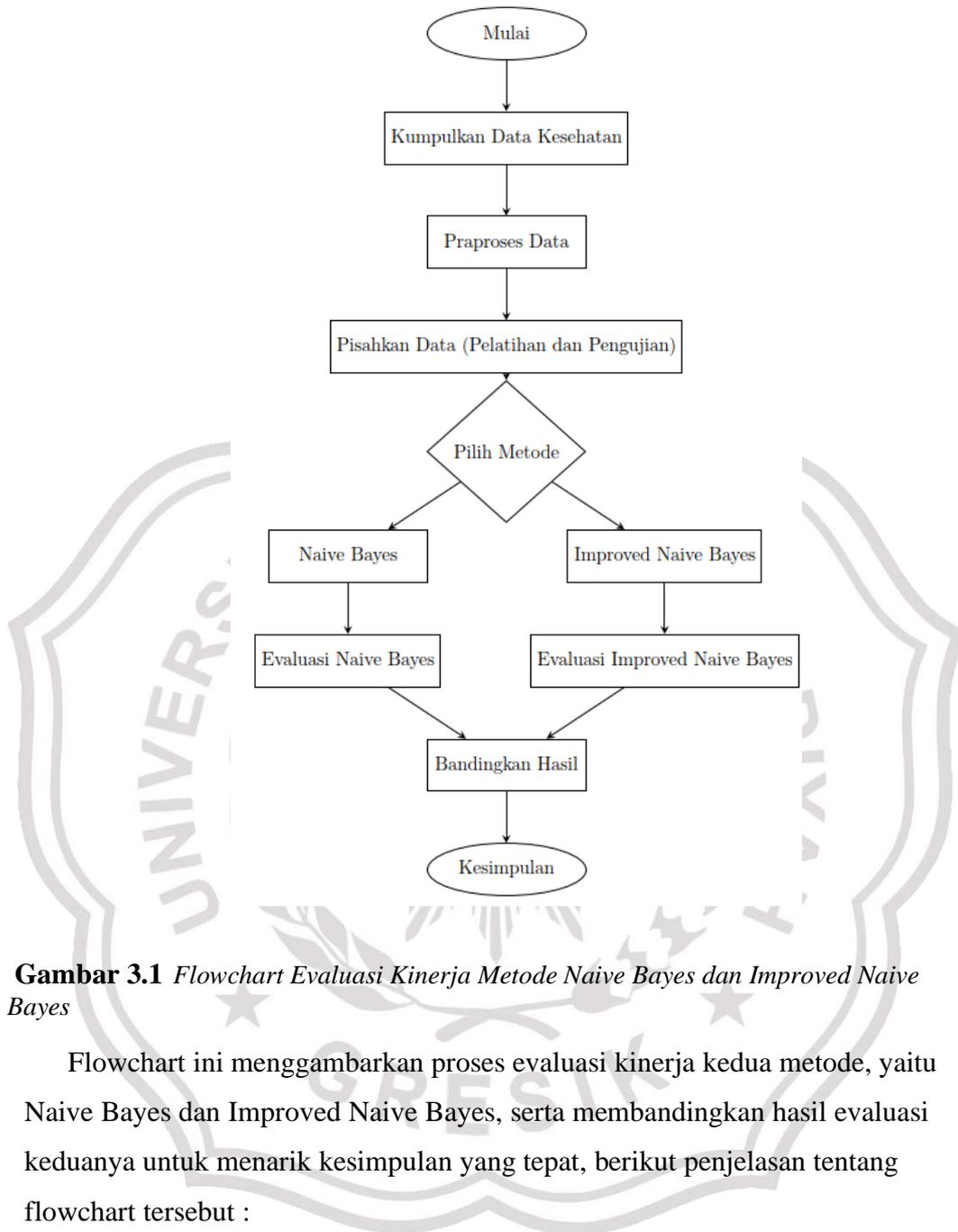
## **BAB 3**

### **ANALISIS DAN PERANCANGAN SISTEM**

#### **3.1. Analisis Sistem**

Kolesterol tinggi, suatu kondisi yang berkaitan erat dengan kesehatan *kardiovaskular*, telah menjadi perhatian serius dalam masyarakat saat ini. Kondisi ini ditandai oleh tingginya kadar kolesterol dalam darah, yang dapat meningkatkan risiko terjadinya penyakit jantung, stroke, dan komplikasi kesehatan lainnya. Peningkatan kadar kolesterol dalam darah yang tidak terkontrol merupakan salah satu faktor yang berkontribusi pada penyakit kardiovaskular (Ardytama Kusuma Yudha, 2023). Oleh karena itu, pemahaman mendalam mengenai tingkat prevalensi dan risiko kolesterol tinggi menjadi sangat penting untuk merancang langkah-langkah pencegahan yang efektif.

Langkah-langkah pencegahan, seperti edukasi masyarakat tentang pola hidup sehat, pemeriksaan rutin, dan pengelolaan pola makan, dapat memainkan peran kunci dalam mengurangi risiko kolesterol tinggi. Dengan memahami situasi ini, dapat diambil tindakan proaktif untuk melindungi kesehatan masyarakat dan mengurangi dampak buruk penyakit kolesterol tinggi di masa depan. Selain itu, langkah-langkah kuantitatif juga dapat memberikan pandangan yang lebih mendalam terkait efektivitas tindakan pencegahan ini dengan mengukur dampaknya secara numerik (Husen et al., 2022). Hal ini memberikan dasar yang kuat untuk pengambilan keputusan dalam upaya meningkatkan kesehatan masyarakat secara menyeluruh, sebagaimana terlihat pada flowchart penelitian atau gambar 3.1.



**Gambar 3.1** Flowchart Evaluasi Kinerja Metode Naive Bayes dan Improved Naive Bayes

Flowchart ini menggambarkan proses evaluasi kinerja kedua metode, yaitu Naive Bayes dan Improved Naive Bayes, serta membandingkan hasil evaluasi keduanya untuk menarik kesimpulan yang tepat, berikut penjelasan tentang flowchart tersebut :

1. **Mulai:** Proses dimulai dengan identifikasi tujuan penelitian dan persiapan awal.
2. **Kumpulkan Data Kesehatan:** Langkah pertama, melakukan pencarian data yang relevan dari situs bank data online yang dikenal dengan nama Kaggle, kemudian melakukan validasi dengan meminta rekomendasi dari

seorang dokter spesialis penyakit dalam di Rumah Sakit Muhammadiyah Gresik

3. **Praproses Data:** Data yang telah dikumpulkan kemudian diproses untuk membersihkan data yang tidak valid, mengisi nilai yang hilang, dan mengonversi format data agar siap untuk analisis lebih lanjut.
4. **Pisahkan Data (Pelatihan dan Pengujian):** Data yang telah diproses dibagi menjadi dua bagian, yaitu data pelatihan (training data) yang digunakan untuk mengembangkan model analisis dan data pengujian (testing data) yang digunakan untuk menguji dan mengevaluasi model tersebut.
5. **Pilih Metode:** Pada tahap ini, peneliti memilih metode analisis yang sesuai dengan tujuan penelitian. Dalam konteks ini, metode yang dipilih adalah Naive Bayes dan Improved Naive Bayes.
6. **Naive Bayes:** Metode Naive Bayes digunakan untuk melakukan klasifikasi data berdasarkan probabilitas dan asumsi independensi antar variabel.
7. **Improved Naive Bayes:** Metode Improved Naive Bayes adalah variasi dari Naive Bayes yang telah ditingkatkan dalam hal akurasi atau efisiensi, mungkin dengan mempertimbangkan faktor-faktor tambahan atau mengatasi beberapa kelemahan yang ada.
8. **Evaluasi Naive Bayes:** Setelah model Naive Bayes dikembangkan, dilakukan evaluasi untuk mengukur kinerja model dalam mengklasifikasikan data pengujian. Evaluasi dapat dilakukan menggunakan metrik seperti akurasi, presisi, recall, atau F1-score.
9. **Evaluasi Improved Naive Bayes:** Model Improved Naive Bayes juga dievaluasi untuk membandingkan kinerja antara kedua metode. Evaluasi ini membantu dalam menentukan apakah peningkatan yang dilakukan pada metode Naive Bayes memberikan hasil yang lebih baik.
10. **Bandungkan Hasil:** Hasil dari evaluasi kedua metode dibandingkan untuk menentukan metode yang lebih efektif atau akurat dalam melakukan klasifikasi data kesehatan.

**11. Kesimpulan:** Berdasarkan hasil analisis dan perbandingan antara Naive Bayes dan Improved Naive Bayes, kesimpulan dibuat untuk menyajikan temuan penelitian dan implikasinya terhadap konteks kesehatan yang diteliti.

### 3.2 Hasil Analisis

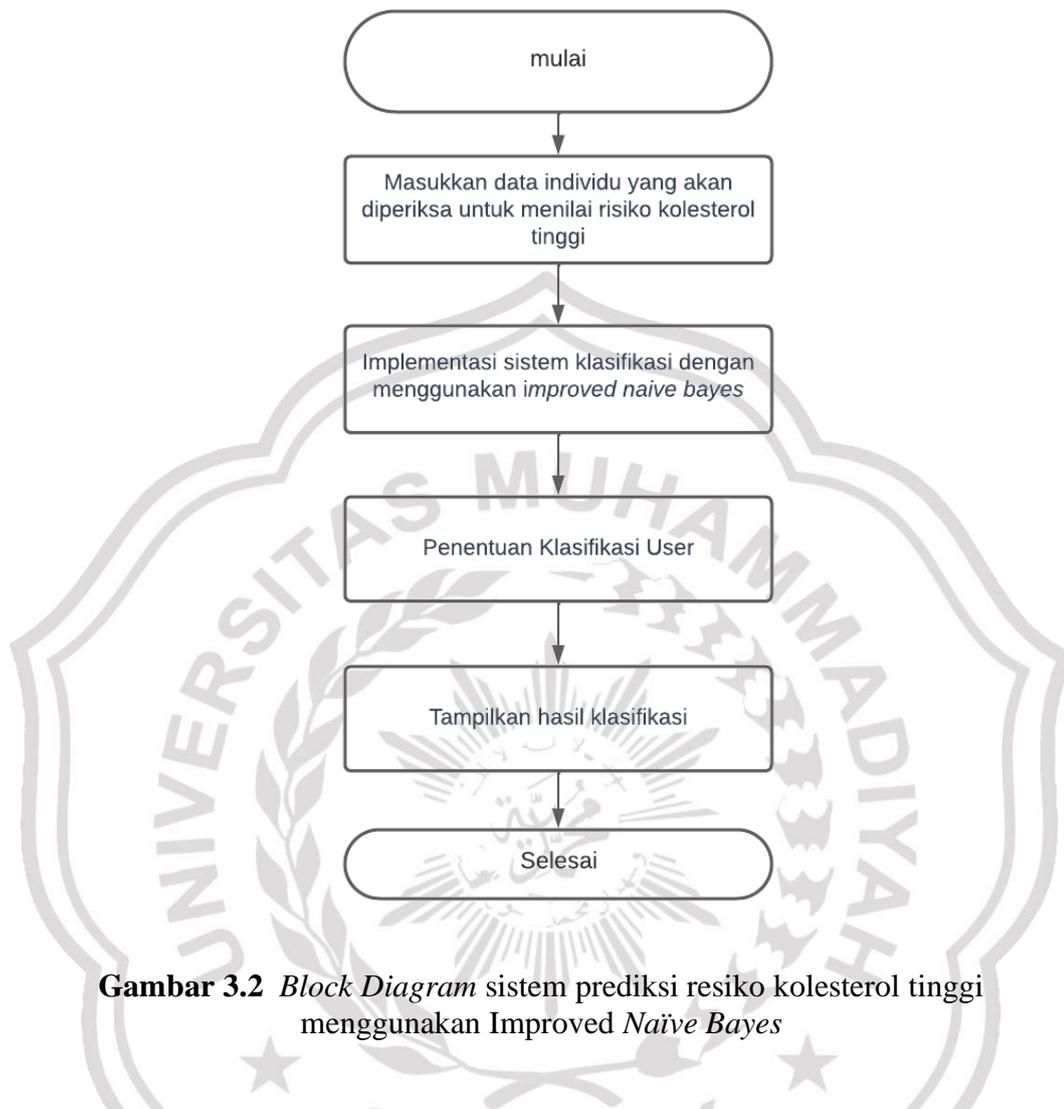
Hasil analisis yang dapat dilakukan dari prediksi risiko kolesterol tinggi menggunakan metode *Improved Naive Bayes* memerlukan sistem yang mampu mengelola data kesehatan yang telah diberi label sesuai dengan kriteria kategori yang ditetapkan. Data ini mencakup berbagai variabel seperti usia, berat badan, tingkat aktivitas fisik, riwayat keluarga, pola makan, dan faktor-faktor lain yang relevan dalam menentukan risiko kolesterol tinggi. Selanjutnya, data yang telah terkumpul akan diolah dan dianalisis menggunakan metode Improved Naive Bayes, yang merupakan pengembangan dari metode Naive Bayes yang lebih kompleks dan mampu mengatasi beberapa kelemahan yang ada. Proses klasifikasi data dilakukan dengan membagi data menjadi kategori risiko yang telah ditentukan, seperti berisiko dan aman atau tidak berisiko. Setiap data akan diberi label sesuai dengan kategori yang paling sesuai berdasarkan nilai-nilai variabel yang diamati. Hasil klasifikasi ini akan menjadi dasar untuk melakukan analisis lebih lanjut terkait faktor-faktor yang dominan memengaruhi tingkat risiko kolesterol tinggi pada populasi tersebut.

Selain itu, hasil analisis ini juga dapat digunakan untuk memberikan informasi yang lebih komprehensif tentang dinamika kesehatan masyarakat terkait risiko kolesterol tinggi. Data yang terkumpul dan hasil analisis yang dihasilkan dapat memberikan pemahaman yang lebih baik tentang faktor-faktor yang perlu diperhatikan dalam upaya pencegahan dan penanganan risiko kolesterol tinggi. Dengan demikian, sistem prediksi risiko kolesterol tinggi berdasarkan data kesehatan dengan metode Improved Naive Bayes Classifier menjadi sebuah instrumen yang penting dalam mendukung upaya pencegahan dan penanganan kolesterol tinggi, serta meningkatkan pemahaman tentang kesehatan masyarakat secara keseluruhan.

### 3.3 Perancangan Sistem

Informasi yang belum memiliki makna bagi penerima dan masih memerlukan proses pengolahan disebut sebagai data. Data dapat berupa berbagai bentuk, seperti keadaan, gambar, huruf, atau angka. Sebelum dapat memberikan makna atau informasi yang berguna, data tersebut perlu melalui suatu tahap pengolahan atau analisis. Proses ini mengubahnya menjadi informasi yang dapat dipahami dan digunakan oleh penerima untuk membuat keputusan atau mendapatkan pemahaman yang lebih mendalam. Oleh karena itu, data menjadi elemen kritis dalam proses pengambilan keputusan dan analisis informasi.

Sistem yang dikembangkan berupa aplikasi atau alat klasifikasi pasien penyakit kolesterol tinggi menggunakan teknik data mining dengan metode *Naive Bayes*. Output sistem ini mencakup prediksi kategori penyakit, yaitu kolesterol tinggi. Beberapa atribut yang dibutuhkan untuk klasifikasi penderita penyakit kolesterol tinggi mencakup usia, berat badan, dan faktor-faktor lain yang signifikan. Proses alur sistem ini dijelaskan pada Gambar 3.2 untuk memberikan gambaran visual bagaimana sistem klasifikasi penyakit kolesterol tinggi dilakukan.



**Gambar 3.2** *Block Diagram* sistem prediksi resiko kolesterol tinggi menggunakan *Improved Naïve Bayes*

Penjelasan pada Gambar 3.2 :

1. Tahap awal melibatkan pengumpulan informasi medis individu yang rentan terhadap risiko kolesterol tinggi dari catatan kesehatan yang ada. Seperti usia, jenis kelamin, indeks massa tubuh, status merokok, tingkat aktivitas fisik, konsumsi buah-buahan, konsumsi sayuran, indikator konsumsi alkohol berat, penilaian umum tentang kesehatan, kesehatan mental, kesehatan fisik, dan status diabetes.
2. Sistem memulai proses perhitungan klasifikasi dengan menggunakan metode Improve Naive Bayes. Proses ini melibatkan pemanfaatan data

pembelajaran untuk setiap kategori, memastikan akurasi dalam menentukan hasil.

3. Setelah proses klasifikasi, sistem menentukan hasil akhir dan mendiagnosis risiko kolesterol tinggi pada individu berdasarkan kelas baru yang dihasilkan dari perhitungan klasifikasi.
4. Langkah terakhir melibatkan penampilan hasil diagnosa risiko kolesterol tinggi pada individu, dengan kemungkinan klasifikasi individu sebagai rentan terhadap risiko kolesterol tinggi yang dapat memicu berbagai komplikasi kesehatan.

### 3.3.1 Sumber Data

Langkah pertama dalam penelitian ini melibatkan persiapan data yang diperoleh dari situs Kaggle atau bank data daring berupa 7.692 tanggapan survei BRFSS 2015. Sebanyak 200 data diambil sesuai dengan rekomendasi dari Dr. Muchammad Fakhruddin Fakhry, Sp.PD, yang bertugas di Rumah Sakit Muhammadiyah Gresik, dengan ketentuan 100 data positif berisiko kolesterol tinggi dan 100 data negatif atau tidak berisiko kolesterol tinggi. Data ini mencakup informasi terkait risiko kolesterol tinggi dengan atribut pendukung seperti usia, jenis kelamin, indeks massa tubuh (BMI), kebiasaan merokok, aktivitas fisik, konsumsi buah, konsumsi sayuran, konsumsi alkohol berat, kondisi kesehatan umum, kesehatan mental, kesehatan fisik, dan diabetes.

Total data set yang digunakan dalam analisis adalah 200 data set. Data ini dibagi menjadi lima bagian yang sama besar untuk penerapan k-fold cross-validation dengan  $k=5$ . Setiap bagian, yang masing-masing berisi 40 data set, digunakan secara bergantian sebagai data uji, sementara empat bagian lainnya, yang berjumlah 160 data set, digunakan sebagai data latih. Pembagian data ini dilakukan tanpa mengubah urutan asli data uji, mengikuti pembagian yang telah ditentukan oleh k-fold cross-validation. Seluruh proses pengujian dilakukan sebanyak lima kali, memastikan bahwa setiap bagian dari 200 data set telah digunakan sebagai data uji setidaknya satu kali. Metode

Improved Naive Bayes digunakan dalam proses perhitungan, melibatkan 12 atribut yang relevan. Informasi terkait variabel tercantum pada Tabel 3.1. Atribut tersebut mencakup *Age*, *Sex*, *BMI*, *Smoker*, *PhysActivity*, *Fruits*, *Veggies*, *HvyAlcoholConsump*, *GenHlth*, *MentHlth*, *PhysHlth*, dan *Diabetes*. Data ini membentuk dasar analisis untuk mengevaluasi risiko kolesterol tinggi dalam populasi survei yang luas.

**Tabel 3.1** Data Atribut

Variable	Keterangan
<i>Age</i>	Variabel ' <i>_AGEG5YR</i> ' menggambarkan kategori usia dalam interval lima tahun, dengan level 1 untuk usia 18-24 tahun, level 9 untuk usia 60-64 tahun, dan level 13 untuk usia 80 tahun ke atas
<i>Sex</i>	jenis kelamin
<i>BMI</i>	index massa tubuh
<i>Smoker</i>	Merokok setidaknya 100 batang rokok sepanjang hidup
<i>PhysActivity</i>	Aktivitas fisik dalam 30 hari terakhir - tidak termasuk pekerjaan.
<i>Fruits</i>	Mengonsumsi buah 1 kali atau lebih per hari.
<i>Veggies</i>	Mengonsumsi sayuran 1 kali atau lebih per hari.
<i>HvyAlcoholConsump</i>	(Orang dewasa pria $\geq 14$ minuman per minggu dan orang dewasa wanita $\geq 7$ minuman per minggu)
<i>GenHlth</i>	kesehatan Secara umum
<i>MentHlth</i>	Jumlah hari dengan kesehatan mental yang buruk, dalam skala 1-30 hari.
<i>PhysHlth</i>	Jumlah hari mengalami penyakit fisik atau cedera dalam 30 hari terakhir, dalam skala 1-30 hari.
<i>Diabetes</i>	menghidam penyakit diabetes

### 3.3.2 Diagram Konteks

Pada fase ini, akan dijelaskan tentang diagram konteks, diagram hirarki, diagram alur data, perancangan basis data, dan antarmuka sistem.



**Gambar 3.3** Diagram Konteks Sistem Prediksi Resiko Kolesterol Tinggi

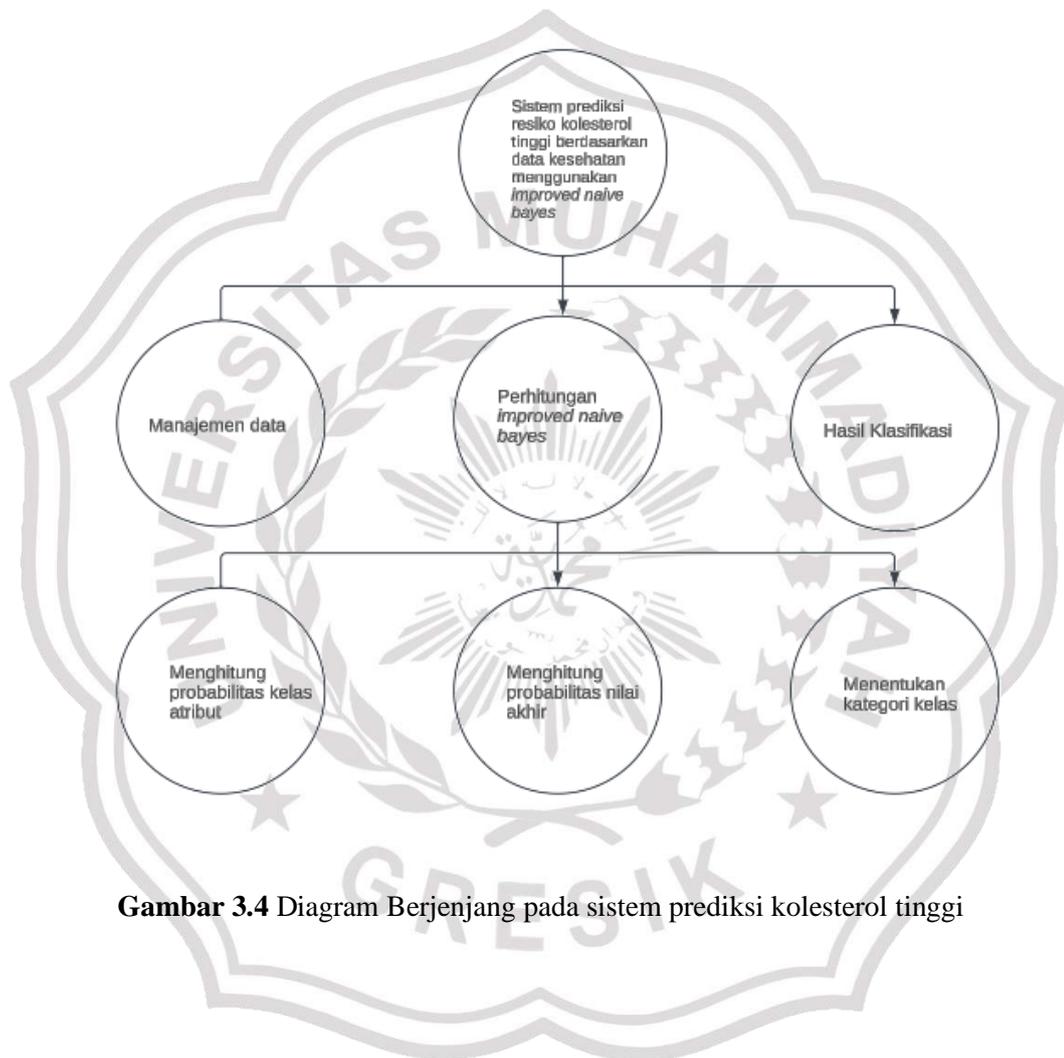
Penjelasan pada gambar 3.3 adalah entitas dalam sistem ini adalah user dan admin. User memasukkan data diri terkait risiko kolesterol tinggi, melibatkan atribut seperti usia, jenis kelamin, pemeriksaan kolesterol, indeks massa tubuh (BMI), kebiasaan merokok, riwayat penyakit jantung atau serangan jantung, aktivitas fisik, konsumsi buah, konsumsi sayuran, konsumsi alkohol berat, kondisi kesehatan umum, kesehatan mental, kesehatan fisik, kesulitan berjalan, riwayat stroke, tekanan darah tinggi, dan diabetes. Data tersebut digunakan sebagai data latih atau data training yang akan dihitung probabilitasnya. Selanjutnya, pengguna dapat memperoleh output berupa hasil klasifikasi. Dan admin hanya dapat menerima output hasil klasifikasi dari semua pengguna, demikian juga dengan data penggunaanya.

### 3.3.3 Diagram Berjenjang

Gambar 3.4 di bawah ini adalah gambaran berjenjang pada sistem klasifikasi penyakit kolesterol tinggi menggunakan metode *Naive Bayes*

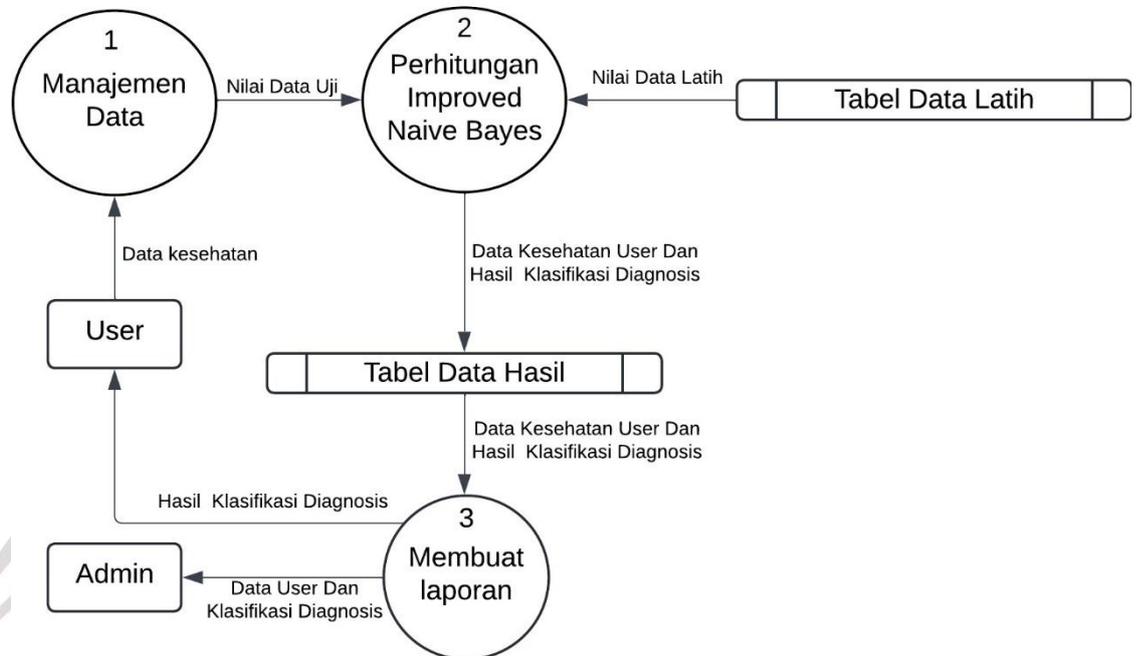
- a. Top Level: Klasifikasi penyakit kolesterol tinggi menggunakan metode *Naive Bayes*.
- b. Level 0:
  1. Manajemen data

2. Perhitungan *Naive Bayes*
  3. Pelaporan hasil klasifikasi kelas
- c. Level 1:
1. Menghitung nilai probabilitas fitur dan kelas
  2. Menghitung nilai probabilitas akhir
  3. Menentukan kategori kelas



**Gambar 3.4** Diagram Berjenjang pada sistem prediksi kolesterol tinggi

### 3.3.4 Data Flow Diagram Level 1

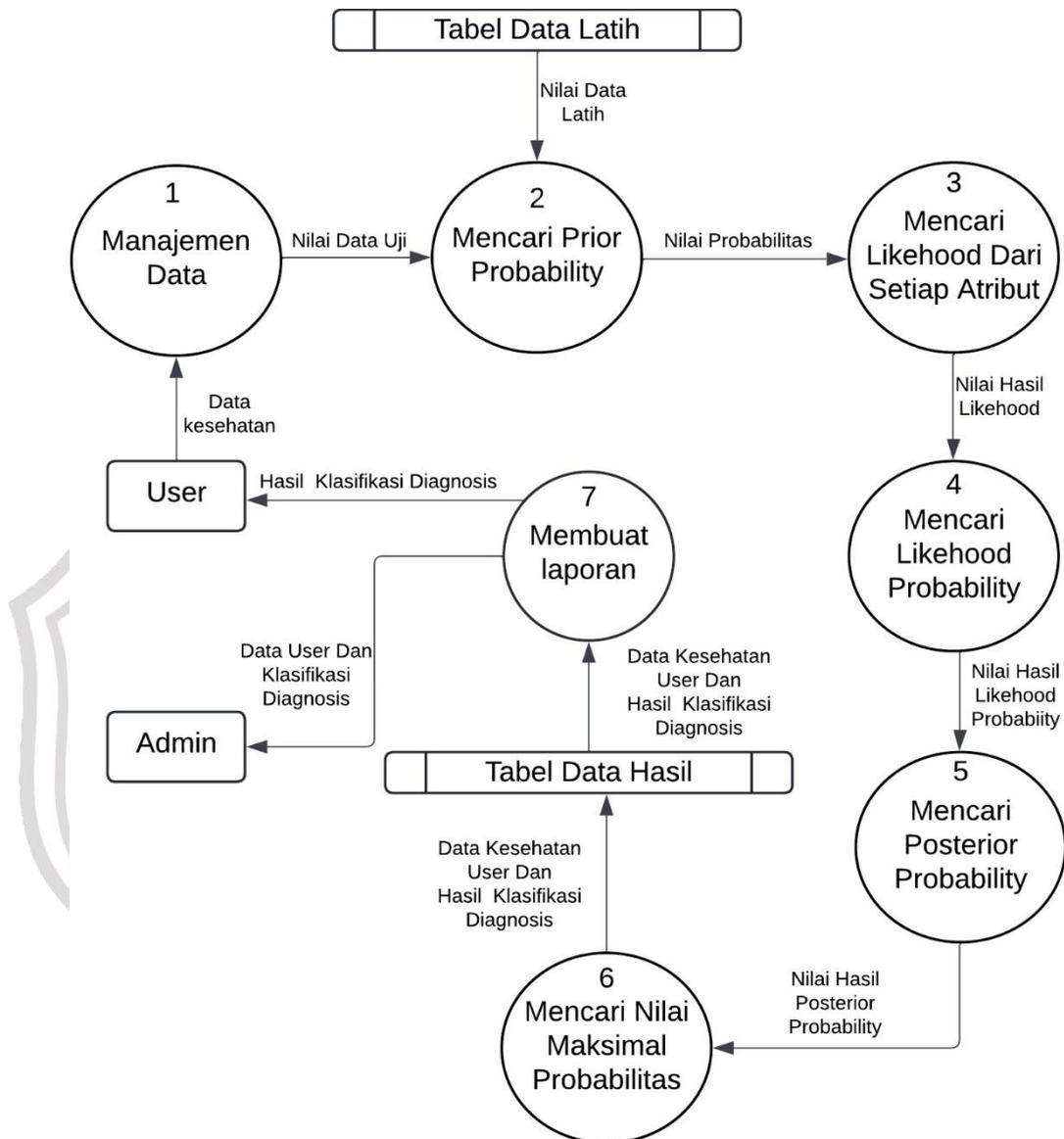


**Gambar 3.5** DFD level 1

Berdasarkan gambar 3.5, dapat dijelaskan bahwa DFD Level 0 menjelaskan beberapa proses yang terjadi. Beberapa proses yang ada pada DFD Level 0 antara lain:

1. Proses 1 yakni manajemen data, Proses ini bertujuan untuk mengubah jenis data yang diinput oleh pengguna agar sesuai dengan jenis data pada data latih.
2. Proses 2 yakni perhitungan *Improved Naive Bayes*, dalam proses ini dilakukan perhitungan klasifikasi menggunakan metode *Improved Naive Bayes*.
3. Proses 3 yakni laporan. Dalam proses ini, dilakukan pembuatan laporan hasil klasifikasi dan juga hasil diagnosis dari pengguna menggunakan metode *Improved Naive Bayes*..

### 3.3.5 Data Flow Diagram Level 2



**Gambar 3.6** DFD Level 2

1. Proses 1 yakni manajemen data, Proses ini bertujuan untuk mengubah jenis data yang diinput oleh pengguna agar sesuai dengan jenis data pada data latih.
2. Proses 2 yakni mencari probabilitas awal. Pada proses ini, dilakukan perhitungan probabilitas awal berdasarkan data kesehatan yang

diinputkan oleh pengguna, yang kemudian digunakan dalam metode *Improved Naive Bayes*.

3. Proses 3 yakni mencari *likelihood* dari setiap atribut. Dalam proses ini, dilakukan perhitungan *likelihood* untuk setiap atribut berdasarkan data latih yang digunakan dalam metode *Improved Naive Bayes*.
4. Proses 4 yakni mencari *likelihood probability*. Pada tahap ini, dilakukan perhitungan probabilitas keseluruhan dari kelas tertentu berdasarkan kombinasi dari semua nilai atribut yang diamati dalam data latih, yang merupakan bagian integral dari metode *Improved Naive Bayes*.
5. Proses 5 yakni mencari *posterior probability*. Dalam proses ini, dilakukan perhitungan probabilitas *posterior* berdasarkan *likelihood probability* dan *prior probability* yang telah dihitung sebelumnya.
6. Proses 6, yaitu pencarian nilai maksimal probabilitas. Pada tahap ini, dilakukan penentuan nilai maksimal dari probabilitas *posterior* yang telah dihitung sebelumnya, yang menjadi hasil akhir dari metode *Improved Naive Bayes*.
7. Proses 7 yakni laporan. Dalam proses ini, dilakukan pembuatan laporan hasil klasifikasi dan juga hasil diagnosis dari pengguna menggunakan metode *Improved Naive Bayes*.

Hasil klasifikasi kemudian akan disampaikan kepada pengguna dan administrator mengenai diagnosis kolesterol tinggi, apakah termasuk dalam kategori berisiko atau tidak, sehingga administrator dapat memberikan informasi atau arahan untuk langkah selanjutnya.

### 3.4 Representasi Model

Data yang akan digunakan untuk melakukan klasifikasi terkait penyakit diabetes berasal dari hasil survei BRFSS 2015. Sebelum tahap klasifikasi dapat dilakukan, Metode Naive Bayes dimulai dengan mengumpulkan dan menyiapkan data pelatihan yang telah dikategorikan. Langkah berikutnya adalah menghitung probabilitas prior untuk setiap kelas dalam data pelatihan.

Kemudian, untuk setiap fitur dalam dataset, hitung probabilitas kondisional dari setiap fitur yang diberikan kelas tertentu. Setelah semua probabilitas dihitung, gabungkan probabilitas prior dan probabilitas kondisional untuk setiap kelas menggunakan teorema Bayes untuk mendapatkan probabilitas posterior. Akhirnya, untuk mengklasifikasikan data baru, hitung probabilitas posterior untuk setiap kelas, dan pilih kelas dengan probabilitas tertinggi sebagai prediksi.

Terdapat dua kategori data yang akan dimanfaatkan, yakni data latih dan data uji. Data latih berperan sebagai pendukung untuk memberikan keputusan, sementara data uji digunakan untuk melakukan evaluasi sistem. Tujuannya adalah untuk mengevaluasi kemampuan model dalam menggeneralisasi pola yang telah dipelajari dari data latih, terutama dengan peningkatan yang telah dilakukan pada metode *Naive Bayes*.

Data latih yang digunakan dalam penelitian ini terdiri dari 200 data . Data ini diperoleh dari situs Kaggle atau bank data online sebanyak 7.692 tanggapan survei BRFSS 2015. dengan ketentuan 100 data berisiko tinggi terhadap kolesterol dan 100 data lainnya tidak berisiko. Data latih yang disajikan hanya merupakan bagian kecil dari keseluruhan data latih yang akan dicantumkan secara lengkap dalam lampiran, data latih dan data uji dapat dilihat sebagai berikut :

**Tabel 3.2** Data Latih

No	Age	Sex	BMI	Smoker	PhysActivity	Fruits	Veggies	HvyAlcoholConsum	GenHlth	MentHlth	PhysHlth	Diabetes	Kelas HighChol
1	8.0	0.0	29.0	1.0	1.0	1.0	1.0	0.0	2.0	0.0	0.0	0.0	0.0
2	1.0	0.0	18.0	0.0	1.0	1.0	1.0	0.0	2.0	7.0	0.0	0.0	0.0
3	6.0	1.0	31.0	1.0	0.0	1.0	1.0	0.0	4.0	0.0	0.0	0.0	0.0
4	3.0	0.0	32.0	0.0	1.0	1.0	1.0	0.0	3.0	0.0	0.0	0.0	0.0

5	6.0	1.0	27.0	1.0	0.0	1.0	1.0	0.0	3.0	0.0	6.0	0.0	0.0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
...	...	...	...	...	...	...	...	...	...	...	...	...	...
200	10.0	1.0	30.0	1.0	1.0	0.0	1.0	0.0	2.0	0.0	0.0	1.0	1.0

**Tabel 3.3** Data Uji

No	Age	Sex	BMI	Smoker	PhysActivity	Fruits	Veggies	HvyAlcoholConsum	GenHlth	MentHlth	PhysHlth	Diabetes	Kelas HighChol
1	7	0	53	0	0	1	1	0	3	0	6	0	?
2	2	0	20	0	1	0	1	0	2	5	0	0	?
3	3	1	24	0	1	1	1	0	1	0	0	0	?

Sesuai dengan Tabel 3.2 yang berisi data latih, dan Tabel 3.3 sebagai tabel data uji, terdapat kolom highchol pada tabel uji yang berisikan tanda tanya (“?”) yang memerlukan pengisian untuk menentukan apakah berpotensi mengalami kolesterol tinggi atau tidak. Proses perhitungan dilakukan menggunakan metode *Naive Bayes* yang belum ditingkatkan, lalu dilanjut dengan *Naive bayes* yang sudah di tingkatkan berdasarkan data tersebut.

Rumus teorema Bayes ini digunakan untuk menghitung persentase nilai atribut yang berasal dari perhitungan probabilitas setiap atribut dengan klasifikasi 0 dan 1, berdasarkan informasi dalam dataset yang tersedia. Langkah-langkah perhitungan probabilitas untuk mendiagnosis kolesterol tinggi dengan data training pada tabel 3.3 adalah sebagai berikut:

- Memisahkan 2 Class menjadi J1 untuk “*HighChol = Ya*” / Beresiko dan J2 untuk “*HighChol = Tidak*” / Aman.
- Mencari Prior Probability:

a. Probabilitas Healthy dengan hasil Ya atau (Beresiko)

$$\frac{\text{HighChol dengan hasil (J1)}}{\text{Total Record}} = \frac{100}{200} = 0.5$$

a. Probabilitas HighChol dengan hasil Tidak atau (Aman)

$$\frac{\text{HighChol dengan hasil (J2)}}{\text{Total Record}} = \frac{100}{200} = 0.5$$

• Mencari Likelihood dari setiap atribut

❖ **Data uji Pertama**

a. Age

$$1. \frac{\text{Likelihood Age (ya)}}{J1} = \frac{6}{100} = 0,06$$

$$2. \frac{\text{Likelihood Age (Tidak)}}{J2} = \frac{8}{100} = 0,08$$

b. Sex

$$1. \frac{\text{Likelihood Sex (ya)}}{J1} = \frac{59}{100} = 0,59$$

$$2. \frac{\text{Likelihood Sex (Tidak)}}{J2} = \frac{57}{100} = 0.57$$

c. BMI

$$1. \frac{\text{Likelihood BMI (ya)}}{J1} = \frac{0}{100} = 0$$

$$2. \frac{\text{Likelihood BMI (Tidak)}}{J2} = \frac{0}{100} = 0$$

d. Smoker

$$1. \frac{\text{Likelihood Smoker (ya)}}{J1} = \frac{45}{100} = 0.45$$

$$2. \frac{\text{Likelihood Smoker (Tidak)}}{J2} = \frac{56}{100} = 0.56$$

e. PhysActivity

$$1. \frac{\text{Likelihood PhyActivity (ya)}}{J_1} = \frac{50}{100} = 0.50$$

$$2. \frac{\text{Likelihood PhyActivity (Tidak)}}{J_2} = \frac{21}{100} = 0.21$$

f. Fruits

$$1. \frac{\text{Likelihood Fruits (ya)}}{J_1} = \frac{48}{100} = 0.48$$

$$2. \frac{\text{Likelihood Fruits (Tidak)}}{J_2} = \frac{70}{100} = 0.70$$

g. Veggies

$$1. \frac{\text{Likelihood Veggies (ya)}}{J_1} = \frac{67}{100} = 0.67$$

$$2. \frac{\text{Likelihood Veggies (Tidak)}}{J_2} = \frac{85}{100} = 0.85$$

h. HvyAlcoholConsump

$$1. \frac{\text{Likelihood HvyAlcoholConsump (ya)}}{J_1} = \frac{96}{100} = 0.96$$

$$2. \frac{\text{Likelihood HvyAlcoholConsump (Tidak)}}{J_2} = \frac{94}{100} = 0.94$$

i. GenHlth

$$1. \frac{\text{Likelihood GenHlth (ya)}}{J_1} = \frac{35}{100} = 0.35$$

$$2. \frac{\text{Likelihood GenHelth (Tidak)}}{J_2} = \frac{19}{100} = 0.19$$

j. MentHlth

$$1. \frac{\text{Likelihood MenHlth (ya)}}{J_1} = \frac{60}{100} = 0.60$$

$$2. \frac{\text{Likelihood MenHlth (Tidak)}}{J_2} = \frac{70}{100} = 0.70$$

k. PhysHlth

$$1. \frac{\text{Likelihood PhysHlth (ya)}}{J_1} = \frac{2}{100} = 0.02$$

$$2. \frac{\text{Likelihood PhyHlth (Tidak)}}{J_2} = \frac{1}{100} = 0.01$$

1. Diabetes

$$1. \frac{\text{Likelihood Diabetes (ya)}}{J_1} = \frac{0}{100} = 0$$

$$2. \frac{\text{Likelihood Diabetes (Tidak)}}{J_2} = \frac{100}{100} = 1$$

❖ **Data uji Kedua**

a. Age

$$1. \frac{\text{Likelihood Age (ya)}}{J_1} = \frac{0}{100} = 0$$

$$2. \frac{\text{Likelihood Age (Tidak)}}{J_2} = \frac{5}{100} = 0,05$$

b. Sex

$$1. \frac{\text{Likelihood Sex (ya)}}{J_1} = \frac{59}{100} = 0.59$$

$$2. \frac{\text{Likelihood Sex (Tidak)}}{J_2} = \frac{57}{100} = 0.57$$

c. BMI

$$1. \frac{\text{Likelihood BMI (ya)}}{J_1} = \frac{0}{100} = 0$$

$$2. \frac{\text{Likelihood BMI (Tidak)}}{J_2} = \frac{7}{100} = 0,07$$

d. Smoker

$$1. \frac{\text{Likelihood Smoker (ya)}}{J_1} = \frac{45}{100} = 0,45$$

$$2. \frac{\text{Likelihood Smoker (Tidak)}}{J_2} = \frac{56}{100} = 0,56$$

e. PhysActivity

$$1. \frac{\text{Likelihood PhyActivity (ya)}}{J_1} = \frac{50}{100} = 0.50$$

$$2. \frac{\text{Likelihood PhyActivity (Tidak)}}{J2} = \frac{79}{100} = 0,79$$

## f. Fruits

$$1. \frac{\text{Likelihood Fruits (ya)}}{J1} = \frac{52}{100} = 0,52$$

$$2. \frac{\text{Likelihood Fruits (Tidak)}}{J2} = \frac{30}{100} = 0,30$$

## g. Veggies

$$1. \frac{\text{Likelihood Veggies (ya)}}{J1} = \frac{67}{100} = 0,67$$

$$2. \frac{\text{Likelihood Veggies (Tidak)}}{J2} = \frac{85}{100} = 0,85$$

## h. HvyAlcoholConsump

$$1. \frac{\text{Likelihood HvyAlcoholConsump (ya)}}{J1} = \frac{96}{100} = 0,96$$

$$2. \frac{\text{Likelihood HvyAlcoholConsump (Tidak)}}{J2} = \frac{94}{100} = 0,94$$

## i. GenHlth

$$1. \frac{\text{Likelihood GenHlth (ya)}}{J1} = \frac{16}{100} = 0,16$$

$$2. \frac{\text{Likelihood GenHelth (Tidak)}}{J2} = \frac{48}{100} = 0,48$$

## j. MentHlth

$$1. \frac{\text{Likelihood MenHlth (ya)}}{J1} = \frac{1}{100} = 0,01$$

$$2. \frac{\text{Likelihood MenHlth (Tidak)}}{J2} = \frac{5}{100} = 0,05$$

## k. PhysHlth

$$1. \frac{\text{Likelihood PhysHlth (ya)}}{J1} = \frac{48}{37163} = 0,48$$

$$2. \frac{\text{Likelihood PhyHlth (Tidak)}}{J2} = \frac{74}{33529} = 0,74$$

## 1. Diabetes

$$1. \frac{\text{Likelihood Diabetes (ya)}}{J_1} = \frac{0}{100} = 0$$

$$2. \frac{\text{Likelihood Diabetes (Tidak)}}{J_2} = \frac{100}{100} = 1$$

## ❖ Data uji Ketiga

## a. Age

$$1. \frac{\text{Likelihood Age (ya)}}{J_1} = \frac{0}{100} = 0$$

$$2. \frac{\text{Likelihood Age (Tidak)}}{J_2} = \frac{10}{100} = 0,1$$

## b. Sex

$$1. \frac{\text{Likelihood Sex (ya)}}{J_1} = \frac{41}{100} = 0,41$$

$$2. \frac{\text{Likelihood Sex (Tidak)}}{J_2} = \frac{43}{100} = 0,43$$

## c. BMI

$$1. \frac{\text{Likelihood BMI (ya)}}{J_1} = \frac{4}{100} = 0,04$$

$$2. \frac{\text{Likelihood BMI (Tidak)}}{J_2} = \frac{7}{100} = 0,07$$

## d. Smoker

$$1. \frac{\text{Likelihood Smoker (ya)}}{J_1} = \frac{45}{100} = 0,45$$

$$2. \frac{\text{Likelihood Smoker (Tidak)}}{J_2} = \frac{56}{100} = 0,56$$

## e. PhysActivity

$$1. \frac{\text{Likelihood PhyActivity (ya)}}{J_1} = \frac{50}{100} = 0,5$$

$$2. \frac{\text{Likelihood PhyActivity (Tidak)}}{J_2} = \frac{79}{100} = 0,79$$

## f. Fruits

$$1. \frac{\text{Likelihood Fruits (ya)}}{J_1} = \frac{48}{100} = 0,48$$

$$2. \frac{\text{Likelihood Fruits (Tidak)}}{J_2} = \frac{70}{100} = 0,7$$

g. Veggies

$$1. \frac{\text{Likelihood Veggies (ya)}}{J_1} = \frac{67}{100} = 0,67$$

$$2. \frac{\text{Likelihood Veggies (Tidak)}}{J_2} = \frac{85}{100} = 0,85$$

h. HvyAlcoholConsump

$$1. \frac{\text{Likelihood HvyAlcoholConsump (ya)}}{J_1} = \frac{94}{100} = 0,94$$

$$2. \frac{\text{Likelihood HvyAlcoholConsump (Tidak)}}{J_2} = \frac{96}{100} = 0,96$$

i. GenHlth

$$1. \frac{\text{Likelihood GenHlth (ya)}}{J_1} = \frac{3}{100} = 0,03$$

$$2. \frac{\text{Likelihood GenHelth (Tidak)}}{J_2} = \frac{26}{100} = 0,26$$

j. MentHlth

$$1. \frac{\text{Likelihood MenHlth (ya)}}{J_1} = \frac{60}{100} = 0,6$$

$$2. \frac{\text{Likelihood MenHlth (Tidak)}}{J_2} = \frac{70}{100} = 0,7$$

k. PhysHlth

$$1. \frac{\text{Likelihood PhysHlth (ya)}}{J_1} = \frac{48}{100} = 0,48$$

$$2. \frac{\text{Likelihood PhyHlth (Tidak)}}{J_2} = \frac{74}{100} = 0,74$$

l. Diabetes

$$1. \frac{\text{Likelihood Diabetes (ya)}}{J_1} = \frac{0}{100} = 0$$

$$2. \frac{\text{Likelihood Diabetes (Tidak)}}{J_2} = \frac{100}{100} = 1$$

#### 4. Mencari *Likelihood Probability*:

##### ❖ Data uji Pertama

- a. Probabilitas Diagnosis dengan hasil Healthy (Ya / beresiko) berdasarkan 12 Likelihood Probability dari setiap atribut.  
 $0.06 * 0.59 * 0 * 0.45 * 0.50 * 0.48 * 0.67 * 0.96 * 0.35 * 0.60 * 0.02 * 0 = 0$
- b. Probabilitas Diagnosis dengan hasil Healthy (Tidak / Aman) berdasarkan 12 Likelihood Probability dari setiap atribut.  
 $0.08 * 0.57 * 0 * 0.56 * 0.21 * 0.70 * 0.85 * 0.94 * 0.19 * 0.70 * 0.01 * 0.1 = 0$

##### ❖ Data uji Kedua

- a. Probabilitas Diagnosis dengan hasil Healthy (Ya / beresiko) berdasarkan 12 Likelihood Probability dari setiap atribut.  
 $0 * 0.59 * 0 * 0.45 * 0.50 * 0.52 * 0.67 * 0.96 * 0.16 * 0.01 * 0.48 * 0 = 0$
- b. Probabilitas Diagnosis dengan hasil Healthy (Tidak / Aman) berdasarkan 12 Likelihood Probability dari setiap atribut.  
 $0.05 * 0.57 * 0.07 * 0.56 * 0.79 * 0.30 * 0.85 * 0.94 * 0.48 * 0.05 * 0.74 * 1 = 0.0000037562154243$

##### ❖ Data uji KeTiga

- a. Probabilitas Diagnosis dengan hasil Healthy (Ya / beresiko) berdasarkan 12 Likelihood Probability dari setiap atribut.  
 $0 * 0.41 * 0.04 * 0.45 * 0.50 * 0.52 * 0.67 * 0.96 * 0.03 * 0.6 * 0.48 * 0 = 0$
- b. Probabilitas Diagnosis dengan hasil Healthy (Tidak / Aman) berdasarkan 12 Likelihood Probability dari setiap atribut.  
 $0.1 * 0.43 * 0.07 * 0.56 * 0.79 * 0.30 * 0.85 * 0.94 * 0.26 * 0.7 * 0.74 * 1 = 0.0000429885459407$

#### 5. Menghitung *Posterior Probability*:

##### ❖ Data Uji Pertama

- a. Likelihood  $J_1 * J_1$

$$0 * 0.5 = 0$$

b. Likelihood J2 \* J2

$$0 * 0.5 = 0$$

❖ **Data Uji Ke Dua**

a. Likelihood J1 \* J1

$$0 * 0.5 = 0$$

b. Likelihood J2 \* J2

$$0.0000037562154243 * 0.5 = 0.00000187810771215$$

❖ **Data Uji Ke Tiga**

a. Likelihood J1 \* J1

$$0 * 0.5 = 0$$

b. Likelihood J2 \* J2

$$0.0000429885459407 * 0.5 = 0.000000000002157$$

Dari perhitungan di atas, dapat diambil kesimpulan melalui tabel berikut. Tabel tersebut menunjukkan hasil perkalian likelihood dari dua jenis uji, J1 dan J2, yang mewakili dua kondisi yang diuji. Penilaian risiko atau keamanan didasarkan pada perbandingan nilai likelihood masing-masing uji. Ketika nilai likelihood J1 \* J1 lebih besar, kategori "Beresiko" diterapkan, dan sebaliknya, jika nilai likelihood J2 \* J2 lebih besar, kategori "Aman" diterapkan. Kategorisasi ini memberikan gambaran kecenderungan hasil uji terhadap kondisi berisiko atau aman. Namun, perlu diperhatikan bahwa hasil ini harus ditafsirkan dengan hati-hati karena model Naive Bayes yang digunakan belum ditingkatkan. Terdapat ketidakakuratan atau kesalahan yang mungkin timbul terutama ketika data memiliki nilai nol. Oleh karena itu, hasil kesimpulan ini sebaiknya dianggap sebagai panduan awal dan tidak sepenuhnya akurat.

**Tabel 3.4** Hasil prediksi perhitungan Naive Bayes Biasa

Data Uji	Probabilitas Diagnosis (Ya/Beresiko)	Probabilitas Diagnosis (Tidak/Aman)	Hasil
1	0	0	?

2	0	0.00000187810771215	Aman
3	0	0.000000000002157	Aman

Selanjutnya, peningkatan model Naive Bayes dapat dilakukan untuk mengatasi masalah ketidakakuratan, meningkatkan akurasi, dan meminimalkan risiko kesalahan dalam klasifikasi. Hal ini menjadi penting terutama jika data yang diuji memiliki nilai nol yang signifikan. Selama peningkatan model, evaluasi dan validasi secara menyeluruh perlu dilakukan untuk memastikan hasil yang lebih dapat diandalkan dan akurat. Sehingga, kesimpulan dari tabel perhitungan ini harus diartikan sebagai suatu indikasi awal dengan peringatan atas potensi ketidakpastian.

Pada proses ini, perlu memberikan perhatian khusus terhadap nilai probabilitas yang sebelumnya bernilai 0. Setelah mendapatkan nilai probabilitas awal dari Likelihood Probability untuk masing-masing atribut, dapat dilanjutkan dengan penerapan teknik Laplacian Correction. Teknik ini digunakan untuk menghindari nilai probabilitas yang nol, yang dapat menyebabkan masalah dalam perhitungan Naive Bayes.

### 3.4.1 Laplacian Correction

*Laplacian correction* biasanya diterapkan pada kasus di mana nilai probabilitas Likelihood Probability bernilai nol. Ini dilakukan dengan menambahkan suatu nilai yang kecil (biasanya 1) pada jumlah frekuensi masing-masing kategori dan pembagi. Dalam konteks ini, dapat digunakan rumus umum *Laplacian correction* sebagai langkah tambahan dalam perhitungan yang sebelumnya dilakukan.

Mari terapkan *Laplacian correction* pada beberapa percobaan yang memiliki nilai 0, agar metode Naive Bayes menjadi lebih akurat:

#### ❖ Data uji Pertama

a. Age

$$1. \frac{\text{Likelihood Age (ya)}}{J_1} = \frac{6+1}{100+12} = 0.0625$$

$$2. \frac{\text{Likelihood Age (Tidak)}}{J_2} = \frac{8+1}{100+12} = 0.0804$$

## b. Sex

$$1. \frac{Likelihood\ Sex\ (ya)}{J_1} = \frac{59+1}{100+12} = 0.5357$$

$$2. \frac{Likelihood\ Sex\ (Tidak)}{J_2} = \frac{57+1}{100+12} = 0.5179$$

## c. BMI

$$1. \frac{Likelihood\ BMI\ (ya)}{J_1} = \frac{0+1}{100+12} = 0.0089$$

$$2. \frac{Likelihood\ BMI\ (Tidak)}{J_2} = \frac{0+1}{100+12} = 0.0089$$

## d. Smoker

$$1. \frac{Likelihood\ Smoker\ (ya)}{J_1} = \frac{45+1}{100+12} = 0.4107$$

$$2. \frac{Likelihood\ Smoker\ (Tidak)}{J_2} = \frac{56+1}{100+12} = 0.5089$$

## e. PhysActivity

$$1. \frac{Likelihood\ PhyActivity\ (ya)}{J_1} = \frac{50+1}{100+12} = 0.4554$$

$$2. \frac{Likelihood\ PhyActivity\ (Tidak)}{J_2} = \frac{21+1}{100+12} = 0.1964$$

## f. Fruits

$$1. \frac{Likelihood\ Fruits\ (ya)}{J_1} = \frac{48+1}{100+12} = 0.4375$$

$$2. \frac{Likelihood\ Fruits\ (Tidak)}{J_2} = \frac{70+1}{100+12} = 0.6339$$

## g. Veggies

$$1. \frac{Likelihood\ Veggies\ (ya)}{J_1} = \frac{67+1}{100+12} = 0.6071$$

$$2. \frac{Likelihood\ Veggies\ (Tidak)}{J_2} = \frac{85+1}{100+12} = 0.7679$$

## h. HvyAlcoholConsump

$$1. \frac{Likelihood\ HvyAlcoholConsump\ (ya)}{J_1} = \frac{96+1}{100+12} = 0.8661$$

$$2. \frac{Likelihood\ HvyAlcoholConsump\ (Tidak)}{J_2} = \frac{94+1}{100+12} = 0.8482$$

## i. GenHlth

$$1. \frac{\text{Likelihood GenHlth (ya)}}{J_1} = \frac{35+1}{100+12} = 0.3214$$

$$2. \frac{\text{Likelihood GenHelth (Tidak)}}{J_2} = \frac{19+1}{100+12} = 0.1786$$

## j. MentHlth

$$1. \frac{\text{Likelihood MenHlth (ya)}}{J_1} = \frac{60+1}{100+12} = 0.5446$$

$$2. \frac{\text{Likelihood MenHlth (Tidak)}}{J_2} = \frac{70+1}{100+12} = 0.6339$$

## k. PhysHlth

$$1. \frac{\text{Likelihood PhysHlth (ya)}}{J_1} = \frac{2+1}{100+12} = 0.0268$$

$$2. \frac{\text{Likelihood PhyHlth (Tidak)}}{J_2} = \frac{1+1}{100+12} = 0.0167$$

## l. Diabetes

$$1. \frac{\text{Likelihood Diabetes (ya)}}{J_1} = \frac{0+1}{100+12} = 0.0089$$

$$2. \frac{\text{Likelihood Diabetes (Tidak)}}{J_2} = \frac{100+1}{100+12} = 0.8868$$

## ❖ Data uji Kedua

## a. Age

$$1. \frac{\text{Likelihood Age (ya)}}{J_1} = \frac{0+1}{100+12} = 0.0089$$

$$2. \frac{\text{Likelihood Age (Tidak)}}{J_2} = \frac{5}{100} = 0,05$$

## b. Sex

$$1. \frac{\text{Likelihood Sex (ya)}}{J_1} = \frac{59+1}{100+12} = 0.5357$$

$$2. \frac{\text{Likelihood Sex (Tidak)}}{J_2} = \frac{57}{100} = 0.57$$

## c. BMI

$$1. \frac{\text{Likelihood BMI (ya)}}{J_1} = \frac{0+1}{100+12} = 0.0089$$

$$2. \frac{\text{Likelihood BMI (Tidak)}}{J_2} = \frac{7}{100} = 0,07$$

d. Smoker

$$1. \frac{\text{Likelihood Smoker (ya)}}{J_1} = \frac{45+1}{100+12} = 0,4107$$

$$2. \frac{\text{Likelihood Smoker (Tidak)}}{J_2} = \frac{56}{100} = 0,56$$

e. PhysActivity

$$1. \frac{\text{Likelihood PhyActivity (ya)}}{J_1} = \frac{50+1}{100+12} = 0.4554$$

$$2. \frac{\text{Likelihood PhyActivity (Tidak)}}{J_2} = \frac{79}{100} = 0.79$$

f. Fruits

$$1. \frac{\text{Likelihood Fruits (ya)}}{J_1} = \frac{52+1}{100+12} = 0.4732$$

$$2. \frac{\text{Likelihood Fruits (Tidak)}}{J_2} = \frac{30}{100} = 0,30$$

g. Veggies

$$1. \frac{\text{Likelihood Veggies (ya)}}{J_1} = \frac{67+1}{100+12} = 0.6071$$

$$2. \frac{\text{Likelihood Veggies (Tidak)}}{J_2} = \frac{85}{100} = 0,85$$

h. HvyAlcoholConsump

$$1. \frac{\text{Likelihood HvyAlcoholConsump (ya)}}{J_1} = \frac{96+1}{100+12} = 0.8661$$

$$2. \frac{\text{Likelihood HvyAlcoholConsump (Tidak)}}{J_2} = \frac{94}{100} = 0,94$$

i. GenHlth

$$1. \frac{\text{Likelihood GenHlth (ya)}}{J_1} = \frac{16+1}{100+12} = 0.1579$$

$$2. \frac{\text{Likelihood GenHelth (Tidak)}}{J_2} = \frac{48}{100} = 0,48$$

j. MentHlth

$$1. \frac{\text{Likelihood MenHlth (ya)}}{J_1} = \frac{1+1}{100+12} = 0.0167$$

$$2. \frac{\text{Likelihood MenHlth (Tidak)}}{J_2} = \frac{5}{100} = 0,05$$

k. PhysHlth

$$1. \frac{\text{Likelihood PhysHlth (ya)}}{J_1} = \frac{48+1}{37163+12} = 0.4375$$

$$2. \frac{\text{Likelihood PhyHlth (Tidak)}}{J_2} = \frac{74}{33529} = 0,74$$

l. Diabetes

$$1. \frac{\text{Likelihood Diabetes (ya)}}{J_1} = \frac{0+1}{100+12} = 0.0089$$

$$2. \frac{\text{Likelihood Diabetes (Tidak)}}{J_2} = \frac{100}{100} = 1$$

#### ❖ Data uji Ketiga

a. Age

$$1. \frac{\text{Likelihood Age (ya)}}{J_1} = \frac{0+1}{100+12} = 0.0089$$

$$2. \frac{\text{Likelihood Age (Tidak)}}{J_2} = \frac{10}{100} = 0,1$$

b. Sex

$$1. \frac{\text{Likelihood Sex (ya)}}{J_1} = \frac{41+1}{100+12} = 0.3864$$

$$2. \frac{\text{Likelihood Sex (Tidak)}}{J_2} = \frac{43}{100} = 0.43$$

c. BMI

$$1. \frac{\text{Likelihood BMI (ya)}}{J_1} = \frac{4+1}{100+12} = 0.0374$$

$$2. \frac{\text{Likelihood BMI (Tidak)}}{J_2} = \frac{7}{100} = 0.07$$

## d. Smoker

$$1. \frac{\text{Likelihood Smoker (ya)}}{J_1} = \frac{45+1}{100+12} = 0.4107$$

$$2. \frac{\text{Likelihood Smoker (Tidak)}}{J_2} = \frac{56}{100} = 0.56$$

## e. PhysActivity

$$1. \frac{\text{Likelihood PhyActivity (ya)}}{J_1} = \frac{50+1}{100+12} = 0.4554$$

$$2. \frac{\text{Likelihood PhyActivity (Tidak)}}{J_2} = \frac{79}{100} = 0,79$$

## f. Fruits

$$1. \frac{\text{Likelihood Fruits (ya)}}{J_1} = \frac{48+1}{100+12} = 0.4375$$

$$2. \frac{\text{Likelihood Fruits (Tidak)}}{J_2} = \frac{70}{100} = 0,7$$

## g. Veggies

$$1. \frac{\text{Likelihood Veggies (ya)}}{J_1} = \frac{67+1}{100+12} = 0.6071$$

$$2. \frac{\text{Likelihood Veggies (Tidak)}}{J_2} = \frac{85}{100} = 0.85$$

## h. HvyAlcoholConsump

$$1. \frac{\text{Likelihood HvyAlcoholConsump (ya)}}{J_1} = \frac{94+1}{100+12} = 0.8482$$

$$2. \frac{\text{Likelihood HvyAlcoholConsump (Tidak)}}{J_2} = \frac{96}{100} = 0.96$$

## i. GenHlth

$$1. \frac{\text{Likelihood GenHlth (ya)}}{J_1} = \frac{3+1}{100+12} = 0.0357$$

$$2. \frac{\text{Likelihood GenHelth (Tidak)}}{J_2} = \frac{26}{100} = 0,26$$

## j. MentHlth

$$1. \frac{\text{Likelihood MenHlth (ya)}}{J_1} = \frac{60+1}{100+12} = 0.5446$$

$$2. \frac{\text{Likelihood MenHlth (Tidak)}}{J_2} = \frac{70}{100} = 0,7$$

k. PhysHlth

$$1. \frac{\text{Likelihood PhysHlth (ya)}}{J_1} = \frac{48+1}{100+12} = 0.4375$$

$$2. \frac{\text{Likelihood PhyHlth (Tidak)}}{J_2} = \frac{74}{100} = 0,74$$

l. Diabetes

$$1. \frac{\text{Likelihood Diabetes (ya)}}{J_1} = \frac{0+1}{100+12} = 0.0089$$

$$2. \frac{\text{Likelihood Diabetes (Tidak)}}{J_2} = \frac{100}{100} = 1$$

1. Mencari *Likelihood Probability* yang baru

❖ **Data uji Pertama**

a. Probabilitas Diagnosis dengan hasil Healthy (Ya / beresiko) berdasarkan 17 Likelihood Probability dari setiap atribut.

$$0.0625 * 0.5357 * 0.0089 * 0.4107 * 0.4554 * 0.4375 * 0.6071 * 0.8661 * 0.3214 * 0.5446 * 0.0268 * 0.0089 = 0.00000000007734$$

b. Probabilitas Diagnosis dengan hasil Healthy (Tidak / Aman) berdasarkan 17 Likelihood Probability dari setiap atribut.

$$0.0804 * 0.5179 * 0.0089 * 0.5089 * 0.1964 * 0.6339 * 0.7679 * 0.8482 * 0.1786 * 0.6339 * 0.0167 * 0.8868 = 0.0000000338386$$

❖ **Data uji Kedua**

a. Probabilitas Diagnosis dengan hasil Healthy (Ya / beresiko) berdasarkan 12 Likelihood Probability dari setiap atribut.

$$0.0089 * 0.5357 * 0.0089 * 0.4107 * 0.4554 * 0.4732 * 0.6071 * 0.8661 * 0.1579 * 0.0167 * 0.4375 * 0.0089 = 0.0000000148522$$

- b. Probabilitas Diagnosis dengan hasil Healthy (Tidak / Aman) berdasarkan 12 Likelihood Probability dari setiap atribut.

$$0.05 * 0.57 * 0.07 * 0.56 * 0.79 * 0.30 * 0.85 * 0.94 * 0.48 * 0.05 * 0.74 * 1 = 0.0000037562154243$$

❖ **Data uji KeTiga**

- a. Probabilitas Diagnosis dengan hasil Healthy (Ya / beresiko) berdasarkan 12 Likelihood Probability dari setiap atribut.

$$0.0089 * 0.3864 * 0.0374 * 0.4107 * 0.4554 * 0.4375 * 0.6071 * 0.8482 * 0.0357 * 0.5446 * 0.4375 * 0.0089 = 0.0000000167203$$

- b. Probabilitas Diagnosis dengan hasil Healthy (Tidak / Aman) berdasarkan 12 Likelihood Probability dari setiap atribut.

$$0.1 * 0.43 * 0.07 * 0.56 * 0.79 * 0.30 * 0.85 * 0.94 * 0.26 * 0.7 * 0.74 * 1 = 0.0000429885459407$$

2. **Menghitung *Posterior Probability* yang baru**

❖ **Data Uji Pertama**

- b. Likelihood J1 \* J1

$$0.00000000007734 * 0.5 = 0.00000000003867$$

- c. Likelihood J2 \* J2

$$0.0000000338386 * 0.5 = 0.0000000169193$$

❖ **Data Uji Ke Dua**

- a. Likelihood J1 \* J1

$$0.0000000148522 * 0.5 = 0.0000000074261$$

- c. Likelihood J2 \* J2

$$0.0000037562154243 * 0.5 = 0.00000187810771215$$

❖ **Data Uji Ke Tiga**

- c. Likelihood J1 \* J1

$$0.0000000167203 * 0.5 = 0.00000000836015$$

- d. Likelihood J2 \* J2

$$0.0000429885459407 * 0.5 = 0.00000000002157$$

Setelah mengimplementasikan perbaikan pada metode Naive Bayes, dapat menyimpulkan bahwa perubahan tersebut memberikan kontribusi signifikan terhadap akurasi analisis. Dengan menggunakan pendekatan yang lebih canggih, dapat mendapatkan hasil yang lebih dapat diandalkan, sehingga memungkinkan pengambilan keputusan yang lebih baik. Proses peningkatan ini melibatkan penyesuaian parameter dan optimalisasi model, yang secara keseluruhan meningkatkan performa algoritma.

Selain itu, perubahan yang telah diterapkan memungkinkan untuk melakukan revisi pada tabel yang sebelumnya telah dibuat. Tabel tersebut menjadi lebih representatif dan menggambarkan hasil analisis dengan lebih akurat. Dengan adanya perubahan ini, dapat mengandalkan data yang dihasilkan untuk mendukung keputusan yang diambil. Untuk Hasil prediksi perhitungan Naïve Bayes setelah ditingkatkan dapat dilihat pada Tabel 3.5, sementara perbandingan antara Naïve Bayes biasa dan Improved Naïve Bayes disajikan dalam Tabel 3.6.

**Tabel 3.5** Hasil prediksi perhitungan Naïve Bayes setelah ditingkatkan

Data Uji	Probabilitas Diagnosis (Ya/Beresiko)	Probabilitas Diagnosis (Tidak/Aman)	Hasil
1	0.000000000003867	0.0000429885459407	Aman
2	0.0000000074261	0.00000187810771215	Aman
3	0.00000000836015	0.000000000002157	Beresiko

**Tabel 3.6** Perbandingan hasil antara *Naive Bayes* Biasa dengan *Improved Naive Bayes*

Data Uji	Probabilitas Diagnosis (Ya/Beresiko)		Probabilitas Diagnosis (Tidak /Aman)		Hasil	
	Naïve Bayes Biasa	Improved Naive Bayes	Naïve Bayes Biasa	Improved Naive Bayes	Naïve Bayes Biasa	Improved Naive Bayes
1	0	0.000000000003867	0	0.0000429885459407	?	Aman
2	0	0.0000000074261	0.00000187810771215	0.00000187810771215	Aman	Aman
3	0	0.00000000836015	0.000000000002157	0.000000000002157	Aman	Beresiko

Peningkatan pada metode analisis seperti *Naive Bayes* tidak hanya berdampak pada akurasi, tetapi juga membuka peluang untuk eksplorasi dan pengembangan lebih lanjut. Keberhasilan implementasi ini menunjukkan pentingnya terus-menerus menggali inovasi dalam bidang analisis data untuk mendapatkan pemahaman yang lebih dalam dan mendukung pengambilan keputusan yang lebih cerdas.

### 3.5 Struktur Tabel

Struktur tabel ini merinci struktur dan lokasi penyimpanan data yang digunakan untuk keperluan sistem yang akan dikembangkan. Berikut adalah gambaran struktur dari tabel yang akan digunakan.

#### 1. Tabel Admin

Tabel admin ini dibuat supaya bisa mengakses menu admin. Data dari admin tersebut tersimpan dalam tabel admin. Struktur tabel ini dapat dilihat pada tabel 3.7

**Tabel 3.7** admin

Name	Type	Length	Key
id	int	11	primary key
Username	varchar	50	
Password	varchar	150	

#### 2. Tabel Data Pengguna

Tabel ini berisi informasi pengguna beserta hasil klasifikasi risiko kolesterol tinggi. Detail datanya terdapat dalam tabel data\_pengguna yang dapat ditemukan pada Tabel 3.8

**Tabel 3.8** data\_pengguna

Variable	Type	Length	Key
No	int	11	primary key
Name	varchar	50	
Age	float	-	
Sex	float	-	

BMI	float	-	
Smoker	float	-	
PhysActivity	float	-	
Fruits	float	-	
Veggies	float	-	
HvyAlcoholConsump	float	-	
GenHlth	float	-	
MentHlth	float	-	
PhysHlth	float	-	
Diabetes	float	-	

### 3. Tabel Data *Training*

Tabel ini berperan sebagai penyimpan data pengujian untuk menguji tingkat akurasi probabilitas terbesar, sebagaimana tercantum pada Tabel

3.9

**Tabel 3.9** data\_training

Variable	Type	Length	Key
No	int	11	primary key
Age	float	-	
Sex	float	-	
BMI	float	-	
Smoker	float	-	
PhysActivity	float	-	
Fruits	float	-	
Veggies	float	-	
HvyAlcoholConsump	float	-	
GenHlth	float	-	
MentHlth	float	-	
PhysHlth	float	-	
Diabetes	float	-	

### 3.6 Desain Antarmuka

#### A. Halaman Perkenalan

Halaman perkenalan atau intro di sebuah situs web adalah bagian yang dirancang untuk memberikan gambaran singkat kepada pengunjung tentang isi, tujuan, dan nilai yang disajikan oleh situs tersebut. Fungsi utamanya adalah untuk menyambut pengunjung dan memberikan pengenalan yang jelas tentang apa yang dapat diharapkan dari pengalaman mereka di situs tersebut.

Pada halaman perkenalan, biasanya terdapat informasi singkat mengenai misi atau visi situs, layanan atau produk yang ditawarkan, dan mungkin juga sejarah singkat atau latar belakang perusahaan atau individu di balik situs tersebut. Desain dan pesan pada halaman intro dirancang untuk menarik perhatian pengunjung dan memberikan navigasi yang jelas untuk memandu mereka ke bagian-bagian penting dari situs.

Halaman perkenalan yang efektif dapat menciptakan kesan positif pertama kali, memotivasi pengunjung untuk menjelajahi lebih lanjut, dan membangun keterlibatan awal dengan konten atau layanan yang ditawarkan oleh situs web tersebut. Untuk halaman perkenalan dapat dilihat pada gambar 3.7.



**Gambar 3.7** halaman perkenalan

## B. Halaman *Home*

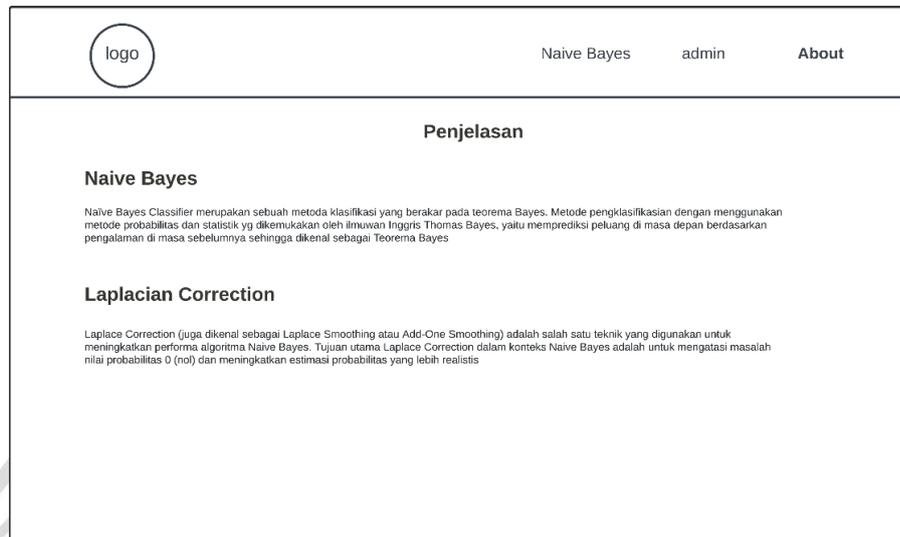
Halaman *home* menampilkan 13 pertanyaan yang harus diisi oleh pengguna aplikasi web untuk memprediksi risiko kolesterol tinggi. Data input ini menjadi data uji untuk perhitungan. Pertanyaan terbagi ke beberapa slide dengan tombol "Selanjutnya" untuk pertanyaan berikutnya dan "Sebelumnya" untuk kembali ke pertanyaan sebelumnya. Tampilan halaman *home* dapat dilihat pada Gambar 3.8.

The screenshot shows a web interface for predicting high cholesterol risk. At the top, there is a navigation bar with a circular logo containing the word 'logo', and three links: 'Naive Bayes', 'admin', and 'About'. The main content area is titled 'Prediksi Resiko Kolesterol Tinggi Berdasarkan Data Kesehatan'. It contains several input fields and dropdown menus: 'Nama :', 'Usia :', 'Jenis Kelamin :', and a question 'Apakah anda pernah mengecek kolesterol anda dalam 5 tahun terakhir :'. Below these is the 'Index Massa Tubuh (BMI)' section with 'Tinggi (cm)' and 'Berat (kg)' input fields. At the bottom of the form are two buttons: 'Sebelumnya' and 'Selanjutnya'.

**Gambar 3.8** halaman home

## C. Halaman *About*

Halaman *About* menjelaskan secara rinci konsep metode *Naive Bayes* dan teknik *Laplacian correction* yang digunakan dalam aplikasi ini. Ini memberikan pemahaman yang lebih mendalam kepada pengunjung mengenai prinsip dasar yang mendasari fungsionalitas aplikasi, membantu mereka memahami bagaimana algoritma ini digunakan untuk memprediksi risiko kolesterol tinggi. Tampilan halaman *About* dapat dilihat pada Gambar 3.9.

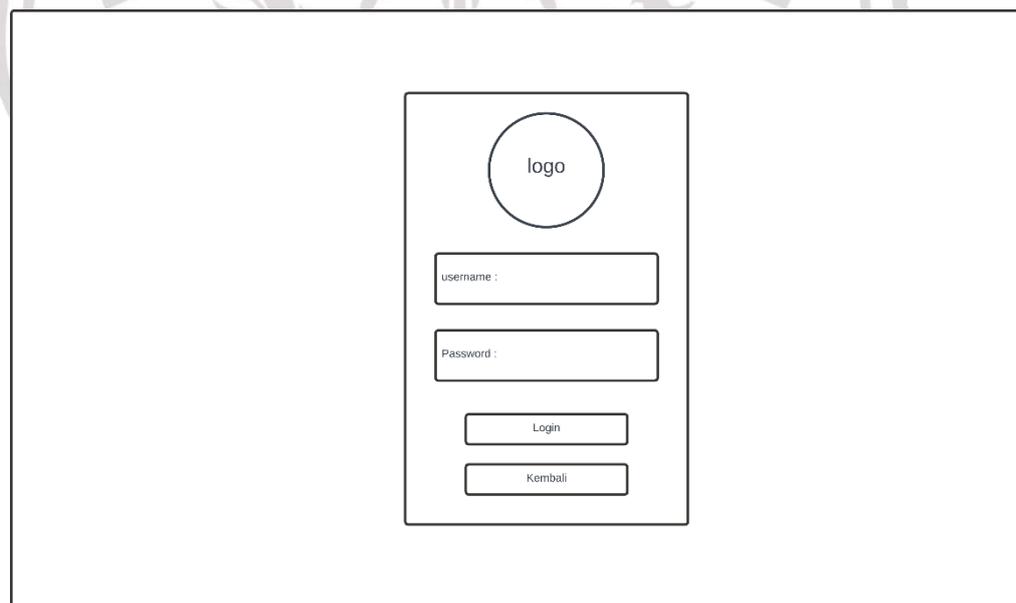


**Gambar 3.9** halaman About

#### D. Halaman Login Admin

Halaman login admin dibuat eksklusif untuk administrator. Hanya admin yang memiliki akses ke menu admin, memastikan keamanan data dan kontrol yang ketat terhadap fungsionalitas dan informasi pengguna.

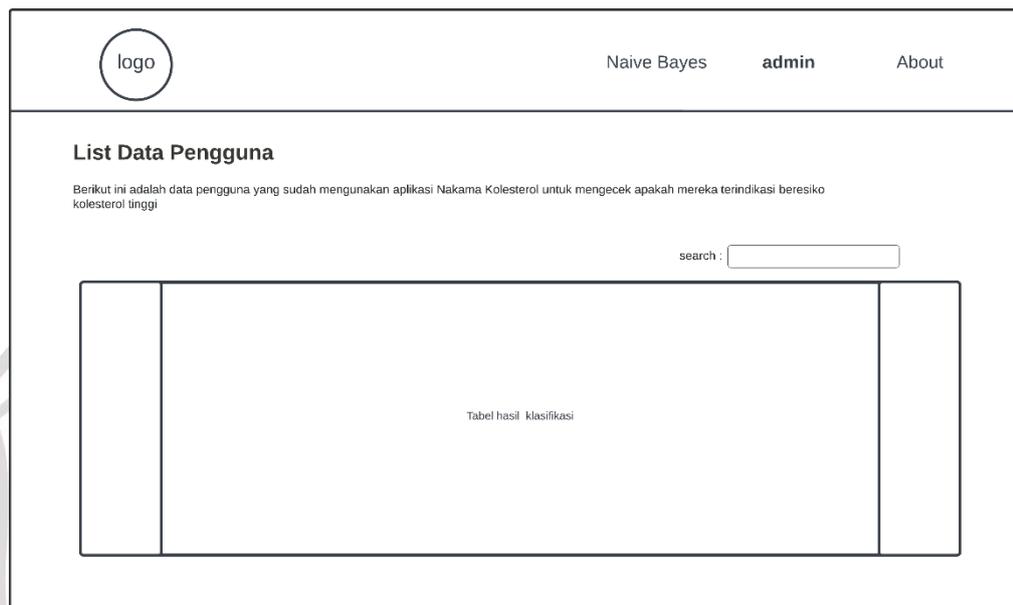
Tampilan halaman Login Admin ditampilkan pada Gambar 3.10.



**Gambar 3.10** halaman Login Admin

### E. Halaman Admin

Halaman admin memuat data pengguna yang menggunakan aplikasi web. Dipersembahkan dalam tabel, halaman ini eksklusif untuk admin guna memastikan privasi dan kontrol penuh terhadap informasi pengguna. Tampilan halaman Admin ditampilkan pada Gambar 3.11.



**Gambar 3.11** halaman Admin

### 3.7 Spesifikasi Kebutuhan Sistem

Kebutuhan dalam pembuatan sistem Prediksi resiko kolesterol menggunakan data kesehatan dengan metode *Improve Naive Bayes* dibutuhkan spesifikasi perangkat keras dan perangkat lunak, diantaranya adalah

#### 3.7.1 Kebutuhan Perangkat Keras

Sistem perangkat keras adalah komponen-komponen pendukung kinerja dari sistem komputer. Komponen-komponen yang dapat dapat dipakai untuk menjalankan sistem sebagai berikut :

1. Prosesor AMD RYZEN 5
2. Memory RAM 4 GB atau lebih
3. Monitor 14 inch
4. Solid State Drive 512 GB

### 3.7.2 Kebutuhan perangkat Lunak

1. Sistem Operasi windows 10
2. Web server XAMPP V.3.2.4
3. Basis data My SQL
4. Software Development Visual Studio Code

## 3.8 Skenario Pengujian

Digunakan teknik K-fold cross-validation sebagai pendekatan utama untuk mengukur kinerja model klasifikasi. K-fold cross-validation adalah metode pengujian yang umum digunakan untuk mengevaluasi sejauh mana model dapat memberikan prediksi yang konsisten dan dapat diterapkan secara umum.

### 3.8.1 K-fold Cross-Validation

Teknik *K-fold cross-validation* melibatkan pembagian dataset menjadi k subset atau "*fold*," di mana setiap subset berperan sebagai data uji satu kali, sedangkan k-1 subset lainnya digunakan sebagai data pelatihan. Pada penelitian ini, teknik 5-fold cross-validation digunakan untuk menguji model. Teknik ini melibatkan pembagian dataset menjadi lima subset yang sama besar. Satu fold digunakan sebagai data uji secara bergantian, sedangkan empat fold sisanya digunakan sebagai data latih. Proses ini diulangi selama lima kali, sehingga setiap fold memiliki kesempatan untuk menjadi data uji tepat satu kali. Pada setiap iterasi, metrik evaluasi seperti akurasi, presisi, recall, dan F1 score dihitung. Nilai rata-rata dari metrik evaluasi di semua iterasi dijadikan sebagai hasil akhir untuk menilai kinerja model secara keseluruhan.

Penerapan K-fold cross-validation memberikan gambaran yang lebih kuat tentang kemampuan model untuk generalisasi, mengurangi risiko overfitting atau underfitting. Dengan menggunakan teknik ini, pengujian metode *improve Naive Bayes* dapat memberikan hasil yang lebih solid dan reliabel dalam mengevaluasi performa model terhadap dataset yang beragam.

Penting untuk dicatat bahwa data yang digunakan dalam pengujian hanya mencakup 200 data dari hasil survei BRFSS (*Behavioral Risk Factor*

*Surveillance System*) 2015. Rencana pengujian akurasi dapat dilihat pada Tabel 3.10 yang memuat informasi mengenai atribut akurasi.

**Tabel 3.10** Data Atribut Akurasi

Skenario	Pembagian Data		Akurasi
	Latih	Uji	
1	41-200	1-40	
2	1-40, 81-200	41-80	
3	1-80, 121-200	81-120	
4	1-120, 161-200	121-160	
5	1-160	161-200	

### 3.8.2 Pengujian Sistem

Dalam pengembangan sistem dengan Metode Extreme Programming (XP), fokus utama adalah kualitas dan keandalan. XP menekankan kolaborasi, fleksibilitas, dan tanggapan cepat terhadap perubahan. Pada tahap pengujian sistem, digunakan metode blackbox testing untuk evaluasi fungsionalitas tanpa memerlukan pemahaman mendalam terhadap struktur internal. Pengujian sistem dalam XP mencakup verifikasi fungsi kunci dan memastikan kesesuaian dengan kebutuhan pengguna. Prioritas diberikan pada pengujian sepanjang siklus pengembangan, memastikan deteksi dini dan perbaikan cepat terhadap potensi cacat.

Pengujian dilakukan dengan prinsip blackbox testing, fokus pada evaluasi eksternal sistem dengan skenario penggunaan umum, input valid/tidak valid, dan verifikasi respons sistem terhadap berbagai kondisi. Penekanan pada blackbox testing menciptakan keunggulan dalam pengujian fungsionalitas, memastikan output sesuai dengan harapan pengguna. Oleh karena itu, tahap pengujian sistem dalam XP menjadi landasan penting untuk memastikan kualitas akhir perangkat lunak yang dikembangkan. Rencana pengujian black box dapat dilihat pada Tabel 3.11.

**Tabel 3.11** Tabel Blackbox

Test Case Number	Test Case	Input	Action	Expected Output	Status
1	Input data user	1. Name 2. Age 3. Sex 4. BMI 5. Smoker 6. PhysActivity 7. Fruits 8. Veggies 9. HvyAlcoholConsump 10. GenHlth 11. MentHlth 12. PhysHlth 13. Diabetes	Masukan Semua data Yang dibutuhkan, sesuai pertanyaan.	data tersimpan ke database	
2	Perhitungan Resiko	-	Klick Tombol "submit"	Menampilkan hasil perhitungan dan presentasi resiko dari penyakit kolesterol tinggi	
3	Login Admin	1. Username 2. Password	Masukan Semua data Yang dibutuhkan, sesuai pertanyaan.	Masuk ke Menu Admin yang menampilkan data dari pengguna	

				yang sudah menggunakan aplikasi website	
--	--	--	--	---	--

