

## BAB IV

### IMPLEMENTASI

Bab ini membahas mengenai tahap pemahaman data, tahap *data cleaning*, tahap *data integration*, tahap *data selection*, tahap *data transformation*, tahap *data mining*, tahap *pattern evaluation*, dan analisis data.

#### 4.1 Tahap Pemahaman Data

Pada tahapan pemahaman data dijelaskan data apa saja yang dipakai dalam penelitian, jumlah data, dan dari mana data diperoleh.

##### 4.1.1 Penjelasan Data

Gejala dan penyakit yang ditemukan dalam dataset penelitian ini menggunakan bahasa Inggris, sehingga dalam penelitian ini semua gejala dan penyakit diterjemahkan ke dalam bahasa Indonesia. Dataset dari penyakit yang digunakan pada penelitian ini memiliki format file CSV (.csv) dan format file DATA (.data). Gambar 4.1 dan 4.2 menunjukkan bentuk awal dari data yang dikumpulkan.

Fever	Runny_Nose	Sore_Throat	Nasal_Symptoms	Non_PetAge_0-4	Age_10-19	Age_20-29	Age_30-39	Age_40-49	Gender_Male	Gender_Female	Severity_Low	Severity_High	Contact_Contract	Contact_XContract	Y_Country
0	1	1	1	0	1	1	1	0	0	1	0	0	0	0	0
1	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
2	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
3	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
4	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
5	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
6	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
7	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
8	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
9	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
10	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
11	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
12	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
13	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
14	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
15	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
16	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
17	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
18	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
19	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
20	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
21	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
22	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
23	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
24	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
25	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
26	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
27	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
28	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
29	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
30	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
31	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
32	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
33	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
34	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
35	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
36	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
37	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
38	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
39	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
40	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
41	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
42	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
43	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
44	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
45	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
46	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
47	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
48	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
49	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
50	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
51	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
52	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
53	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
54	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
55	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
56	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
57	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
58	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
59	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
60	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
61	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
62	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
63	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
64	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
65	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
66	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
67	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
68	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
69	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
70	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
71	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
72	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
73	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
74	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
75	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
76	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
77	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
78	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
79	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
80	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
81	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
82	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
83	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
84	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
85	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
86	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
87	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
88	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
89	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
90	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
91	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
92	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
93	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
94	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
95	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
96	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
97	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0
98	1	1	1	0	1	1	1	0	1	0	1	0	0	0	0

Gambar 4.2 Bentuk Data Awal dengan Format .data

Data yang digunakan dalam penelitian berjumlah 338.600 data, dengan rincian dataset penyakit obesitas memiliki data sebanyak 2.111 data, dataset penyakit tiroid memiliki data sebanyak 12.944 data, dataset penyakit kanker payudara memiliki data sebanyak 569 data, dataset penyakit umum memiliki data sebanyak 4.962 data, dataset penyakit jantung memiliki data sebanyak 1.214 data, dan dataset penyakit COVID-19 memiliki data sebanyak 316.800 data. Tabel 4.1 menjelaskan jenis penyakit dan jumlah data yang digunakan.

Tabel 4.1 Tabel Jumlah Data

Jenis Penyakit	Jumlah Data
Obesitas	2.111
Tiroid	12.944
Kanker Payudara	569
Umum	4.962
Jantung	1.214
COVID-19	316.800
<b>TOTAL</b>	<b>338.600</b>

Dataset penyakit obesitas didapatkan berdasarkan estimasi dari tingkat obesitas dari individu – individu yang berasal dari Meksiko, Peru dan Kolombia.

Data - data ini diambil berdasarkan gaya hidup, kebiasaan makan dan kondisi fisik responden. Responden terdiri dari laki – laki dan perempuan dengan rentang umur 17 sampai 50 tahun. Dataset ini memiliki 16 kolom gejala yang digunakan untuk mendiagnosis termasuk ke dalam kategori apa responden tersebut. Kolom – kolom gejala tersebut memiliki nama berupa kode sesuai dengan kuesioner yang telah dibuat oleh Palenchor dan Manotas [38]. Responden mengisi kuesioner tersebut dengan jawaban yang sudah disediakan. Tabel 4.2 menjelaskan pertanyaan dalam kuesioner, kemungkinan jawab responden, dan kodifikasi pertanyaan menjadi kolom gejala dalam dataset.

Tabel 4.2 Pertanyaan Kuesioner, Kemungkinan Jawaban, dan Kodifikasi Pertanyaan dalam Dataset Penyakit Obesitas

Pertanyaan	Kemungkinan Jawaban	Kodifikasi Pertanyaan dalam dataset
<i>What is your gender?</i>	<ul style="list-style-type: none"> <li>• <i>Female</i></li> <li>• <i>Male</i></li> </ul>	<i>Gender</i>
<i>what is your age?</i>	<ul style="list-style-type: none"> <li>• <i>Numeric value</i></li> </ul>	<i>Age</i>
<i>What is your height?</i>	<ul style="list-style-type: none"> <li>• <i>Numeric value in meters</i></li> </ul>	<i>Height</i>
<i>What is your weight?</i>	<ul style="list-style-type: none"> <li>• <i>Numeric value in kilograms</i></li> </ul>	<i>Weight</i>
<i>Has a family member suffered or suffers from overweight?</i>	<ul style="list-style-type: none"> <li>• <i>Yes</i></li> <li>• <i>No</i></li> </ul>	<i>family_history_with_overweight</i>
<i>Do you eat high caloric food frequently?</i>	<ul style="list-style-type: none"> <li>• <i>Yes</i></li> <li>• <i>No</i></li> </ul>	<i>Frequent consumption of high caloric food (FAVC)</i>
<i>Do you usually eat vegetables in your meals?</i>	<ul style="list-style-type: none"> <li>• <i>Never</i></li> <li>• <i>Sometimes</i></li> <li>• <i>Always</i></li> </ul>	<i>Frequency of consumption of vegetables (FCVC)</i>
<i>How many main meals do you have daily?</i>	<ul style="list-style-type: none"> <li>• <i>Between 1 y 2</i></li> <li>• <i>Three</i></li> <li>• <i>More than three</i></li> </ul>	<i>Number of main meals (NCP)</i>
<i>Do you eat any food between meals?</i>	<ul style="list-style-type: none"> <li>• <i>No</i></li> <li>• <i>Sometimes</i></li> <li>• <i>Frequently</i></li> <li>• <i>Always</i></li> </ul>	<i>Consumption of food between meals (CAEC)</i>

Tabel 4.2 Pertanyaan Kuesioner, Kemungkinan Jawaban, dan Kodifikasi Pertanyaan dalam Dataset Penyakit Obesitas (lanjutan)

Pertanyaan	Kemungkinan Jawaban	Kodifikasi Pertanyaan dalam dataset
<i>Do you smoke?</i>	<ul style="list-style-type: none"> <li>• Yes</li> <li>• No</li> </ul>	<i>SMOKE</i>
<i>How much water do you drink daily?</i>	<ul style="list-style-type: none"> <li>• Less than a liter</li> <li>• Between 1 and 2 L</li> <li>• More than 2 L</li> </ul>	<i>Consumption of water daily (CH20)</i>
<i>Do you monitor the calories you eat daily?</i>	<ul style="list-style-type: none"> <li>• Yes</li> <li>• No</li> </ul>	<i>Calories consumption monitoring (SCC)</i>
<i>How often do you have physical activity?</i>	<ul style="list-style-type: none"> <li>• I do not have</li> <li>• 1 or 2 days</li> <li>• 2 or 4 days</li> <li>• 4 or 5 days</li> </ul>	<i>Physical activity frequency (FAF)</i>
<i>How much time do you use technological devices such as cell phone, videogames, television, computer and others?</i>	<ul style="list-style-type: none"> <li>• 0–2 hours</li> <li>• 3–5 hours</li> <li>• More than 5 hours</li> </ul>	<i>Time using technology devices (TUE)</i>
<i>How often do you drink alcohol?</i>	<ul style="list-style-type: none"> <li>• I do not drink</li> <li>• Sometimes</li> <li>• Frequently</li> <li>• Always</li> </ul>	<i>Consumption of alcohol (CALC).</i>
<i>Which transportation do you usually use?</i>	<ul style="list-style-type: none"> <li>• Automobile</li> <li>• Motorbike</li> <li>• Bike</li> <li>• Public Transportation</li> <li>• Walking</li> </ul>	<i>Transportation used (MTRANS)</i>

Sumber : Palenchor dan Manotas [38]

Dataset penyakit tiroid memiliki 7 kolom gejala yang digunakan dalam memprediksi penyakit. Beberapa indikator yang digunakan merupakan pengukuran dari hormon yang berkaitan dengan kelenjar tiroid, seperti total hormon tiroksin (T4), triiodotironin (T3), *thyroid-binding globulin* (TBG), dan *thyroid stimulating hormone* (TSH). Selain itu terdapat pengukuran *free T4 index* (FTI) yang berkaitan dengan pengecekan hormon yang menstimulasi tiroid (TSH, *thyroid-stimulating*

*hormone*). Data yang didapatkan bersumber dari STMW, SVHC, SVI, SVHD, dan organisasi – organisasi kesehatan lain di Australia. Data diambil dari tahun 1984 hingga 1987.

Dataset penyakit kanker payudara memiliki 30 kolom gejala yang mendeskripsikan karakteristik dari sebuah sel otot pada payudara. Kolom – kolom gejala ini merupakan total pengukuran 10 tipe ukuran dari sel dengan jenis *mean*, *standard error* (SE), dan *worst*. Kolom – kolom gejala tersebut adalah *radius*, *texture*, *perimeter*, *area*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, dan *fractal dimension*.

Pada dataset penyakit umum, terdapat 132 kolom gejala yang diteliti, dengan jumlah penyakit yang didiagnosis sebanyak 41 tipe. Gejala – gejala umum yang diukur seperti gatal – gatal, sakit perut, kelelahan, berat badan menurun, dan lain – lain.

Dataset penyakit jantung memiliki 13 kolom gejala yang digunakan untuk memprediksi penyakit. Responden data terdiri dari laki – laki dan perempuan yang berumur 29 – 77 tahun. Beberapa kolom gejala yang digunakan seperti *trestbp*, *chol*, *fasting blood sugar*, dan lain – lain. Pada dataset ini, beberapa kolom gejala memiliki tipe kategorial sehingga nilai data berada dalam suatu rentang angka. Seperti pada kolom *cp* (*chest pain*) terdapat 4 jenis nilai data, yaitu *typical* (bernilai 1), *atypical* (bernilai 2), *non-anginal* (bernilai 3), dan *asymptomatic* (bernilai 4). Kolom *restecg* memiliki 3 jenis nilai data, yaitu normal (bernilai 0), gelombang ST-T abnormal (bernilai 1), dan menunjukkan kondisi hipertropi ventrikular kiri menurut kriteria Estes (bernilai 2). Kolom *slope* memiliki 3 jenis nilai data, yaitu *upsloping/naik* (bernilai 1), *flat* (bernilai 2), dan *downsloping/turun* (bernilai 3). Kolom *thal* memiliki 3 jenis nilai data, yaitu normal (bernilai 3), *fixed defect* (bernilai 6), dan *reversible defect* (bernilai 7).

Data penyakit COVID-19 memiliki 7 indikator yang digunakan untuk memprediksi penyakit. Responden data terdiri dari laki – laki, perempuan, dan transgender. Responden data dibagi berdasarkan kategori umur, yaitu umur 0 – 9 tahun, 10 – 19 tahun, 20 – 24 tahun, 25 – 59 tahun, dan diatas 60 tahun. Dalam data penyakit COVID-19 terdapat indikator *severity* atau tingkat keparahan dari pasien,

yang terdiri dari kecil, sedang, dan tinggi. Selain itu, terdapat pula indikator negara yang mengindikasikan apakah responden pernah mengunjungi suatu negara. Negara – negara tersebut antara lain Cina, Perancis, Jerman, Iran, Italia, Korea Selatan, Spanyol, Uni Emirat Arab, negara – negara lain di Eropa (dilabelkan dengan *Other-EUR*), dan negara – negara di luar Eropa (dilabelkan dengan *Other*).

#### 4.1.2 Sumber Data

Dataset yang digunakan dalam penelitian berasal dari *website* yang menyediakan dataset online secara gratis seperti Kaggle dan UCI Machine Learning. Kedua website ini menyediakan dataset yang dapat didownload dan digunakan untuk meneliti topik – topik tertentu, seperti kesehatan, manajemen, olahraga, kelautan, dan lain – lain. *User* yang berkontribusi dalam mengunggah dataset dalam website ini terdiri dari pegawai publik, instansi pendidikan, peneliti mandiri, dan lain – lain. Tabel 4.3 mendeskripsikan pemilik dari data yang digunakan dan penyakit yang diteliti.

Tabel 4.3 Tabel Sumber Data

Jenis Penyakit	Sumber/Donor
Obesitas	Fabio Mendoza Palechor Email: fmendoza1 '@' cuc.edu.co Alexis de la Hoz Manotas Email: akdelahoz '@' gmail.com
Kanker Payudara	Dr. William H. Wolberg, General Surgery Dept., University of Wisconsin W. Nick Street, Computer Sciences Dept., University of Wisconsin Olvi L. Mangasarian, Computer Sciences Dept., University of Wisconsin Donor : Nick Street
Jantung	Andras Janosi, M.D. Hungarian Institute of Cardiology, Budapest William Steinbrunn, M.D. University Hospital, Zurich, Switzerland Matthias Pfisterer, M.D. University Hospital, Basel, Switzerland Robert Detrano, M.D., Ph.D. V.A. Medical Center, Long Beach and Cleveland Clinic Foundation Donor: David W. Aha (aha '@' ics.uci.edu)

Tabel 4.3 Tabel Sumber Data (lanjutan)

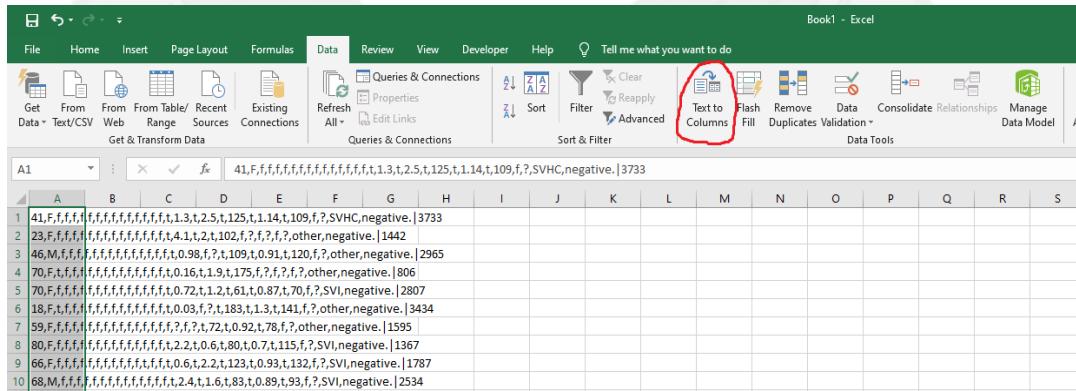
Jenis Penyakit	Sumber/Donor
Umum	KAUSHIL268
Tiroid	Garavan Institute and J. Ross Quinlan New South Wales Institute, Sydney, Australia
COVID-19	Bilal H. Hungund K.

## 4.2 Tahap Data Cleaning

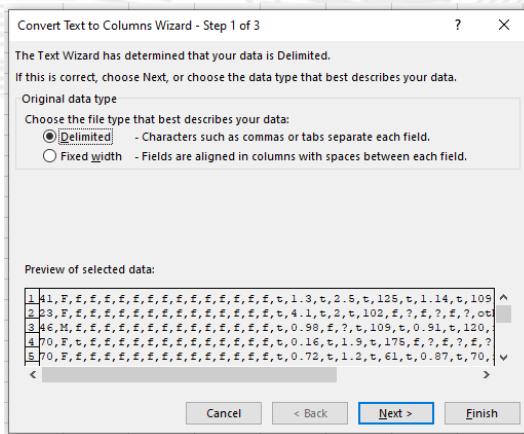
Tahap pembersihan data (data cleaning) dilakukan untuk membersihkan dataset yang sudah dikumpulkan dari nilai – nilai kosong, outlier, dan nilai – nilai yang tidak konsisten. Pada tahapan ini, dataset dipersiapkan sebelum memasuki tahapan *preprocessing*. Pembersihan dataset dilakukan dengan menghapus data yang memiliki nilai tanda tanya (?), nilai kosong atau *blank*. Beberapa dataset juga memiliki nilai data kosong yang direpresentasikan dengan angka -9 sehingga nilai juga dihapus.

Untuk kemudahan pengolahan data, semua data yang dikumpulkan diubah ke dalam format file excel (.xlxs). Proses mengubah bentuk data dari .data menjadi file excel dilakukan dengan cara mengopi teks dari .data yang dibuka menggunakan program Notepad dan menyalinnya pada file excel. Saat menyalin, data akan berada pada satu kolom. Untuk membaginya menjadi kolom – kolom, yang perlu dilakukan adalah menyeleksi seluruh data, lalu klik menu *Data*, klik *Text to Columns* pada bagian *Data Tools*, centang pilihan “*Delimited*” untuk memisahkan data berdasarkan tanda baca, lalu klik *Next*. Di tampilan selanjutnya, centang “*comma*” untuk memisahkan data berdasarkan koma, lalu klik *Finish*. Gambar 4.3 hingga gambar 4.6 menjelaskan tahapan dalam mengubah bentuk data secara lebih jelas. Untuk proses pengubahan bentuk data dari .csv tidak sulit karena hanya memerlukan penyimpanan file ulang dengan format file excel.

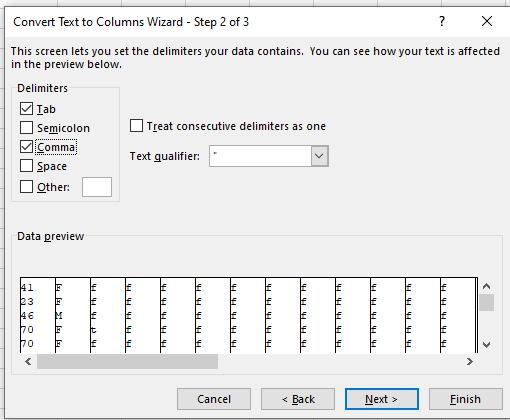
Gambar 4.3 Hasil Penyalinan File .data di File Excel



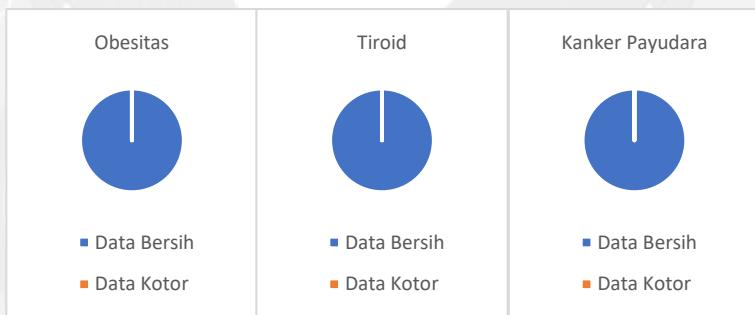
Gambar 4.4 Menu Text to Columns Untuk Memisahkan Data



Gambar 4.5 Tampilan Pemisah Data Berdasarkan Tanda Baca



Gambar 4.6 Tampilan Pemisah Data Menggunakan Tanda Baca Koma

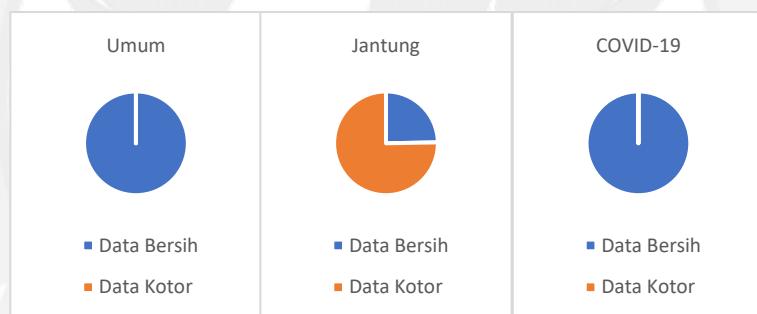


Gambar 4.7 Perbandingan Hasil Pembersihan Data Penyakit Obesitas, Tiroid, dan Kanker Payudara

Setelah seluruh dataset diubah ke dalam bentuk file excel, dilakukanlah proses *data cleaning* pada dataset. Gambar 4.7 menunjukkan perbandingan hasil perbersihan dataset penyakit obesitas, tiroid, dan kanker payudara. Pada proses pembersihan dataset penyakit obesitas, proses ini menghasilkan data 2.111 data bersih dan 0 data kotor dari total 2.111 data. Proses pembersihan dataset penyakit tiroid menghasilkan data 12.944 data bersih dan 0 data kotor dari 12.944 data. Pada proses pembersihan data penyakit tiroid ini, terdapat nilai kosong berupa tanda tanya (?). Pada penelitian ini, nilai – nilai kosong pada dataset penyakit tiroid dapat diganti dengan nilai rata – rata dari keseluruhan dataset, seperti pada kolom *age* dan *sex*. Selain itu juga terdapat nilai kosong pada kolom *TSH*, *T3*, *TT4*, *T4U*, *FTI*, dan *TBG*. Kolom – kolom ini berhubungan dengan kolom *TSH measured*, *T3 measured*, *TT4 measured*, *T4U measured*, *FTI measured*, dan *TBG measured*.

Apabila kolom dengan *measured* memiliki nilai f, maka nilai dari kolom pengukuran yang bersangkutan akan bernilai tanda tanya. Maka daripada itu, penelitian ini akan mengganti tanda tanya pada kolom TSH, T3, TT4, T4U, FTI, dan TBG menjadi 0. Proses pembersihan data penyakit kanker payudara menghasilkan 569 data bersih dan 0 data kotor dari 569 data.

Gambar 4.8 menunjukkan perbandingan hasil pembersihan data penyakit umum, penyakit jantung, dan penyakit COVID-19. Proses pembersihan data penyakit umum menghasilkan 4.962 data bersih dan 0 data kotor dari 4.962 data. Pembersihan data penyakit jantung menghasilkan 300 data bersih dan 914 data kotor dari 1.214 data. Pembersihan data penyakit COVID-19 menghasilkan 316.800 data bersih dan 0 data kotor dari 316.800 data. Pada proses pembersihan dataset penyakit COVID-19 juga dilakukan proses menghapus kolom data yang hanya memiliki nilai 0.



Gambar 4.8 Perbandingan Hasil Pembersihan Data Penyakit Umum, Jantung, dan COVID-19

### 4.3 Tahap Data Integration

Tahap pengintegrasian data (*data integration*) merupakan tahapan menggabungkan data dari berbagai sumber yang berbeda menjadi satu format data yang dipakai saat melakukan data mining. Karena tipe dataset dari masing – masing penyakit yang dikumpulkan hanya memiliki satu format saja, yaitu antara file CSV dan file .data, maka tahap *data integration* tidak perlu dilakukan.

## 4.4 Tahap Data Selection

Tahap pemilihan data (*data selection*) memilih data yang relevan untuk dilakukannya analisis. Subbab ini menentukan setiap kolom gejala dan label yang akan dipakai pada penelitian ini. Setiap kolom gejala dan label diagnosis akan dijelaskan per dataset penyakit.

### 4.4.1 Penjelasan Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Obesitas

Pada dataset penyakit obesitas, kolom gejala yang digunakan untuk mendiagnosis adalah *gender*, *age*, *height*, *weight*, *family\_history\_with\_overweight*, FAVC, FCVC, NCP, CAEC, SMOKE, CH20, SCC, FAF, TUE, CALC, dan MTRANS.

Label diagnosis yang digunakan dalam dataset obesitas berjumlah 7 label, yaitu *Normal\_Weight*, *Insufficient\_Weight*, *Overweight\_Level\_I*, *Overweight\_Level\_II*, *Obesity\_Type\_I*, *Obesity\_Type\_II*, dan *Obesity\_Type\_III*. Tabel 4.4 menunjukkan pemilihan kolom gejala dan label diagnosis terpilih dari dataset penyakit obesitas.

Tabel 4.4 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Obesitas

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>gender</i>
Kolom Gejala	<i>age</i>
Kolom Gejala	<i>height</i>
Kolom Gejala	<i>weight</i>
Kolom Gejala	<i>family_history_with_overweight</i>
Kolom Gejala	FAVC
Kolom Gejala	FCVC
Kolom Gejala	NCP
Kolom Gejala	CAEC
Kolom Gejala	SMOKE

Tabel 4.4 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Obesitas (lanjutan)

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	CH20
Kolom Gejala	SCC
Kolom Gejala	FAF
Kolom Gejala	TUE
Kolom Gejala	CALC
Kolom Gejala	MTRANS
Label Diagnosis	<i>Normal_Weight</i>
Label Diagnosis	<i>Insufficient_Weight</i>
Label Diagnosis	<i>Overweight_Level_I</i>
Label Diagnosis	<i>Overweight_Level_I</i>
Label Diagnosis	<i>Obesity_Level_I</i>
Label Diagnosis	<i>Obesity_Level_II</i>
Label Diagnosis	<i>Obesity_Level_III</i>

#### 4.4.2 Penjelasan Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Tiroid

Pada dataset penyakit tiroid, terdapat 27 kolom gejala yang akan digunakan untuk mendiagnosis penyakit tiroid. Kolom – kolom tersebut adalah *age*, *sex*, *on thyroxine*, *query on thyroxine*, *on antithyroid medication*, *sick*, *pregnant*, *thyroid surgery*, *I131 treatment*, *query hypothyroid*, *query hyperthyroid*, *lithium*, *goitre*, *tumor*, *hypopituitary*, *psych*, *TSH measured*, *TSH*, *T3 measured*, *TT4 measured*, *TT4, T4U measured*, *T4U*, *FTI measured*, *FTI*, dan *TBG measured*.

Label diagnosis dalam dataset penyakit tiroid terbagi menjadi 3 tipe, yaitu label diagnosis dengan tipe hipotiroid, label diagnosis dengan tipe hipertiroid dan berupa kode. Label diagnosis yang masih berupa kode akan dipilih dengan catatan label kode setelah ditafsirkan masih termasuk dalam kategori hipotiroid atau

hipertiroid. Tabel 4.5 menunjukkan kolom gejala dan label diagnosis terpilih dari dataset penyakit tiroid.

Tabel 4.5 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Tiroid

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>age</i>
Kolom Gejala	<i>sex</i>
Kolom Gejala	<i>on thyroxine</i>
Kolom Gejala	<i>query on thyroxine</i>
Kolom Gejala	<i>on antithyroid medication</i>
Kolom Gejala	<i>sick</i>
Kolom Gejala	<i>pregnant</i>
Kolom Gejala	<i>thyroid surgery</i>
Kolom Gejala	<i>I131 treatment</i>
Kolom Gejala	<i>query hypothyroid.</i>
Kolom Gejala	<i>query hyperthyroid</i>
Kolom Gejala	<i>lithium</i>
Kolom Gejala	<i>goitre</i>
Kolom Gejala	<i>tumor</i>
Kolom Gejala	<i>hypopituitary</i>
Kolom Gejala	<i>psych</i>
Kolom Gejala	<i>TSH measured</i>
Kolom Gejala	<i>TSH</i>
Kolom Gejala	<i>T3 measured</i>
Kolom Gejala	<i>TT4 measured</i>
Kolom Gejala	<i>TT4</i>
Kolom Gejala	<i>T4U measured</i>

Tabel 4.5 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Tiroid (lanjutan)

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>T4U</i>
Kolom Gejala	<i>FTI measured</i>
Kolom Gejala	<i>FTI</i>
Kolom Gejala	<i>TBG measured</i>
Label Diagnosis	<i>negative</i>
Label Diagnosis	<i>compensated hypothyroid</i>
Label Diagnosis	<i>primary hypothyroid</i>
Label Diagnosis	<i>hyperthyroid</i>
Label Diagnosis	<i>goitre</i>
Label Diagnosis	<i>T3 toxic (Kode B)</i>
Label Diagnosis	<i>toxic goitre (Kode C)</i>
Label Diagnosis	<i>secondary toxic (Kode D)</i>
Label Diagnosis	<i>concurrent non-thyroidal illness (Kode K)</i>

#### 4.4.3 Penjelasan Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Kanker Payudara

Pada dataset penyakit kanker payudara dipilih 30 kolom gejala yang digunakan untuk mendiagnosis penyakit kanker payudara. Adapun 30 gejala tersebut adalah *radius\_mean*, *texture\_mean*, *perimeter\_mean*, *area\_mean*, *smoothness\_mean*, *compactness\_mean*, *concavity\_mean*, *concave points\_mean*, *symmetry\_mean*, *fractal\_dimension\_mean*, *radius\_se*, *texture\_se*, *perimeter\_se*, *area\_se*, *smoothness\_se*, *compactness\_se*, *concavity\_se*, *concave points\_se*, *symmetry\_se*, *fractal\_dimension\_se*, *radius\_worst*, *texture\_worst*, *perimeter\_worst*, *area\_worst*, *smoothness\_worst*, *compactness\_worst*, *concavity\_worst*, *concave points\_worst*, *symmetry\_worst*, dan *fractal\_dimension\_worst*.

Dalam dataset penyakit kanker payudara terdapat 2 label diagnosis yang dihasilkan yaitu B atau *Benign* (Tumor tidak membahayakan tubuh) dan M atau *Malignant* (Tumor membahayakan tubuh). Tabel 4.6 menunjukkan kolom gejala dan label diagnosis terpilih dari dataset penyakit kanker payudara.

Tabel 4.6 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Kanker Payudara

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>radius_mean</i>
Kolom Gejala	<i>texture_mean</i>
Kolom Gejala	<i>perimeter_mean</i>
Kolom Gejala	<i>area_mean</i>
Kolom Gejala	<i>smoothness_mean</i>
Kolom Gejala	<i>compactness_mean</i>
Kolom Gejala	<i>concavity_mean</i>
Kolom Gejala	<i>concave points_mean</i>
Kolom Gejala	<i>symmetry_mean</i>
Kolom Gejala	<i>fractal_dimension_mean</i>
Kolom Gejala	<i>radius_se</i>
Kolom Gejala	<i>texture_se</i>
Kolom Gejala	<i>perimeter_se</i>
Kolom Gejala	<i>area_se</i>
Kolom Gejala	<i>smoothness_se</i>
Kolom Gejala	<i>compactness_se</i>
Kolom Gejala	<i>concavity_se</i>
Kolom Gejala	<i>concave points_se</i>
Kolom Gejala	<i>symmetry_se</i>
Kolom Gejala	<i>fractal_dimension_se</i>

Tabel 4.6 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Kanker Payudara (lanjutan)

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>radius_worst</i>
Kolom Gejala	<i>texture_worst</i>
Kolom Gejala	<i>perimeter_worst</i>
Kolom Gejala	<i>area_worst</i>
Kolom Gejala	<i>smoothness_worst</i>
Kolom Gejala	<i>compactness_worst</i>
Kolom Gejala	<i>concavity_worst</i>
Kolom Gejala	<i>concave points_worst</i>
Kolom Gejala	<i>symmetry_worst</i>
Kolom Gejala	<i>fractal_dimension_worst</i>
Label Diagnosis	<i>B</i>
Label Diagnosis	<i>M</i>

#### 4.4.4 Penjelasan Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum

Pada dataset penyakit umum, terdapat 116 kolom gejala yang akan digunakan untuk mendiagnosis penyakit umum. Kolom – kolom tersebut adalah *itching*, *skin\_rash*, *continuous\_sneezing*, *shivering*, *chills*, *joint\_pain*, *acidity*, *ulcers\_on\_tongue*, *muscle\_wasting*, *vomiting*, *burning\_micturition*, *spotting\_urination*, *fatigue*, *weight\_gain*, *cold\_hands\_and\_feets*, *mood\_swings*, *weight\_loss*, *restlessness*, *patches\_in\_throat*, *irregular\_sugar\_level*, *cough*, *high\_fever*, *sunken\_eyes*, *breathlessness*, *sweating*, *dehydration*, *indigestion*, *headache*, *yellowish\_skin*, *dark\_urine*, *nausea*, *loss\_of\_appetite*, *pain\_behind\_the\_eyes*, *back\_pain*, *constipation*, *abdominal\_pain*, *diarrhoea*, *mild\_fever*, *yellow\_urine*, *yellowing\_of\_eyes*, *acute\_liver\_failure*, *swelling\_of\_stomach*, *swelled\_lymph\_nodes*, *malaise*,

*blurred\_and\_distorted\_vision, phlegm, throat\_irritation, redness\_of\_eyes, sinus\_pressure, runny\_nose, congestion, chest\_pain, weakness\_in\_limbs, fast\_heart\_rate, pain\_during\_bowel\_movements, pain\_in\_anal\_region, bloody\_stool, irritation\_in\_anus, neck\_pain, dizziness, cramps, bruising, obesity, swollen\_legs, swollen\_blood\_vessels, puffy\_face\_and\_eyes, brittle\_nails, swollen\_extremeties, excessive\_hunger, extra\_marital\_contacts, drying\_and\_tingling\_lips, slurred\_speech, knee\_pain, hip\_joint\_pain, muscle\_weakness, stiff\_neck, swelling\_joints, movement\_stiffness, spinning\_movements, loss\_of\_balance, weakness\_of\_one\_body\_side, loss\_of\_smell, bladder\_discomfort, foul\_smell\_of\_urine, continuous\_feel\_of\_urine, internal\_itching, toxic\_look\_(typhos), depression, irritability, muscle\_pain, red\_spots\_over\_body, abnormal\_menstruation, dischromic\_patches, watering\_from\_eyes, increased\_appetite, polyuria, family\_history, mucoid\_sputum, rusty\_sputum, lack\_of\_concentration, visual\_disturbances, receiving\_blood\_transfusion, history\_of\_alcohol\_consumption, fluid\_overload, blood\_in\_sputum, prominent\_veins\_on\_calf, palpitations, painful\_walking, pus\_filled\_pimples, blackheads, skin\_peeling, small\_dents\_in\_nails, inflammatory\_nails, blister, red\_sore\_around\_nose, dan yellow\_crust\_ooze.* Beberapa kolom gejala memiliki pengertian yang kurang lebih sama. Pada tahapan ini, kolom gejala yang memiliki arti yang kurang lebih sama dilihat perbandingan nilai 0 dan 1 antar kolom gejala tersebut. Kolom yang memiliki nilai 1 lebih banyak dipakai dalam penelitian ini.

Pada dataset penyakit umum, terdapat 40 label diagnosis yang dihasilkan. Label - label tersebut antara lain (*vertigo*) *Paroxysmal Positional Vertigo, Fungal infection, Allergy, GERD, Chronic Cholestasis, Acne, AIDS, Alcoholic hepatitis, Arthritis, Bronchial Asthma, Cervical Spondylosis, Chicken Pox, Common Cold, Dengue, Diabetes, Dimorphic Hemorrhoids (piles), Drug Reaction, Gastroenteritis, Heart Attack, Hepatitis A, Hepatitis B, Hepatitis C, Hepatitis D, Hepatitis E, Hypertension, Hyperthyroidism, Hypothyroidism, Impetigo, Jaundice, Malaria, Migraine, Osteoarthritis, Paralysis (Brain Hemorrhage), Peptic Ulcer Disease, Pneumonia, Psoriasis, Tuberculosis, Typhoid, Urinary Tract Infection,*

dan *Varicose Veins*. Tabel 4.7 menunjukkan kolom gejala dan label diagnosis terpilih dari dataset penyakit umum.

Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>itching</i>
Kolom Gejala	<i>skin_rash</i>
Kolom Gejala	<i>continuous_sneezing</i>
Kolom Gejala	<i>shivering</i>
Kolom Gejala	<i>chills</i>
Kolom Gejala	<i>joint_pain</i>
Kolom Gejala	<i>acidity</i>
Kolom Gejala	<i>ulcers_on_tongue</i>
Kolom Gejala	<i>muscle_wasting</i>
Kolom Gejala	<i>vomiting</i>
Kolom Gejala	<i>burning_micturition</i>
Kolom Gejala	<i>spotting_urination</i>
Kolom Gejala	<i>fatigue</i>
Kolom Gejala	<i>weight_gain</i>
Kolom Gejala	<i>cold_hands_and_feets</i>
Kolom Gejala	<i>mood_swings</i>
Kolom Gejala	<i>weight_loss</i>
Kolom Gejala	<i>restlessness</i>
Kolom Gejala	<i>patches_in_throat</i>
Kolom Gejala	<i>irregular_sugar_level</i>
Kolom Gejala	<i>cough</i>
Kolom Gejala	<i>high_fever</i>

Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>sunken_eyes</i>
Kolom Gejala	<i>breathlessness</i>
Kolom Gejala	<i>sweating</i>
Kolom Gejala	<i>dehydration</i>
Kolom Gejala	<i>indigestion</i>
Kolom Gejala	<i>headache</i>
Kolom Gejala	<i>yellowish_skin</i>
Kolom Gejala	<i>dark_urine</i>
Kolom Gejala	<i>nausea</i>
Kolom Gejala	<i>loss_of_appetite</i>
Kolom Gejala	<i>pain_behind_the_eyes</i>
Kolom Gejala	<i>back_pain</i>
Kolom Gejala	<i>constipation</i>
Kolom Gejala	<i>abdominal_pain</i>
Kolom Gejala	<i>diarrhoea</i>
Kolom Gejala	<i>mild_fever</i>
Kolom Gejala	<i>yellow_urine</i>
Kolom Gejala	<i>yellowing_of_eyes</i>
Kolom Gejala	<i>acute_liver_failure</i>
Kolom Gejala	<i>swelling_of_stomach</i>
Kolom Gejala	<i>swelled_lymph_nodes</i>
Kolom Gejala	<i>malaise</i>
Kolom Gejala	<i>blurred_and_distorted_vision</i>
Kolom Gejala	<i>phlegm</i>
Kolom Gejala	<i>throat_irritation</i>

Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>redness_of_eyes</i>
Kolom Gejala	<i>sinus_pressure</i>
Kolom Gejala	<i>runny_nose</i>
Kolom Gejala	<i>congestion</i>
Kolom Gejala	<i>chest_pain</i>
Kolom Gejala	<i>weakness_in_limbs</i>
Kolom Gejala	<i>fast_heart_rate</i>
Kolom Gejala	<i>pain_during_bowel_movements</i>
Kolom Gejala	<i>pain_in_anal_region</i>
Kolom Gejala	<i>bloody_stool</i>
Kolom Gejala	<i>irritation_in_anus</i>
Kolom Gejala	<i>neck_pain</i>
Kolom Gejala	<i>dizziness</i>
Kolom Gejala	<i>cramps</i>
Kolom Gejala	<i>bruising</i>
Kolom Gejala	<i>obesity</i>
Kolom Gejala	<i>swollen_legs</i>
Kolom Gejala	<i>swollen_blood_vessels</i>
Kolom Gejala	<i>puffy_face_and_eyes</i>
Kolom Gejala	<i>brittle_nails</i>
Kolom Gejala	<i>swollen_extremeties</i>
Kolom Gejala	<i>excessive_hunger</i>
Kolom Gejala	<i>extra_marital_contacts</i>
Kolom Gejala	<i>drying_and_tingling_lips</i>
Kolom Gejala	<i>slurred_speech</i>

Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>knee_pain</i>
Kolom Gejala	<i>hip_joint_pain</i>
Kolom Gejala	<i>muscle_weakness</i>
Kolom Gejala	<i>stiff_neck</i>
Kolom Gejala	<i>swelling_joints</i>
Kolom Gejala	<i>movement_stiffness</i>
Kolom Gejala	<i>spinning_movements</i>
Kolom Gejala	<i>loss_of_balance</i>
Kolom Gejala	<i>weakness_of_one_body_side</i>
Kolom Gejala	<i>loss_of_smell</i>
Kolom Gejala	<i>bladder_discomfort</i>
Kolom Gejala	<i>foul_smell_of_urine</i>
Kolom Gejala	<i>continuous_feel_of_urine</i>
Kolom Gejala	<i>internal_itching</i>
Kolom Gejala	<i>toxic_look_(typhos)</i>
Kolom Gejala	<i>depression</i>
Kolom Gejala	<i>irritability</i>
Kolom Gejala	<i>muscle_pain</i>
Kolom Gejala	<i>red_spots_over_body</i>
Kolom Gejala	<i>abnormal_menstruation</i>
Kolom Gejala	<i>dischromic_patches</i>
Kolom Gejala	<i>watering_from_eyes</i>
Kolom Gejala	<i>increased_appetite</i>
Kolom Gejala	<i>polyuria</i>
Kolom Gejala	<i>family_history</i>

Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>mucoid_sputum</i>
Kolom Gejala	<i>rusty_sputum</i>
Kolom Gejala	<i>lack_of_concentration</i>
Kolom Gejala	<i>visual_disturbances</i>
Kolom Gejala	<i>receiving_blood_transfusion</i>
Kolom Gejala	<i>history_of_alcohol_consumption</i>
Kolom Gejala	<i>fluid_overload</i>
Kolom Gejala	<i>blood_in_sputum</i>
Kolom Gejala	<i>prominent_veins_on_calf</i>
Kolom Gejala	<i>palpitations</i>
Kolom Gejala	<i>painful_walking</i>
Kolom Gejala	<i>pus_filled_pimples</i>
Kolom Gejala	<i>blackheads</i>
Kolom Gejala	<i>skin_peeling</i>
Kolom Gejala	<i>small_dents_in_nails</i>
Kolom Gejala	<i>inflammatory_nails</i>
Kolom Gejala	<i>blister</i>
Kolom Gejala	<i>red_sore_around_nose</i>
Kolom Gejala	<i>yellow_crust_ooze</i>
Label Diagnosis	<i>(vertigo) Paroxysmal Positional Vertigo</i>
Label Diagnosis	<i>Fungal_infection</i>
Label Diagnosis	<i>Allergy</i>
Label Diagnosis	<i>GERD</i>
Label Diagnosis	<i>Chronic Cholestasis</i>
Label Diagnosis	<i>Acne</i>

Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)

Jenis Kolom	Nama Kolom/Label
Label Diagnosis	<i>AIDS</i>
Label Diagnosis	<i>Alcoholic hepatitis</i>
Label Diagnosis	<i>Arthritis</i>
Label Diagnosis	<i>Bronchial Asthma</i>
Label Diagnosis	<i>Cervical Spondylosis</i>
Label Diagnosis	<i>Chicken Pox</i>
Label Diagnosis	<i>Common Cold</i>
Label Diagnosis	<i>Dengue</i>
Label Diagnosis	<i>Diabetes</i>
Label Diagnosis	<i>Dimorphic Hemmorhoids (piles)</i>
Label Diagnosis	<i>Drug Reaction</i>
Label Diagnosis	<i>Gastroenteritis</i>
Label Diagnosis	<i>Heart Attack</i>
Label Diagnosis	<i>Hepatitis A</i>
Label Diagnosis	<i>Hepatitis B</i>
Label Diagnosis	<i>Hepatitis C</i>
Label Diagnosis	<i>Hepatitis D</i>
Label Diagnosis	<i>Hepatitis E</i>
Label Diagnosis	<i>Hypertension</i>
Label Diagnosis	<i>Hypoglycemia</i>
Label Diagnosis	<i>Hyperthyroidism</i>
Label Diagnosis	<i>Hypothyroidism</i>
Label Diagnosis	<i>Impetigo</i>
Label Diagnosis	<i>Jaundice</i>
Label Diagnosis	<i>Malaria</i>

Tabel 4.7 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Umum (lanjutan)

Jenis Kolom	Nama Kolom/Label
Label Diagnosis	<i>Migraine</i>
Label Diagnosis	<i>Osteoarthristis</i>
Label Diagnosis	<i>Paralysis (Brain Hemorrhage)</i>
Label Diagnosis	<i>Peptic Ulcer Disease</i>
Label Diagnosis	<i>Pneumonia</i>
Label Diagnosis	<i>Psoriasis</i>
Label Diagnosis	<i>Tuberculosis</i>
Label Diagnosis	<i>Typhoid</i>
Label Diagnosis	<i>Urinary Tract Infection</i>
Label Diagnosis	<i>Varicose Veins</i>

#### 4.4.5 Penjelasan Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Jantung

Pada dataset penyakit jantung, terdapat 13 kolom gejala yang akan digunakan untuk mendiagnosis penyakit jantung. Kolom gejala tersebut adalah *age*, *sex*, *cp*, *trestbp*, *chol*, *fbs*, *restecg*, *thalach*, *exang*, *oldpeak*, *slope*, *ca*, dan *thal*.

Pada dataset penyakit jantung terdapat label hasil yang ditandai dengan angka 0 (pasien tidak mengalami penyakit jantung) hingga angka 4 yang menyatakan tingkat terparah dari penyakit jantung yang dideteksi. Tabel 4.8 menunjukkan kolom gejala dan label diagnosis terpilih dari dataset penyakit jantung.

Tabel 4.8 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Jantung

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>age</i>
Kolom Gejala	<i>sex</i>
Kolom Gejala	<i>cp</i>

Tabel 4.8 Tabel Pemilihan Kolom Gejala dan Label Diagnosis Dataset Penyakit Jantung (lanjutan)

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>trestbp</i>
Kolom Gejala	<i>chol</i>
Kolom Gejala	<i>fbs</i>
Kolom Gejala	<i>restecg</i>
Kolom Gejala	<i>thalach</i>
Kolom Gejala	<i>exang</i>
Kolom Gejala	<i>oldpeak</i>
Kolom Gejala	<i>slope</i>
Kolom Gejala	<i>ca</i>
Kolom Gejala	<i>thal</i>
Label Diagnosis	<i>0</i>
Label Diagnosis	<i>1</i>
Label Diagnosis	<i>2</i>
Label Diagnosis	<i>3</i>
Label Diagnosis	<i>4</i>

#### 4.4.6 Penjelasan Pemilihan Kolom Gejala Dataset Penyakit COVID-19

Pada dataset penyakit COVID-19, terdapat 21 kolom gejala yang akan digunakan untuk mendiagnosis penyakit COVID-19. Kolom - kolom tersebut adalah *Fever*, *Tiredness*, *Dry-Cough*, *Difficulty-In-Breathing*, *Sore-Throat*, *Pains*, *Nasal-Congestion*, *Runny-Nose*, *Diarrhea*, *Age\_0-9*, *Age\_10-19*, *Age\_20-24*, *Age\_25-59*, *Age\_60+*, *Gender\_Female*, *Gender-Male*, *Gender\_Transgender*, *Contact\_Yes*, *Contact\_No*, dan *Country*. Dataset penyakit COVID-19 tidak memiliki kolom label diagnosis. Tabel 4.9 menunjukkan kolom gejala terpilih dari dataset penyakit COVID-19.

Tabel 4.9 Tabel Pemilihan Kolom Gejala Dataset Penyakit COVID-19

Jenis Kolom	Nama Kolom/Label
Kolom Gejala	<i>Fever</i>
Kolom Gejala	<i>Tiredness</i>
Kolom Gejala	<i>Dry-Cough</i>
Kolom Gejala	<i>Difficulty-In-Breathing</i>
Kolom Gejala	<i>Sore-Throat</i>
Kolom Gejala	<i>Pains</i>
Kolom Gejala	<i>Nasal-Congestion</i>
Kolom Gejala	<i>Runny-Nose</i>
Kolom Gejala	<i>Diarrhea</i>
Kolom Gejala	<i>Age_0-9</i>
Kolom Gejala	<i>Age_10-19</i>
Kolom Gejala	<i>Age_20-24</i>
Kolom Gejala	<i>Age_25-59</i>
Kolom Gejala	<i>Age_60+</i>
Kolom Gejala	<i>Gender_Female</i>
Kolom Gejala	<i>Gender-Male</i>
Kolom Gejala	<i>Gender_Transgender</i>
Kolom Gejala	<i>Contact_DontKnow</i>
Kolom Gejala	<i>Contact_Yes</i>
Kolom Gejala	<i>Contact_No</i>
Kolom Gejala	<i>Country</i>

#### 4.5 Tahap Data Transformation

Pada subbab ini dijelaskan tahapan pengubahan data (*data transformation*) yang dilakukan untuk dapat dilakukannya proses *data mining*. Tahapan *data transformation* ini bertujuan untuk mengubah data dapat terproses dengan baik saat

proses *data mining*. Beberapa tahapan yang akan dijelaskan adalah tahapan penghitungan total data yang akan dipakai dalam penelitian, tahapan pengubahan nilai data yang masih bersifat kategorial menjadi nilai 0 dan 1, dan tahapan pengubahan bahasa yang digunakan dari bahasa Inggris ke bahasa Indonesia.

#### 4.5.1 Tahap Pengubahan Nilai Data Kategorial dan Label Diagnosis

Pada beberapa kolom gejala dan kolom label diagnosis dalam dataset masih terdapat nilai data yang bersifat kategorial, sebagai contoh simbol F dan M pada data kolom jenis kelamin. Tahap ini akan menjelaskan proses pengubahan nilai data kategorial pada semua dataset penyakit.

AB	AC	AD	F	AA	AB	AC	AD
ce hypo	hyper	with code		id referral source	hypo	hyper	with code
negative. 3733	negative. 3733			other	compensated hypothyroid. 1593	negative. 1593	
negative. 2965	negative. 2965			SVI	negative. 1974	negative. 1974	
negative. 2807	negative. 2807			other	negative. 1493	negative. 1493	
negative. 3434	negative. 3434			SVI	negative. 1252	negative. 1252	
negative. 1367	negative. 1367			other	negative. 2008	negative. 2008	
negative. 1787	negative. 1787			other			F[840815067]
negative. 2534	negative. 2534			other			{-840815068}
negative. 1485	negative. 1485			other			{-840816001}
negative. 3448	negative. 3448			other			{-840816002}
negative. 1027	negative. 1027			other			{-840816003}
negative. 3331	negative. 3331			other			{-840816005}
negative. 2043	negative. 2043			SVI			{-840816007}
compensated hypothyroid. 3169	negative. 3169			SVI			AK[840816008]
negative. 2755	negative. 2755			SVI			{-840816009}
negative. 1010	negative. 1010			SVI			{-840816010}
negative. 1803	negative. 1803			SVI			{-840816011}
negative. 2297	negative. 2297			SVI			R[840816013]
negative. 3564	negative. 3564			other			I[840816014]
negative. 152	negative. 152			other			{-840816022}
negative. 936	negative. 936			other			{-840816025}
negative. 716	negative. 716			SVI			{-840816028}
negative. 1933	negative. 1933			SVI			{-840816029}
negative. 3445	hyperthyroid. 3445			other			M[840816047]
negative. 3724	negative. 3724			other			N[840816049]
negative. 1966	negative. 1966			other			{-840816052}
negative. 1091	negative. 1091			other			F[840816061]
negative. 583	negative. 583			other			{-840816063}
negative. 2137	negative. 2137			SVI			{-840816069}
negative. 1400	negative. 1400			SVI			{-840816070}
negative. 1815	negative. 1815			SVI			{-840816071}
negative. 3659	negative. 3659			SVI			{-840816072}
negative. 2797	negative. 2797			other			{-840816076}
negative. 3317	negative. 3317			other			{-840816077}
negative. 1566	negative. 1566			other			{-840816093}
primary hypothyroid. 2427	negative. 2427			other			M[840816095]
negative. 1433	negative. 1433			SVI			{-840817006}
...	...	...		other			{-840817008}

Gambar 4.9 Tampilan Awal Kolom Diagnosis Dataset Penyakit Tiroid

Pada dataset penyakit tiroid terdapat nilai f dan t pada data kolom Mengkonsumsi Thryroxine, Memiliki Resep Thyroxine, Mengkonsumsi Obat Antithyroid, Ingin Muntah, Hamil, Pernah Menjalani Operasi Terkait Thyroid, Menjalani Terapi I-131, (Query Hypothyroid), (Query Hyperthyroid), Mengkonsumsi Obat Lithium, Memiliki Benjolan di Leher, (Tumor), (Hypopituitary), (Psych), Pengukuran TSH, Pengukuran T3, Pengukuran TT4, Pengukuran T4U (T4U Measured), Pengukuran FTI, Pengukuran TBG. Selain itu, pada kolom jenis kelamin dan sumber referensi juga memiliki nilai kategorial yaitu F & M (untuk kolom jenis kelamin) dan SVHC, SVI, STMW, dan Other (untuk

kolom sumber referensi). Pada kolom diagnosis dari dataset penyakit tiroid, peneliti menggabungkan kolom – kolom diagnosis yang awalnya berjumlah 3 kolom menjadi 1 kolom untuk memudahkan dilakukannya *data mining*. Label diagnosis pada kolom diagnosis diubah menjadi “None”, “hipertiroid”, “keracunan T3”, “toxic goitre”, “secondary toxic”, “hipotiroid utama”, dan “hipotiroid terkompensasi” untuk label diagnosis awal. Gambar 4.9 dan 4.10 memperlihatkan perbandingan awal dari kolom diagnosis sebelum dan setelah dilakukannya transformasi.

	AA	AB	AC
referral source	diagnosis		
SVHC	None		
other	None		
SVI	None		
other	None		
SVI	None		
other	None		
other	None		
SVI	None		
SVHC	None		
other	compensated hypothyroid		
other	None		
SVI	None		
SVI	None		
SVHC	None		
other	None		
other	None		
SVI	None		
SVI	None		
other	hyperthyroid		
other	None		
SVI	None		
other	None		
other	None		
SVI	None		
SVI	None		
other	None		
other	None		
SVHC	None		
other	None		
other	None		
SVI	primary hypothyroid		
other	None		

Gambar 4.10 Tampilan Kolom Diagnosis Dataset Penyakit Tiroid Setelah Proses Transformasi

Pada dataset penyakit COVID-19 terdapat nilai kategorial pada kolom negara. Nilai – nilai tersebut antara lain adalah *China, Italy, Iran, Republic of Korean, France, Spain, Germany, UAE, Other-EUR*, dan *Other*. Dataset penyakit COVID-19 tidak memiliki kolom diagnosis, sehingga pada penelitian ini dibuatkan sebuah kolom diagnosis berdasarkan penelitian yang dibuat oleh Priyantono, Rachmawan, Budi, dan Kirana. Priyantono et al. membuat sebuah sistem prediksi virus korona menggunakan metode *forward chaining* [39]. Peneliti menggunakan *rule* yang dibuat untuk menentukan nilai diagnosis dari tiap data. Hasil diagnosis yang didapatkan ditulis dengan nilai “besar”, “sedang”, dan “kecil” berdasarkan

gejala dan kriteria dari *rule* yang diambil. Tabel 4.10 sampai 4.12 menunjukkan kode dan rule yang digunakan dalam menentukan hasil diagnosis.

Tabel 4.10 Tabel Kode Gejala Virus COVID-19

Kode	Keterangan
G1	Demam
G2	Pusing
G3	Bersin – Bersin
G4	Batuk
G5	Sakit Tenggorokan
G6	Kelelahan
G7	Nyeri Dada
G01	Tidak Demam
G02	Tidak Pusing
G03	Tidak Bersin – Bersin
G04	Tidak Batuk
G05	Tidak Sakit Tenggorokan
G06	Tidak Kelelahan
G07	Tidak Nyeri Dada

Sumber : Priyantono et al. [39]

Tabel 4.11 menunjukkan kode hasil keputusan diagnosis. Kode hasil keputusan dituliskan dengan huruf H diikuti dengan angka urutan hasil keputusan.

Tabel 4.11 Tabel Kode Hasil Keputusan Diagnosis

Kode	Keterangan
H1	Berpotensi besar positif Korona
H2	Berpotensi kecil positif Korona
H3	Beresiko

Sumber : Priyantono et al. [39]

Tabel 4.12 menunjukkan kombinasi kode gejala dan kode hasil keputusan. Kombinasi yang terbentuk terdiri dari kode gejala berdasarkan tabel 4.10 dan kode hasil keputusan yang dijelaskan pada tabel 4.11.

Tabel 4.12 Tabel Kombinasi Kode Gejala dan Hasil Kode Keputusan

Rule	Kode Gejala	Kode Keputusan
1	G1, G2, G3, G4, G5, G6, G7	H1
2	G1, G02, G3, G4, G5, G6, G7	H1
3	G1, G02, G3, G4, G05, G6, G7	H1
4	G1, G02, G3, G4, G05, G6, G07	H1
5	G1, G02, G3, G4, G5, G06, G07	H2

6	G1, G2, G3, G04, G5, G06, G07	H2
7	G01, G2, G3, G4, G5, G06, G07	H2
8	G1, G2, G3, G4, G05, G06, G07	H2
9	G01, G02, G3, G4, G5, G06, G07	H3
10	G01, G2, G3, G04, G5, G06, G07	H3
11	G1, G02, G3, G04, G5, G06, G07	H3
12	G1, G2, G03, G04, G5, G06, G07	H3

Sumber : Priyantono et al. [39]

Pada dataset penyakit obesitas terdapat kolom – kolom gejala yang masih bernilai kategorial, seperti kolom *Gender*, *family\_history\_with\_overweight*, FAVC, CAEC, SMOKE, SCC, CALC, dan MTRANS. Selain itu, label diagnosis dari dataset penyakit obesitas juga masih bersifat kategorial sehingga perlu adanya pengubahan nilai.

Pada dataset penyakit jantung, penyakit kanker payudara dan penyakit umum, tidak ditemukan kolom gejala yang memiliki nilai kategorial sehingga tidak perlu dilakukan pengubahan nilai. Kolom label diagnosis dari penyakit jantung juga tidak perlu dilakukan pengubahan nilai. Kolom label diagnosis dataset penyakit umum memiliki 41 tipe label diagnosis yang masih berupa nilai kategorial sehingga perlu dilakukan pengubahan nilai. Di sisi lain, kolom label diagnosis dataset penyakit kanker payudara memiliki 2 tipe label diagnosis yaitu “B” dan “M” sehingga juga perlu dilakukan pengubahan nilai.

Untuk proses *data mining*, label – label kolom diagnosis dari seluruh dataset penyakit perlu diubah menjadi angka agar bisa diolah. Dalam proses mengolah label – label menjadi angka, untuk kemudahan penelitian dan proses selanjutnya, penelitian ini menggunakan *library scikit-learn* yaitu *LabelEncoder* di dalam file

*Python.* Gambar 4.11 menunjukkan contoh kode yang digunakan dalam mengubah label bernilai kategorial menjadi angka.

```
In [24]: from sklearn.preprocessing import LabelEncoder
labelencoder1 = LabelEncoder()
thyroid["JenisKelamin"] = labelencoder1.fit_transform(thyroid["JenisKelamin"])
thyroid["MengkonsumsiThyroxine"] = labelencoder1.fit_transform(thyroid["MengkonsumsiThyroxine"])
thyroid["ResepThyroxine"] = labelencoder1.fit_transform(thyroid["ResepThyroxine"])
thyroid["MengkonsumsiObatAntitiroid"] = labelencoder1.fit_transform(thyroid["MengkonsumsiObatAntitiroid"])
thyroid["InginMuntah"] = labelencoder1.fit_transform(thyroid["InginMuntah"])
thyroid["Hamil"] = labelencoder1.fit_transform(thyroid["Hamil"])
thyroid["OperasiTerkaitTiroid"] = labelencoder1.fit_transform(thyroid["OperasiTerkaitTiroid"])
thyroid["MenjalaniTerapiI-131"] = labelencoder1.fit_transform(thyroid["MenjalaniTerapiI-131"])
thyroid["PemeriksaanHipotiroid"] = labelencoder1.fit_transform(thyroid["PemeriksaanHipotiroid"])
thyroid["PemeriksaanHipertiroid"] = labelencoder1.fit_transform(thyroid["PemeriksaanHipertiroid"])
thyroid["MengkonsumsiObatLutium"] = labelencoder1.fit_transform(thyroid["MengkonsumsiObatLutium"])
thyroid["BenjolanLeher"] = labelencoder1.fit_transform(thyroid["BenjolanLeher"])
thyroid["PengecekanTumor"] = labelencoder1.fit_transform(thyroid["PengecekanTumor"])
thyroid["PengecekanKondisiHipopituitar"] = labelencoder1.fit_transform(thyroid["PengecekanKondisiHipopituitar"])
thyroid["PengecekanPsikis"] = labelencoder1.fit_transform(thyroid["PengecekanPsikis"])
thyroid["PengukuranTSH"] = labelencoder1.fit_transform(thyroid["PengukuranTSH"])
thyroid["PengukuranT3"] = labelencoder1.fit_transform(thyroid["PengukuranT3"])
thyroid["PengukuranT4"] = labelencoder1.fit_transform(thyroid["PengukuranT4"])
thyroid["PengukuranT4U"] = labelencoder1.fit_transform(thyroid["PengukuranT4U"])
thyroid["PengukuranFTI"] = labelencoder1.fit_transform(thyroid["PengukuranFTI"])
thyroid["PengukuranTBG"] = labelencoder1.fit_transform(thyroid["PengukuranTBG"])
```

Gambar 4.11 Kode Untuk Pelabelan Nilai Kolom Gejala yang Bernilai Kategorial

#### 4.5.2 Tahap Pengubahan Bahasa Kolom Gejala dan Label Diagnosis Dataset Penyakit

Kolom gejala dan label diagnosis dalam dataset penyakit sebagian besar dituliskan dalam bahasa Inggris. Sebagai contoh, dalam dataset penyakit obesitas terdapat kolom gejala *height* dan *weight* yang masing – masing berarti tinggi badan dan berat badan. Beberapa kolom gejala dalam dataset juga masih dituliskan dalam bentuk kode sehingga sulit untuk memahami dataset secara langsung. Misalnya pada dataset penyakit jantung dan penyakit obesitas terdapat kolom gejala dalam bentuk kode. Dataset penyakit jantung memiliki kolom gejala *trestbps* yang sebenarnya mengacu pada pengukuran tekanan darah responden saat beristirahat. Penyakit obesitas memiliki kolom gejala *CH20* yang sebenarnya untuk pengukuran berapa banyak liter air yang diminum per hari. Untuk kemudahan penelitian dan pemahaman data, penelitian ini mengubah semua indikator dan diagnosis menjadi bahasa Indonesia. Label diagnosis yang sudah menggunakan nama medis penyakit tidak akan diubah ke dalam bahasa Indonesia. Beberapa dataset penyakit yang menggunakan kode pada kolom gejalanya juga akan diubah sesuai dengan bahasa Indonesia. Proses pengubahan bahasa menggunakan teknik manual dengan mencari arti dari gejala – gejala dan label diagnosis secara *online* dan menggantinya pada file excel.

Pada dataset penyakit obesitas, berdasarkan tabel 4.2 pada bagian 4.1 tentang penjelasan data, kode kolom gejala dalam dataset penyakit obesitas dirumuskan berdasarkan kuesioner dari penelitian Palechor dan Manotas [38] yang ditulis dalam bahasa Inggris. Melalui proses ini, peneliti menerjemahkan kuesioner tersebut ke dalam bahasa Indonesia dan membuat kode baru berdasarkan kuesioner yang telah diterjemahkan. Kolom gejala yang akan diubah antara lain adalah kolom *Gender*, *Age*, *Height*, *Weight*, *family\_history\_with\_overweight*, FAVC, FCVC, NCP, CAEC, SMOKE, CH20, SCC, FAF, TUE, CALC, dan MTRANS. Selain itu kolom label diagnosis dataset penyakit obesitas masih bernilai “*NObeyesdad*” sehingga akan diubah menjadi “diagnosis”. Nilai label diagnosis dari dataset penyakit obesitas yang terdiri dari *Normal\_Weight*, *Insufficient\_Weight*, *Overweight\_Level\_I*, *Overweight\_Level\_II*, *Obesity\_Type\_I*, *Obesity\_Type\_II*, dan *Obesity\_Type\_III* juga akan diubah ke dalam bentuk bahasa Indonesia. Selain itu, proses ini juga menghilangkan seluruh tanda *underscore* (\_) yang ada pada label diagnosis. Tabel 4.13 menunjukkan pengubahan bahasa yang dilakukan pada dataset penyakit obesitas.

Tabel 4.13 Tabel Pengubahan Bahasa dalam Dataset Penyakit Obesitas

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>Gender</i>	JenisKelamin
Kolom Gejala	<i>Age</i>	Umur
Kolom Gejala	<i>Height</i>	TinggiBadan
Kolom Gejala	<i>Weight</i>	BeratBadan
Kolom Gejala	<i>family_history_with_overweight</i>	RiwayatKegemukan
Kolom Gejala	FAVC	FKMBT (Frekuensi Konsumsi Makanan Berkalori Tinggi)
Kolom Gejala	FCVC	FKS (Frekuensi Konsumsi Sayuran)
Kolom Gejala	NCP	MUS (Makanan Utama dalam Sehari)

Tabel 4.13 Tabel Pengubahan Bahasa dalam Dataset Penyakit Obesitas (lanjutan)

Jenis Kolom	Nama Kolom/Label Diagnosis (Bahasa Inggris)	Nama Kolom/Label Diagnosis (Bahasa Indonesia)
Kolom Gejala	CAEC	FMAMU (Frekuensi Makan diantara Makanan Utama)
Kolom Gejala	SMOKE	ROKOK
Kolom Gejala	CH20	KAS (Konsumsi Air Sehari)
Kolom Gejala	SCC	MK (Monitor Kalori)
Kolom Gejala	FAF	FAF (Frekuensi Aktivitas Fisik)
Kolom Gejala	TUE	PAE (Penggunaan Alat Elektronik)
Kolom Gejala	CALC	FKA (Frekuensi Konsumsi Alkohol)
Kolom Gejala	MTRANS	AlatTransportasi
Label Diagnosis	Normal_Weight	BeratBadanNormal
Label Diagnosis	Insufficient_Weight	BeratBadanKurang
Label Diagnosis	Overweight_Level_I	KegemukanTingkatI
Label Diagnosis	Overweight_Level_II	KegemukanTingkatII
Label Diagnosis	Obesity_Level_I	ObesitasTingkatI
Label Diagnosis	Obesity_Level_II	ObesitasTingkatII
Label Diagnosis	Obesity_Level_III	ObesitasTingkatIII

Pada dataset penyakit tiroid, kolom gejala yang akan diubah menjadi bahasa Indonesia adalah *age, sex, on thyroxine, query on thyroxine, on antithyroid medication, sick, pregnant, thyroid surgery, I131 treatment, query hypothyroid, query hyperthyroid, lithium, goitre, tumor, hypopituitary, psych, TSH measured, T3 measured, TT4 measured, T4U measured, FTI measured, dan TBG measured*. Label diagnosis dari dataset penyakit tiroid yang terdiri dari *negative, compensated hypothyroid, primary hypothyroid, hyperthyroid, secondary hypothyroid, goitre, T3 toxic, toxic goitre, secondary toxic, dan concurrent non-thyroidal illness* juga akan diubah. Tabel 4.14 menunjukkan pengubahan bahasa yang dilakukan pada dataset penyakit tiroid dan keterangan dari tiap kolom dan label diagnosis.

Tabel 4.14 Tabel Pengubahan Bahasa dalam Dataset Penyakit Tiroid

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>age</i>	Umur
Kolom Gejala	<i>sex</i>	JenisKelamin
Kolom Gejala	<i>on thyroxine</i>	MengkonsumsiThyroxine
Kolom Gejala	<i>query on thyroxine</i>	ResepThyroxine
Kolom Gejala	<i>on antithyroid medication</i>	MengkonsumsiObatAntitiroid
Kolom Gejala	<i>sick</i>	InginMuntah
Kolom Gejala	<i>pregnant</i>	Hamil
Kolom Gejala	<i>thyroid surgery</i>	OperasiTerkaitTiroid
Kolom Gejala	<i>I131 treatment</i>	MenjalaniTerapiI-131
Kolom Gejala	<i>query hypothyroid.</i>	PemeriksaanHipotiroid
Kolom Gejala	<i>query hyperthyroid</i>	PemeriksaanHipertiroid
Kolom Gejala	<i>lithium</i>	MengkonsumsiObatLitium
Kolom Gejala	<i>goitre</i>	BenjolanLeher
Kolom Gejala	<i>tumor</i>	PengecekanTumor
Kolom Gejala	<i>hypopituitary</i>	PengecekanKondisiHipopituitari
Kolom Gejala	<i>psych</i>	PengecekanPsikis
Kolom Gejala	<i>TSH measured</i>	PengukuranTSH
Kolom Gejala	<i>TSH</i>	TSH
Kolom Gejala	<i>T3 measured</i>	PengukuranT3
Kolom Gejala	<i>TT4 measured</i>	PengukuranTT4
Kolom Gejala	<i>TT4</i>	TT4
Kolom Gejala	<i>T4U measured</i>	PengukuranT4U
Kolom Gejala	<i>T4U</i>	T4U
Kolom Gejala	<i>FTI measured</i>	PengukuranFTI
Kolom Gejala	<i>FTI</i>	FTI
Kolom Gejala	<i>TBG measured</i>	PengukuranTBG
Label Diagnosis	<i>negative</i>	Negatif
Label Diagnosis	<i>compensated hypothyroid</i>	HipotiroidTerkompensasi
Label Diagnosis	<i>primary hypothyroid</i>	HipotiroidUtama
Label Diagnosis	<i>hyperthyroid</i>	Hipertiroid

Tabel 4.14 Tabel Pengubahan Bahasa dalam Dataset Penyakit Tiroid (lanjutan)

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Label Diagnosis	<i>secondary hypothyroid</i>	HipotiroidSekunder
Label Diagnosis	<i>Goitre</i>	ToxicGoitre
Label Diagnosis	<i>T3 toxic</i> (Kode B)	Tirotoksikosis
Label Diagnosis	<i>toxic goitre</i> (Kode C)	ToxicGoitre
Label Diagnosis	<i>secondary toxic</i> (Kode D)	ToxicMultinodularGoitre
Label Diagnosis	<i>concurrent non-thyroidal illness</i> (Kode K)	EuthyroidSickSyndrome

Pada dataset penyakit kanker payudara, kolom gejala yang akan diubah ke dalam bahasa Indonesia adalah *radius*, *texture*, *perimeter*, *area*, *smoothness*, *compactness*, *concavity*, *concave points*, *symmetry*, dan *fractal dimension*. Seperti yang telah dijelaskan pada bagian 4.1, masing-masing kolom memiliki 3 jenis, yaitu *mean*, *standard error* (SE), dan *worst*. Hanya jenis *worst* yang akan diterjemahkan, sedangkan *mean* dan *standard error* tidak akan diubah. Proses ini juga menghilangkan tanda *underscore* (\_) pada kolom gejala. Tabel 4.15 menunjukkan pengubahan bahasa yang dilakukan pada dataset penyakit kanker payudara.

Tabel 4.15 Pengubahan Bahasa dalam Dataset Penyakit Kanker Payudara

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>radius_mean</i>	RadiusMean
Kolom Gejala	<i>texture_mean</i>	TekturMean
Kolom Gejala	<i>perimeter_mean</i>	Kelilingean
Kolom Gejala	<i>area_mean</i>	LuasMean
Kolom Gejala	<i>smoothness_mean</i>	KehalusanMean

Tabel 4.15 Pengubahan Bahasa dalam Dataset Penyakit Kanker Payudara (lanjutan)

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>compactness_mean</i>	KepadatanMean
Kolom Gejala	<i>concavity_mean</i>	KecekunganMean
Kolom Gejala	<i>concave points_mean</i>	JumlahCekunganMean
Kolom Gejala	<i>symmetry_mean</i>	SimetriMean
Kolom Gejala	<i>fractal_dimension_mean</i>	DimensiFraktalMean
Kolom Gejala	<i>radius_se</i>	RadiusStandarError
Kolom Gejala	<i>texture_se</i>	TekturStandarError
Kolom Gejala	<i>perimeter_se</i>	KelilingStandarError
Kolom Gejala	<i>area_se</i>	LuasStandarError
Kolom Gejala	<i>smoothness_se</i>	KehalusanStandarError
Kolom Gejala	<i>compactness_se</i>	KepadatanStandarError
Kolom Gejala	<i>concavity_se</i>	KecekunganStandarError
Kolom Gejala	<i>concave points_se</i>	JumlahCekunganStandarError
Kolom Gejala	<i>symmetry_se</i>	SimetriStandarError
Kolom Gejala	<i>fractal_dimension_se</i>	DimensiFraktalStandarError
Kolom Gejala	<i>radius_worst</i>	RadiusTerburuk
Kolom Gejala	<i>texture_worst</i>	TekturTerburuk
Kolom Gejala	<i>perimeter_worst</i>	KelilingTerburuk
Kolom Gejala	<i>area_worst</i>	LuasTerburuk
Kolom Gejala	<i>smoothness_worst</i>	KehalusanTerburuk
Kolom Gejala	<i>compactness_worst</i>	KepadatanTerburuk
Label Diagnosis	<i>concavity_worst</i>	KecekunganTerburuk
Label Diagnosis	<i>concave points_worst</i>	JumlahCekunganTerburuk
Label Diagnosis	<i>symmetry_worst</i>	SimetriTerburuk
Label Diagnosis	<i>fractal_dimension_worst</i>	DimensiFraktalTerburuk
Label Diagnosis	<i>B</i>	Jinak
Label Diagnosis	<i>M</i>	Ganas

Pada dataset penyakit umum, seluruh kolom gejala akan diubah ke dalam bahasa Indonesia. Beberapa kolom gejala juga masih memiliki tanda *underscore*

(\_) sehingga tanda tersebut akan dihapus. Beberapa label diagnosis dari dataset penyakit umum akan diubah menjadi bahasa Indonesia yaitu *Fungal infection*, *Allergy*, *GERD*, *Chronic Cholestasis*, *Acne*, *Alcoholic hepatitis*, *Arthritis*, *Bronchial Asthma*, *Cervical Spondylosis*, *Chicken Pox*, *Common Cold*, *Dengue*, *Dimorphic Hemorrhoids (piles)*, *Drug Reaction*, *Heart Attack*, *Hypertension*, *Hyperthyroidism*, *Hypothyroidism*, *Jaundice*, *Migraine*, *Osteoarthristis*, *Paralysis (Brain Hemorrhage)*, *Peptic Ulcer Disease*, *Thyphoid*, *Urinary Tract Infection* dan *Varicose Veins*. Tabel 4.16 menunjukkan pengubahan bahasa yang dilakukan pada dataset penyakit umum.

Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum

Jenis Kolom	Nama Kolom/Label Diagnosis (Bahasa Inggris)	Nama Kolom/Label Diagnosis (Bahasa Indonesia)
Kolom Gejala	<i>itching</i>	Gatal
Kolom Gejala	<i>skin_rash</i>	RuamKulit
Kolom Gejala	<i>continuous_sneezing</i>	BersinTerusMenerus
Kolom Gejala	<i>shivering</i>	Menggigil
Kolom Gejala	<i>chills</i>	Panas_Dingin
Kolom Gejala	<i>joint_pain</i>	NyeriPadaSendi
Kolom Gejala	<i>acidity</i>	AsamLambungNaik
Kolom Gejala	<i>ulcers_on_tongue</i>	Sariawan
Kolom Gejala	<i>muscle_wasting</i>	OtotMatiRasa
Kolom Gejala	<i>vomiting</i>	Muntah-Muntah
Kolom Gejala	<i>burning_micturition</i>	NyeriSaatBuangAirKecil
Kolom Gejala	<i>spotting_urination</i>	TitikDarahDiAirSeni
Kolom Gejala	<i>fatigue</i>	Kelelahan
Kolom Gejala	<i>weight_gain</i>	BeratBadanNaik
Kolom Gejala	<i>cold_hands_and_feets</i>	KakiDanTanganDingin
Kolom Gejala	<i>mood_swings</i>	PerubahanSuasanaHati
Kolom Gejala	<i>weight_loss</i>	BeratBadanTurun
Kolom Gejala	<i>restlessness</i>	Cemas
Kolom Gejala	<i>patches_in_throat</i>	TitikPutihDiTenggorokan
Kolom Gejala	<i>irregular_sugar_level</i>	GulaDarahTidakNormal

Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>cough</i>	Batuk
Kolom Gejala	<i>high_fever</i>	DemamTinggi
Kolom Gejala	<i>sunken_eyes</i>	KantongMata
Kolom Gejala	<i>breathlessness</i>	SulitBernafas
Kolom Gejala	<i>sweating</i>	Berkeringat
Kolom Gejala	<i>dehydration</i>	Dehidrasi
Kolom Gejala	<i>indigestion</i>	RasaTidakNyamanDiPerutBagian Atas
Kolom Gejala	<i>headache</i>	SakitKepala
Kolom Gejala	<i>yellowish_skin</i>	KulitMenguning
Kolom Gejala	<i>dark_urine</i>	AirSeniBerwarnaGelap
Kolom Gejala	<i>nausea</i>	PerasaanInginMuntah
Kolom Gejala	<i>loss_of_appetite</i>	KehilanganNafsuMakan
Kolom Gejala	<i>pain_behind_the_eyes</i>	RasaNyeriDiBelakangMata
Kolom Gejala	<i>back_pain</i>	NyeriPunggung
Kolom Gejala	<i>constipation</i>	Sembelit
Kolom Gejala	<i>abdominal_pain</i>	NyeriPadaPerut
Kolom Gejala	<i>diarrhoea</i>	Diare
Kolom Gejala	<i>mild_fever</i>	DemamRingan
Kolom Gejala	<i>yellow_urine</i>	AirSeniBerwarnaKuning
Kolom Gejala	<i>yellowing_of_eyes</i>	MataMenguning
Kolom Gejala	<i>acute_liver_failure</i>	GagalGinjal
Kolom Gejala	<i>swelling_of_stomach</i>	PerutMembengkak
Kolom Gejala	<i>swelled_lymph_nodes</i>	BenjolanPadaLeher
Kolom Gejala	<i>malaise</i>	RasaTidakEnak
Kolom Gejala	<i>blurred_and_distorted _vision</i>	PenglihatanTidakJelas
Kolom Gejala	<i>phlegm</i>	Dahak
Kolom Gejala	<i>throat_irritation</i>	RasaGatalPadaTenggorokan
Kolom Gejala	<i>redness_of_eyes</i>	MataMemerah
Kolom Gejala	<i>sinus_pressure</i>	HidungTerasaTertekan
Kolom Gejala	<i>runny_nose</i>	Pilek

Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>congestion</i>	HidungTersumbat
Kolom Gejala	<i>chest_pain</i>	DadaTerasaSakit
Kolom Gejala	<i>weakness_in_limbs</i>	KakiTerasaLemas
Kolom Gejala	<i>fast_heart_rate</i>	DetakJantungCepat
Kolom Gejala	<i>pain_during_bowel_movements</i>	NyeriSaatBuangAirBesar
Kolom Gejala	<i>pain_in_anal_region</i>	NyeriSekitarBokong
Kolom Gejala	<i>bloody_stool</i>	DarahPadaFeses
Kolom Gejala	<i>irritation_in_anus</i>	IritasiPadaAnus
Kolom Gejala	<i>neck_pain</i>	NyeriPadaLeher
Kolom Gejala	<i>dizziness</i>	Pusing
Kolom Gejala	<i>cramps</i>	Keram
Kolom Gejala	<i>bruising</i>	Lebam
Kolom Gejala	<i>obesity</i>	BeratBadanBerlebihan
Kolom Gejala	<i>swollen_legs</i>	KakiMembengkak
Kolom Gejala	<i>swollen_blood_vessels</i>	PembuluhDarahMembengkak
Kolom Gejala	<i>puffy_face_and_eyes</i>	WajahDanMataMengembung
Kolom Gejala	<i>brittle_nails</i>	KukuRapuhDanKering
Kolom Gejala	<i>swollen_extremities</i>	PembengkakanEkstrim
Kolom Gejala	<i>excessive_hunger</i>	RasaLaparBerlebihan
Kolom Gejala	<i>extra_marital_contacts</i>	BerhubunganSeksDiLuarPernikahan
Kolom Gejala	<i>drying_and_tingling_lips</i>	BibirKering
Kolom Gejala	<i>slurred_speech</i>	BerbicaraTidakLancar
Kolom Gejala	<i>knee_pain</i>	NyeriPadaLutut
Kolom Gejala	<i>hip_joint_pain</i>	NyeriPadaPinggul
Kolom Gejala	<i>muscle_weakness</i>	OtotMelemah
Kolom Gejala	<i>stiff_neck</i>	LeherTerasaKaku
Kolom Gejala	<i>swelling_joints</i>	SendiMembengkak
Kolom Gejala	<i>movement_stiffness</i>	KakuSaatBergerak
Kolom Gejala	<i>spinning_movements</i>	PusingSepertiBerputar

Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>loss_of_balance</i>	KehilanganKeseimbangan
Kolom Gejala	<i>weakness_of_one_body_side</i>	LemasPadaSatuSisiTubuh
Kolom Gejala	<i>loss_of_smell</i>	KehilanganInderaPenciuman
Kolom Gejala	<i>bladder_discomfort</i>	NyeriPadaSaluranKemih
Kolom Gejala	<i>foul_smell_of_urine</i>	AirSeniBerbauTidakSedap
Kolom Gejala	<i>continuous_feel_of_urine</i>	InginSelaluBuangAirKecil
Kolom Gejala	<i>internal_itching</i>	DalamKulitTerasaGatal
Kolom Gejala	<i>toxic_look_(typhos)</i>	Pucat
Kolom Gejala	<i>depression</i>	Depresi
Kolom Gejala	<i>irritability</i>	CepatMarah
Kolom Gejala	<i>muscle_pain</i>	NyeriPadaOtot
Kolom Gejala	<i>red_spots_over_body</i>	TitikMerahDiSeluruhTubuh
Kolom Gejala	<i>abnormal_menstruation</i>	MenstruasiTidakWajar
Kolom Gejala	<i>dischromic_patches</i>	BercakCokelatPadaKulit
Kolom Gejala	<i>watering_from_eyes</i>	MataBerair
Kolom Gejala	<i>increased_appetite</i>	NafsuMakanMeningkat
Kolom Gejala	<i>polyuria</i>	SeringBuangAirKecil
Kolom Gejala	<i>family_history</i>	RiwayatKeluarga
Kolom Gejala	<i>mucoid_sputum</i>	DahakJernih
Kolom Gejala	<i>rusty_sputum</i>	DahakBerwarnaKuning
Kolom Gejala	<i>lack_of_concentration</i>	KurangKonsentrasi
Kolom Gejala	<i>visual_disturbances</i>	PenglihatanTerganggu
Kolom Gejala	<i>receiving_blood_transfusion</i>	MenerimaTransfusiDarah
Kolom Gejala	<i>history_of_alcohol_consumption</i>	RiwayatKonsumsiAlkohol
Kolom Gejala	<i>fluid_overload</i>	BengkakPadaKakiDanTangan
Kolom Gejala	<i>blood_in_sputum</i>	DahakBerdarah

Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>prominent_veins_on_calf</i>	PembuluhDarahDiBetisTerlihat
Kolom Gejala	<i>palpitations</i>	JantungBerdebarDebar
Kolom Gejala	<i>painful_walking</i>	SakitSaatBerjalan
Kolom Gejala	<i>pus_filled_pimples</i>	JerawatDenganNanah
Kolom Gejala	<i>blackheads</i>	Komedo
Kolom Gejala	<i>skin_peeling</i>	KulitMengelupas
Kolom Gejala	<i>small_dents_in_nails</i>	LekukanKecilDiKuku
Kolom Gejala	<i>inflammatory_nails</i>	KukuMembengkak
Kolom Gejala	<i>blister</i>	BenjolanBerisiCairan
Kolom Gejala	<i>red_sore_around_nose</i>	BenjolanMerahSekitarHidung
Kolom Gejala	<i>yellow_crust_ooze</i>	Nanah
Label Diagnosis	<i>(vertigo) Paroxysmal Positional Vertigo</i>	Vertigo
Label Diagnosis	<i>Fungal infection</i>	InfeksiJamur
Label Diagnosis	<i>Allergy</i>	Alergi
Label Diagnosis	<i>GERD</i>	GERD
Label Diagnosis	<i>Chronic Cholestasis</i>	PenyakitHatiAkibatKolestasis
Label Diagnosis	<i>Acne</i>	Jerawat
Label Diagnosis	<i>AIDS</i>	AIDS
Label Diagnosis	<i>Alcoholic hepatitis</i>	HepatitisMinumanBeralkohol
Label Diagnosis	<i>Arthritis</i>	Arthritis
Label Diagnosis	<i>Bronchial Asthma</i>	Asma
Label Diagnosis	<i>Cervical Spondylosis</i>	SpondilosisServikal
Label Diagnosis	<i>Chicken Pox</i>	CacarAir
Label Diagnosis	<i>Common Cold</i>	Selesma(BatuPilek)
Label Diagnosis	<i>Dengue</i>	DBD
Label Diagnosis	<i>Diabetes</i>	Diabetes
Label Diagnosis	<i>Dimorphic Hemorrhoids (piles)</i>	Ambeien/Wasir
Label Diagnosis	<i>Drug Reaction</i>	ReaksiPemakaianObat
Label Diagnosis	<i>Gastroenteritis</i>	Muntaber

Tabel 4.16 Pengubahan Bahasa dalam Dataset Penyakit Umum (lanjutan)

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Label Diagnosis	<i>Heart Attack</i>	Serangan Jantung
Label Diagnosis	<i>Hepatitis A</i>	Hepatitis A
Label Diagnosis	<i>Hepatitis B</i>	Hepatitis B
Label Diagnosis	<i>Hepatitis C</i>	Hepatitis C
Label Diagnosis	<i>Hepatitis D</i>	Hepatitis D
Label Diagnosis	<i>Hepatitis E</i>	Hepatitis E
Label Diagnosis	<i>Hypertension</i>	Hipertensi
Label Diagnosis	<i>Hypoglycemia</i>	Hipoglisemia
Label Diagnosis	<i>Hyperthyroidism</i>	Hipertiroid
Label Diagnosis	<i>Hypothyroidism</i>	Hipotiroid
Label Diagnosis	<i>Impetigo</i>	Impetigo
Label Diagnosis	<i>Jaundice</i>	Penyakit Kuning
Label Diagnosis	<i>Malaria</i>	Malaria
Label Diagnosis	<i>Migraine</i>	Migren
Label Diagnosis	<i>Osteoarthritis</i>	Osteoartritis
Label Diagnosis	<i>Paralysis (Brain Hemorrhage)</i>	Kelumpuhan
Label Diagnosis	<i>Peptic Ulcer Disease</i>	Tukak Lambung
Label Diagnosis	<i>Pneumonia</i>	Pneumonia
Label Diagnosis	<i>Psoriasis</i>	Psoriasis
Label Diagnosis	<i>Tuberculosis</i>	TBC
Label Diagnosis	<i>Typhoid</i>	Tipes (Infeksi bakteri)
Label Diagnosis	<i>Urinary Tract Infection</i>	Infeksi Saluran Kemih
Label Diagnosis	<i>Varicose Veins</i>	Varises

Pada dataset penyakit jantung, sebagian besar dari kolom gejala memiliki nilai berupa kode berdasarkan kata – kata dalam bahasa Inggris. Seperti yang telah dilakukan pada dataset penyakit obesitas, peneliti menerjemahkan kata - kata tersebut ke dalam bahasa Indonesia dan membuat kode baru berdasarkan kata – kata baru yang terbentuk. Seluruh kolom gejala akan diubah ke dalam bahasa Indonesia.

Tabel 4.17 menunjukkan pengubahan bahasa yang dilakukan pada dataset penyakit jantung.

Tabel 4.17 Pengubahan Bahasa dalam Dataset Penyakit Jantung

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>age</i>	Umur
Kolom Gejala	<i>sex</i>	JenisKelamin
Kolom Gejala	<i>cp</i>	NyeriDada
Kolom Gejala	<i>trestbp</i>	TekananDarahNormal
Kolom Gejala	<i>chol</i>	Kolesterol
Kolom Gejala	<i>Fbs</i>	GulaDarahDiatas120
Kolom Gejala	<i>restecg</i>	ElektrokardiografiNormal
Kolom Gejala	<i>thalach</i>	DetakJantungMaksimal
Kolom Gejala	<i>Exang</i>	NyeriDadaSaatOlahraga
Kolom Gejala	<i>oldpeak</i>	PenurunanSegmenST
Kolom Gejala	<i>slope</i>	TitikNaikSegmenST
Kolom Gejala	<i>ca</i>	PUB (Pembuluh Utama Berwarna Saat Tes Flouroskopi)
Kolom Gejala	<i>thal</i>	ThalliumStressTest

Pada dataset penyakit COVID-19, seluruh kolom gejala akan diubah ke dalam bahasa Indonesia. Sebagian besar kolom gejala juga memiliki tanda *underscore* (\_) sehingga tanda tersebut akan dihapus. Tabel 4.18 menunjukkan pengubahan bahasa yang dilakukan pada dataset penyakit COVID-19.

Tabel 4.18 Pengubahan Bahasa dalam Dataset Penyakit COVID-19

<b>Jenis Kolom</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Inggris)</b>	<b>Nama Kolom/Label Diagnosis (Bahasa Indonesia)</b>
Kolom Gejala	<i>Fever</i>	Demam
Kolom Gejala	<i>Tiredness</i>	Kelelahan
Kolom Gejala	<i>Dry-Cough</i>	BatukKering

Tabel 4.18 Pengubahan Bahasa dalam Dataset Penyakit COVID-19 (lanjutan)

Jenis Kolom	Nama Kolom/Label Diagnosis (Bahasa Inggris)	Nama Kolom/Label Diagnosis (Bahasa Indonesia)
Kolom Gejala	<i>Difficulty-In-Breathing</i>	KesulitanBernafas
Kolom Gejala	<i>Sore-Throat</i>	SakitTenggorokan
Kolom Gejala	<i>Pains</i>	Nyeri
Kolom Gejala	<i>Nasal-Congestion</i>	HidungTersumbat
Kolom Gejala	<i>Runny-Nose</i>	Pilek
Kolom Gejala	<i>Diarrhea</i>	Diare
Kolom Gejala	<i>Age_0-9</i>	Umur0-9
Kolom Gejala	<i>Age_10-19</i>	Umur10-19
Kolom Gejala	<i>Age_20-24</i>	Umur20-24
Kolom Gejala	<i>Age_25-59</i>	Umur25-59
Kolom Gejala	<i>Age_60+</i>	Umur60+
Kolom Gejala	<i>Gender_Female</i>	JenisKelaminPerempuan
Kolom Gejala	<i>Gender-Male</i>	JenisKelaminLaki-Laki
Kolom Gejala	<i>Gender_Transgender</i>	JenisKelaminTransgender
Kolom Gejala	<i>Contact_Dont_Know</i>	KontakTidakTahu
Kolom Gejala	<i>Contact_Yes</i>	AdaKontak
Kolom Gejala	<i>Contact_No</i>	TidakAdaKontak
Kolom Gejala	<i>Country</i>	Negara

#### 4.6 Tahap Data Mining

Pada subbab ini dijelaskan tahapan *data mining* untuk mendapatkan informasi berdasarkan data – data yang telah dikumpulkan dan diproses pada tahapan sebelumnya. Tahapan *data mining* terdiri dari tahap preprocessing, pemisahan dataset, dan pemodelan algoritma klasifikasi.

#### 4.6.1 Preprocessing

```
from sklearn.preprocessing import StandardScaler  
sc = StandardScaler()  
symptoms = sc.fit_transform(symptoms)
```

Gambar 4.12 Baris Kode untuk Melakukan Preprocessing

Tahapan *preprocessing* yang dilakukan dalam penelitian ini adalah standarisasi. Standarisasi merupakan proses dimana dataset yang memiliki nilai pengukuran yang berbeda – beda disetarakan sehingga rentang nilai dari tiap data tidak berbeda jauh. Proses penyetaraan dilakukan dengan cara menghilangkan nilai mean dan melakukan penskalaan ke varians unit [40]. Proses ini menggunakan library *StandardScaler* miliki *scikit-learn*. Gambar 4.12 menunjukkan contoh baris kode yang digunakan untuk melakukan proses standarisasi.

#### 4.6.2 Pemisahan Dataset

```
from sklearn.model_selection import train_test_split  
symp_train, symp_test, diagnosis_train, diagnosis_test = train_test_split(symptoms,  
diagnosis, test_size = 0.25, random_state = 0)
```

Gambar 4.13 Baris Kode untuk Memisahkan Data *Training* dan Data *Testing*

Tahapan pemisahan dataset dilakukan dengan membagi tiap dataset penyakit ke dalam 2 tipe data, yaitu *data training* dan *data testing*. Pemisahan dataset ini menggunakan library *scikit-learn Train\_Test\_Split*. Persentase pembagian dataset yang akan digunakan adalah 75% untuk *Train Dataset* dan 25% untuk *Test Dataset*. Data yang akan menjadi *data training* dan *data testing* dipilih secara acak oleh program. Gambar 4.13 menunjukkan fungsi untuk membagi dataset menjadi *data training* dan *data testing*.

#### 4.6.3 Pemodelan Algoritma Klasifikasi

Tahapan pemodelan algoritma klasifikasi dilakukan dengan membandingkan tiga algoritma, yaitu *Logistic Regression*, *Support Vector*

*Machine*, dan *Gaussian Naïve Bayes*. Pemodelan algoritma klasifikasi ini menggunakan *library scikit-learn*. Pada tahapan ini masing – masing algoritma dilakukan fungsi *import* sehingga dapat dipakai. Gambar 4.14 menunjukkan kode pemodelan algoritma klasifikasi.

```

Model dengan Algoritma Logistic Regression
In [14]: from sklearn.linear_model import LogisticRegression
classifier_LR = LogisticRegression(random_state=0, max_iter=10000)
classifier_LR.fit(symp_train, diagnosis_train)
Out[14]: LogisticRegression(max_iter=10000, random_state=0)

In [15]: diagnosis_pred_LR = classifier_LR.predict(symp_test)

Model dengan Algoritma Support Vector Machine
In [16]: from sklearn.svm import LinearSVC
classifier_SVM = LinearSVC(max_iter=100000, random_state = 0)
classifier_SVM.fit(symp_train, diagnosis_train)
Out[16]: LinearSVC(max_iter=100000, random_state=0)

In [17]: diagnosis_pred_SVM = classifier_SVM.predict(symp_test)

Model dengan Algoritma Gaussian Naïve Bayes
In [18]: from sklearn.naive_bayes import GaussianNB
classifier_GNB = GaussianNB()
classifier_GNB.fit(symp_train, diagnosis_train)
Out[18]: GaussianNB()

In [19]: diagnosis_pred_GNB = classifier_GNB.predict(symp_test)

```

Gambar 4.14 Kode untuk Memodelkan Algoritma Klasifikasi

## 4.7 Tahap Pattern Evaluation

Tahap *pattern evaluation* akan melakukan evaluasi terhadap pemodelan algoritma klasifikasi dalam melakukan *data mining*. Nilai evaluasi yang akan dibandingkan yaitu akurasi, *recall*, *error rate*, sensitivitas, presisi (*precision*), dan *F1-score*. Evaluasi model algoritma klasifikasi ini menggunakan *library scikit-learn confusion matrix* dan juga fungsi – fungsi untuk menghasilkan nilai evaluasi yang akan dibandingkan. Gambar 4.15 hingga 4.17 menunjukkan baris kode yang digunakan untuk melakukan evaluasi.

```

In [20]: from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.metrics import recall_score
from sklearn.metrics import precision_score
from sklearn.metrics import f1_score
from sklearn.metrics import plot_confusion_matrix
cm_LR = confusion_matrix(diagnosis_test, diagnosis_pred_LR)
cm_LR

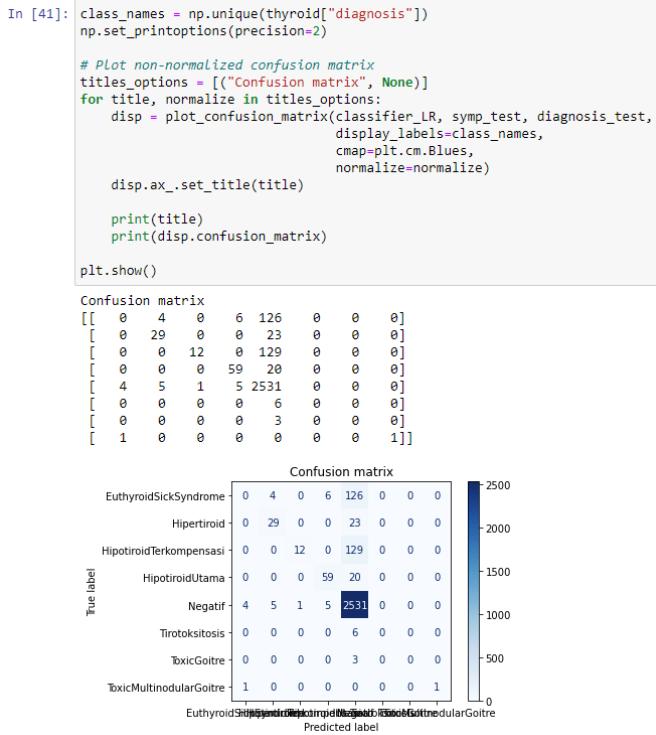
```

```

Out[20]: array([[ 0,  4,  0,  6, 126,  0,  0,  0],
   [ 0, 29,  0,  0, 23,  0,  0,  0],
   [ 0,  0, 12,  0, 129,  0,  0,  0],
   [ 0,  0,  0, 59, 20,  0,  0,  0],
   [ 4,  5,  1,  5, 2531,  0,  0,  0],
   [ 0,  0,  0,  0,  6,  0,  0,  0],
   [ 0,  0,  0,  0,  3,  0,  0,  0],
   [ 1,  0,  0,  0,  0,  0,  0, 1]], dtype=int64)

```

Gambar 4.15 Baris Kode untuk Menunjukkan *Confusion Matrix*



Gambar 4.16 Baris Kode untuk Menunjukkan Plot Confusion Matrix

```

In [22]: # Accuracy
accuracy_score(diagnosis_test, diagnosis_pred_LR)
Out[22]: 0.887689713322091

In [23]: # Recall
recall_score(diagnosis_test, diagnosis_pred_LR, average='weighted')
Out[23]: 0.887689713322091

In [24]: # Precision
precision_score(diagnosis_test, diagnosis_pred_LR, average='weighted', zero_division=0)
Out[24]: 0.8462893774326452

In [25]: # F1 Score
f1_score(diagnosis_test, diagnosis_pred_LR, average='weighted')
Out[25]: 0.8475932636517688

In [26]: # Error rate
error_rate = 1 - accuracy_score(diagnosis_test, diagnosis_pred_LR)
error_rate
Out[26]: 0.11231028667790899

```

Gambar 4.17 Baris Kode untuk Menunjukkan Nilai Evaluasi Algoritma

## 4.8 Analisa Data

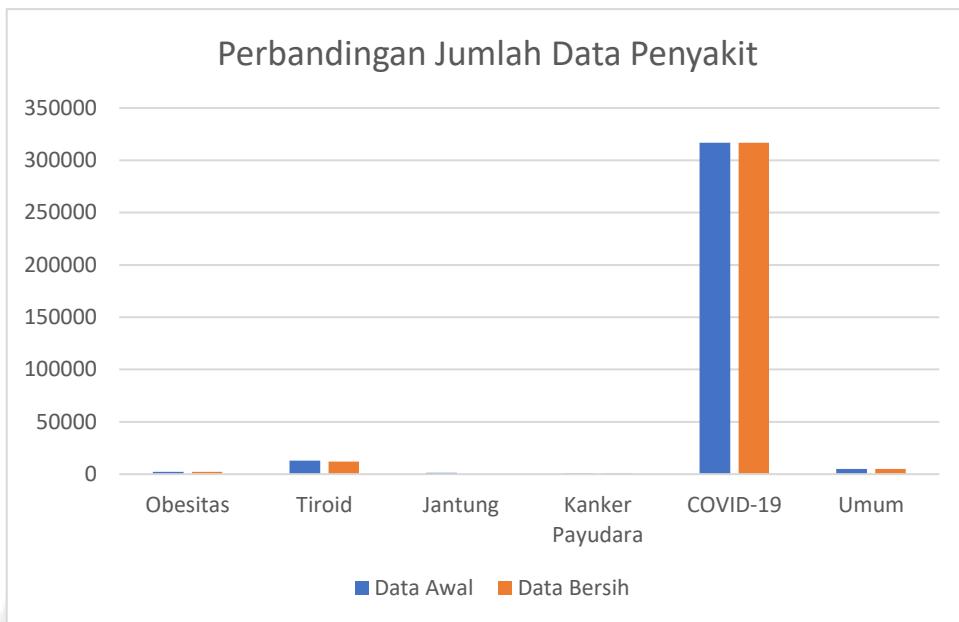
Tahapan ini akan menjelaskan pembahasan analisis dan hasil evaluasi dari pemodelan algoritma klasifikasi pada subbab sebelumnya. Tahap ini terdiri dari tahap eksplorasi data yang digunakan pada penelitian, analisis perbandingan algoritma klasifikasi, dan diskusi hasil penelitian.

#### 4.8.1 Eksplorasi Data

Pada tahapan ini, dilakukan rekapitulasi terakhir mengenai jumlah data yang dipakai dalam penelitian ini. Pada bagian 4.1.1 tentang penjelasan data, total data awal yang ditemukan berjumlah 338.600 data, dengan rincian dataset penyakit obesitas memiliki data sebanyak 2.111 data, dataset penyakit tiroid memiliki data sebanyak 12.944 data, dataset penyakit kanker payudara memiliki data sebanyak 569 data, dataset penyakit umum memiliki data sebanyak 4.962 data, dataset penyakit jantung memiliki data sebanyak 1.214 data, dan dataset penyakit COVID-19 memiliki data sebanyak 316.800 data. Setelah dilakukannya proses *data cleaning*, *data integration*, dan *data selection*, didapatkan data sebanyak 337.686 data. Data ini terdiri dari data penyakit obesitas sebanyak 2.111 data, data penyakit tiroid sebanyak 11.860 data, data penyakit kanker payudara sebanyak 569 data, data penyakit umum sebanyak 4.962 data, data penyakit jantung sebanyak 1.214 data, dan data penyakit COVID-19 sebanyak 316.800 data. Penjelasan mengenai data dan jumlah dari data dapat dilihat pada tabel 4.19.

Tabel 4.19 Tabel Rekapitulasi Data Penelitian

Jenis Penyakit	Jumlah Data
Obesitas	2.111
Tiroid	11.860
Kanker Payudara	569
Umum	4.962
Jantung	300
COVID-19	316.800
<b>TOTAL</b>	<b>336.602</b>

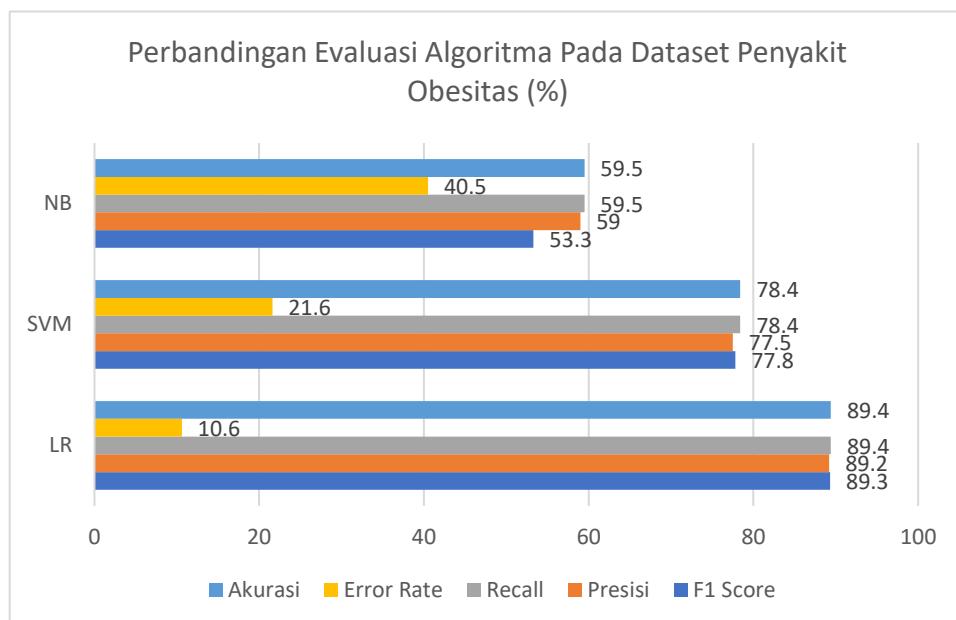


Gambar 4.18 Grafik Perbandingan Jumlah Data Penyakit

Gambar 4.18 menunjukkan perbandingan jumlah data dari tiap penyakit. Melalui gambar 4.18 dapat dilihat bahwa sebagian besar data yang digunakan dalam penelitian ini merupakan data penyakit COVID-19. Data penyakit tiroid merupakan dataset kedua terbanyak dalam penelitian, dilanjutkan dengan data penyakit umum, penyakit obesitas, penyakit kanker payudara, dan penyakit jantung.

#### 4.8.2 Analisis Perbandingan Algoritma Klasifikasi

Subbab ini menjelaskan pemilihan algoritma klasifikasi pada masing – masing dataset penyakit dengan melihat aspek - aspek evaluasi dari performa tiap algoritma. Aspek evaluasi yang digunakan pada penelitian ini adalah nilai akurasi, *error rate*, sensitivitas (*recall*), *specificity*, dan presisi (*precision*) yang akan diuji terhadap ketiga algoritma, yaitu *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Naïve Bayes* (NB). Subbab ini akan menjelaskan seluruh aspek - aspek evaluasi yang telah disebutkan pada performa dari ketiga algoritma dalam mendiagnosis penyakit.

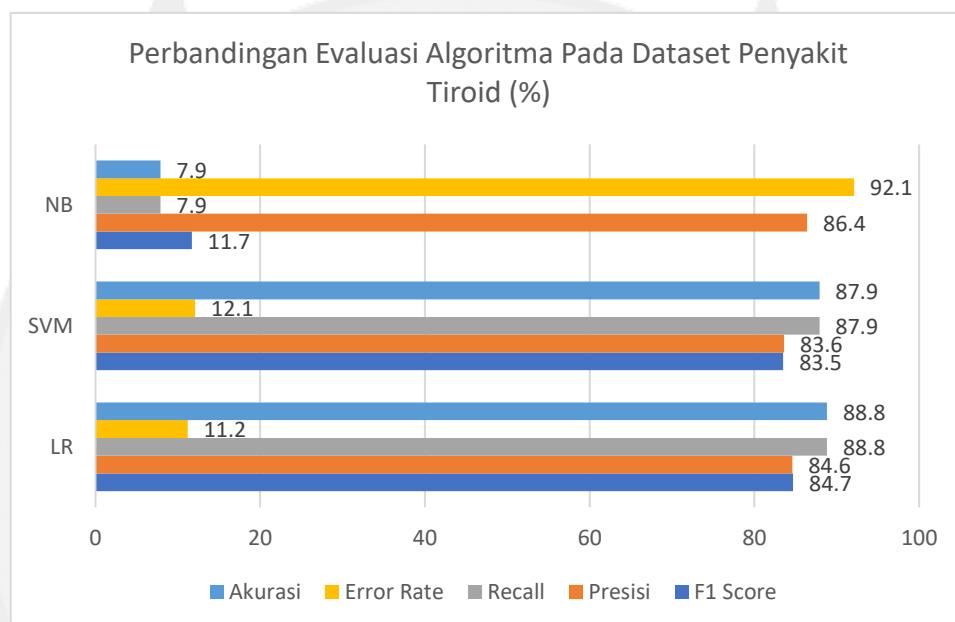


Gambar 4.19 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Obesitas

Gambar 4.19 menunjukkan perbandingan hasil evaluasi algoritma *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Naïve Bayes* (NB) pada dataset penyakit obesitas. Hasil evaluasi dari performa algoritma berdasarkan gambar 4.18 adalah algoritma *Logistic Regression* (LR) menjadi algoritma yang paling baik untuk digunakan dengan nilai akurasi 89,4%, *error rate* 10,6%, presisi 89,2%, *recall* 89,4%, dan *F<sub>1</sub> Score* 89,3%. Algoritma *Support Vector Machine* (SVM) menjadi algoritma kedua terbaik dengan nilai akurasi 78,4%, *error rate* 21,6%, presisi 77,5%, *recall* 78,4%, dan *F<sub>1</sub> Score* 78,4%. Algoritma *Naïve Bayes* (NB) mendapatkan nilai evaluasi terendah dengan nilai akurasi 59,5%, *error rate* 40,5%, presisi 59%, *recall* 59,5%, dan *F<sub>1</sub> Score* 53,3%. Berdasarkan hasil ini maka model sistem pakar yang dibuat akan menggunakan algoritma *Logistic Regression* untuk mendiagnosa penyakit obesitas.

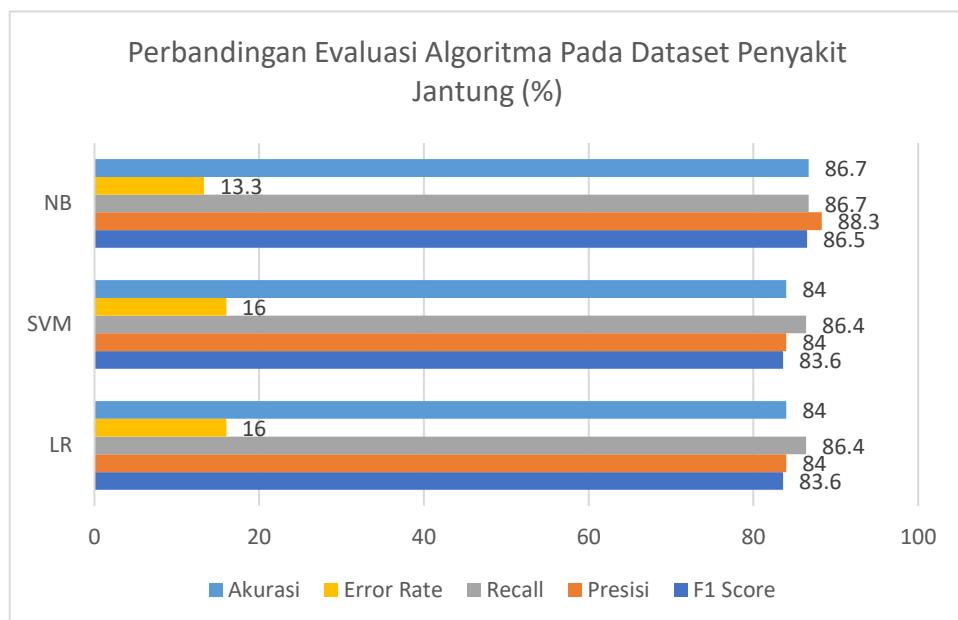
Gambar 4.20 menunjukkan perbandingan hasil evaluasi algoritma *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Naïve Bayes* (NB) pada dataset penyakit tiroid. Hasil evaluasi dari performa algoritma berdasarkan gambar 4.19 adalah algoritma *Logistic Regression* (LR) menjadi algoritma yang paling baik untuk digunakan dengan nilai akurasi 88,8%, *error rate* 11,2%, presisi 84,6%, *recall* 88,8%, dan *F<sub>1</sub> Score* 84,7%. Algoritma *Support Vector Machine* (SVM)

menjadi algoritma kedua terbaik dengan nilai akurasi 87,9%, *error rate* 12,1%, presisi 83,6%, *recall* 87,9%, dan *F<sub>1</sub> Score* 83,5%. Algoritma *Naïve Bayes* (NB) mendapatkan nilai evaluasi terendah dengan nilai akurasi 7,9%, *error rate* 92,1%, presisi 86,4%, *recall* 7,9%, dan *F<sub>1</sub> Score* 11,7%. Berdasarkan hasil ini maka model sistem pakar yang dibuat akan menggunakan algoritma *Logistic Regression* untuk mendiagnosis penyakit tiroid.



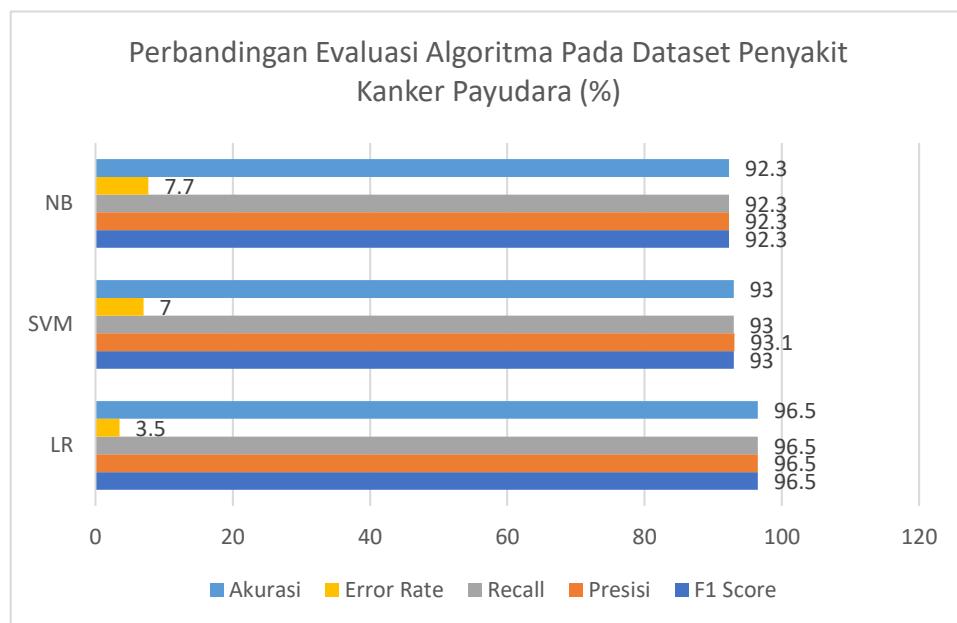
Gambar 4.20 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Tiroid

Gambar 4.21 menunjukkan perbandingan hasil evaluasi algoritma *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Naïve Bayes* (NB) pada dataset penyakit jantung. Hasil evaluasi dari performa algoritma berdasarkan gambar 4.20 adalah algoritma *Naïve Bayes* (NB) menjadi algoritma yang paling baik untuk digunakan dengan nilai akurasi 86,7%, *error rate* 13,3%, presisi 88,3%, *recall* 86,7%, dan *F<sub>1</sub> Score* 86,5%. Algoritma *Support Vector Machine* (SVM) dan *Logistic Regression* memiliki hasil evaluasi yang sama dengan nilai akurasi 84,0%, *error rate* 16,0%, presisi 86,4%, *recall* 84%, dan *F<sub>1</sub> Score* 83,6%. Berdasarkan hasil ini maka model sistem pakar yang dibuat akan menggunakan algoritma *Naïve Bayes* untuk mendiagnosis penyakit jantung.

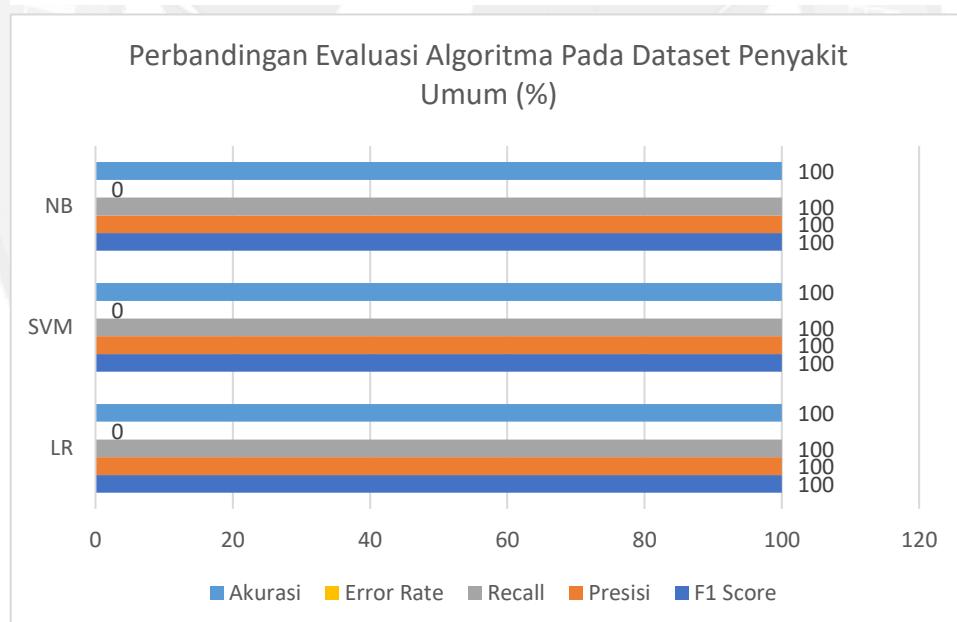


Gambar 4.21 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Jantung

Gambar 4.22 menunjukkan perbandingan hasil evaluasi algoritma *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Naïve Bayes* (NB) pada dataset penyakit kanker payudara. Hasil evaluasi dari performa algoritma berdasarkan gambar 4.21 adalah algoritma *Logistic Regression* (LR) menjadi algoritma yang paling baik untuk digunakan dengan nilai akurasi 96,5%, *error rate* 3,5%, presisi 96,5%, *recall* 96,5%, dan *F1 Score* 96,5%. Algoritma Support Vector Machine (SVM) menjadi algoritma kedua terbaik dengan nilai akurasi 93%, *error rate* 7%, presisi 93,1%, *recall* 93%, dan *F1 Score* 93%. Algoritma *Naïve Bayes* (NB) mendapatkan nilai evaluasi yang hampir sama dengan algoritma *Support Vector Machine* dengan perbedaan hanya pada nilai presisi. Algoritma *Naïve Bayes* memiliki nilai akurasi 92,3%, *error rate* 7,7%, presisi 92,3%, *recall* 92,3%, dan *F1 Score* 92,3%. Berdasarkan hasil ini maka model sistem pakar yang dibuat akan menggunakan algoritma *Logistic Regression* untuk mendiagnosis penyakit kanker payudara.



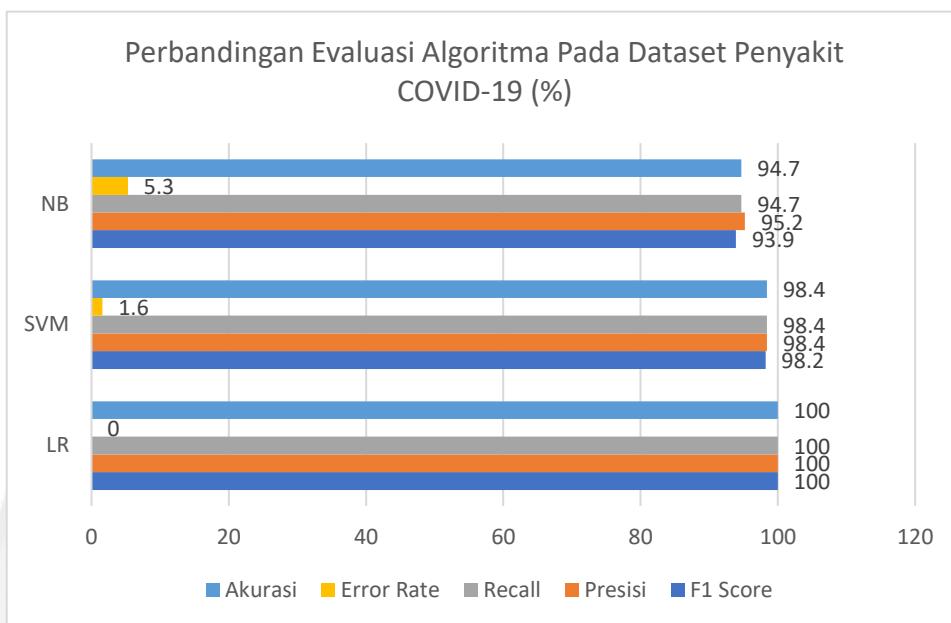
Gambar 4.23 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Kanker Payudara



Gambar 4.22 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit Umum

Gambar 4.23 menunjukkan perbandingan hasil evaluasi algoritma *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Naïve Bayes* (NB) pada dataset penyakit umum. Hasil evaluasi dari performa algoritma berdasarkan gambar 4.22 adalah ketiga algoritma memiliki hasil evaluasi yang sama yaitu nilai akurasi 100%, *error rate* 0%, presisi 100%, *recall* 100%, dan *F<sub>1</sub> Score* 100%. Berdasarkan

hasil ini maka dipilih salah satu algoritma yaitu algoritma *Logistic Regression* yang akan digunakan dalam sistem pakar untuk mendiagnosis penyakit umum.



Gambar 4.24 Perbandingan Hasil Evaluasi Algoritma Pada Dataset Penyakit COVID-19

Gambar 4.24 menunjukkan perbandingan hasil evaluasi algoritma *Logistic Regression* (LR), *Support Vector Machine* (SVM), dan *Naïve Bayes* (NB) pada dataset penyakit COVID-19. Hasil evaluasi dari performa algoritma berdasarkan gambar 4.23 adalah algoritma *Logistic Regression* (LR) menjadi algoritma yang paling baik untuk digunakan dengan nilai akurasi 100%, *error rate* 0%, presisi 100%, *recall* 100%, dan *F<sub>1</sub> Score* 100%. Algoritma *Support Vector Machine* (SVM) menjadi algoritma kedua terbaik dengan nilai akurasi 98,4%, *error rate* 1,6%, presisi 98,4%, *recall* 98,4%, dan *F<sub>1</sub> Score* 98,2%. Algoritma *Naïve Bayes* (NB) mendapatkan nilai evaluasi terendah dengan nilai akurasi 94,7%, *error rate* 5,3%, presisi 95,2%, *recall* 94,7%, dan *F<sub>1</sub> Score* 93,9%. Berdasarkan hasil ini maka model sistem pakar yang dibuat akan menggunakan algoritma *Logistic Regression* untuk mendiagnosis penyakit COVID-19.

#### 4.8.3 Analisa Hasil Penelitian

Pertanyaan penelitian dari penelitian ini adalah: “Bagaimana cara merancang dan mengembangkan sebuah sistem pakar menggunakan metode

klasifikasi *data mining* yang dapat membantu pengguna mengetahui penyakit berdasarkan gejala – gejala yang diketahui?”. Hasil dari penelitian menunjukkan bahwa perancangan sebuah sistem pakar menggunakan metode klasifikasi *data mining* yang dapat memprediksi penyakit berdasarkan gejala – gejala yang diberikan dapat dilakukan menggunakan notasi BPMN dan pengembangan dari sistem pakar dapat dibuat menggunakan bahasa pemrograman *Python*. Diagnosis dengan dibantu oleh komputer merupakan sebuah pendekatan yang berguna untuk meningkatkan akurasi diagnosis dan menggabungkan informasi yang berguna dari seorang ahli manusia [22]. Penggunaan proses *data mining* pada data – data gejala penyakit sudah membuktikan performanya dalam melakukan penyesuaian data – data kosong, yang memang menjadi salah satu karakteristik pada data medis [2]. Selain itu proses *data mining* mampu memproses data gejala penyakit dan menyesuaikannya ke dalam bentuk yang dapat diolah, dan menghasilkan diagnosis berdasarkan data. Oleh sebab itu, perancangan dan pengembangan sistem pakar yang dapat mendiagnosis penyakit sangat terbantu dengan pemanfaatan proses *data mining* untuk menghasilkan diagnosis penyakit.

Penelitian ini menggunakan metode klasifikasi *data mining* untuk mengembangkan *knowledge base* dari sistem pakar yang dibuat. Melalui penelitian yang dilakukan, dapat dilihat bahwa metode klasifikasi *data mining* dapat menjadi *knowledge base* setelah dilakukannya perbandingan hasil evaluasi dari tiga model klasifikasi yang masing – masing model menggunakan algoritma *data mining* yang berbeda. Model terbaik dibuat menjadi sebuah file *classifier* menggunakan *library Joblib* untuk digunakan pada sistem pakar.

Penelitian ini melakukan percobaan dengan membandingkan tiga algoritma klasifikasi (*Logistics Regression*, *Support Vector Machine*, dan *Naïve Bayes*) untuk mendapatkan model klasifikasi yang paling baik dalam menghasilkan diagnosis pada enam tipe penyakit, yaitu penyakit obesitas, penyakit kanker payudara, penyakit jantung, penyakit tiroid, penyakit umum, dan penyakit COVID-19. Ketiga algoritma ini diuji dalam menghasilkan diagnosis dan dievaluasi berdasarkan nilai hasil akurasi, presisi, *recall*, *error rate*, dan *F<sub>1</sub> score*. Berdasarkan hasil penelitian

ini, model klasifikasi yang paling baik digunakan untuk melakukan diagnosis antara lain:

1. Penyakit obesitas menggunakan algoritma *Logistic Regression* yang mendapatkan nilai akurasi 89,4%, nilai *error rate* 10,6%, nilai presisi 89,2%, nilai *recall* 89,4%, dan nilai *F<sub>1</sub> Score* 89,3%.
2. Penyakit tiroid menggunakan algoritma *Logistic Regression* yang mendapatkan nilai akurasi 88,8%, nilai *error rate* 11,2%, nilai presisi 84,6%, nilai *recall* 88,8%, dan nilai *F<sub>1</sub> Score* 84,7%.
3. Penyakit jantung menggunakan algoritma *Naïve Bayes* yang mendapatkan nilai akurasi 86,7%, nilai *error rate* 13,3%, nilai presisi 88,3%, nilai *recall* 86,7%, dan nilai *F<sub>1</sub> Score* 86,5%.
4. Penyakit kanker payudara menggunakan algoritma *Logistic Regression* yang mendapatkan nilai akurasi 96,5%, nilai *error rate* 3,5%, nilai presisi 96,5%, nilai *recall* 96,5%, dan nilai *F<sub>1</sub> Score* 96,5%.
5. Penyakit umum menggunakan algoritma *Logistic Regression* yang mendapatkan nilai akurasi 100%, nilai *error rate* 0%, nilai presisi 100%, nilai *recall* 100%, dan nilai *F<sub>1</sub> Score* 100%.
6. Penyakit COVID-19 menggunakan algoritma *Logistic Regression* yang mendapatkan nilai akurasi 100%, nilai *error rate* 0%, nilai presisi 100%, nilai *recall* 100%, dan nilai *F<sub>1</sub> Score* 100%.

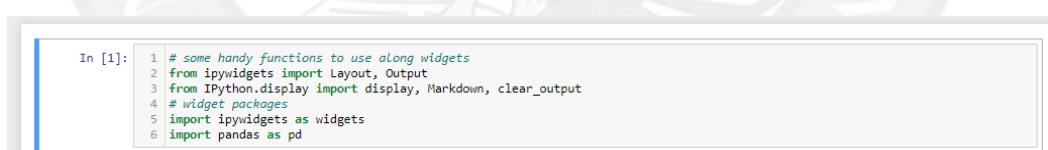
Menggunakan algoritma – algoritma dengan hasil nilai evaluasi terbaik, penelitian ini membentuk sebuah model sistem pakar (*expert system*) yang dapat mendiagnosis penyakit – penyakit yang telah diteliti. Salah satu keuntungan dari penggunaan sistem pakar dalam mendeteksi atau mendiagnosis penyakit adalah sistem pakar dapat mendeteksi penyakit secara cepat berdasarkan data gejala yang diberikan. Pada penelitian ini, peneliti mencatat waktu yang dibutuhkan pengguna untuk mengisi data gejala dan waktu yang dibutuhkan oleh sistem pakar dalam menghasilkan diagnosis. Berdasarkan pengukuran yang dilakukan, didapatkan hasil sebagai berikut:

1. Pengguna membutuhkan waktu sekitar 1 menit 4 detik untuk mengisi data penyakit obesitas dan sistem pakar penyakit obesitas membutuhkan waktu kurang lebih 9 detik dalam menghasilkan diagnosis.
2. Pengguna membutuhkan waktu sekitar 1 menit 13 detik untuk mengisi data penyakit tiroid dan sistem pakar penyakit tiroid membutuhkan waktu kurang lebih 9 detik dalam menghasilkan diagnosis.
3. Pengguna membutuhkan waktu sekitar 1 menit 7 detik untuk mengisi data penyakit jantung dan sistem pakar penyakit jantung membutuhkan waktu kurang lebih 10 detik dalam menghasilkan diagnosis.
4. Pengguna membutuhkan waktu sekitar 1 menit 4 detik untuk mengisi data penyakit kanker payudara dan sistem pakar penyakit kanker payudara membutuhkan waktu kurang lebih 10 detik dalam menghasilkan diagnosis.
5. Pengguna membutuhkan waktu sekitar 1 menit 33 detik untuk mengisi data penyakit umum dan sistem pakar penyakit umum membutuhkan waktu kurang lebih 10 detik dalam menghasilkan diagnosis.
6. Pengguna membutuhkan waktu sekitar 1 menit untuk mengisi data penyakit COVID-19 dan sistem pakar penyakit COVID-19 membutuhkan waktu kurang lebih 16 detik dalam menghasilkan diagnosis.

Berdasarkan penghitungan waktu yang dilakukan, dapat dilihat bahwa sistem pakar dapat membantu pengguna dalam menghasilkan diagnosis dengan cepat. Hasil perhitungan waktu diagnosis yang didapatkan memperlihatkan seluruh penyakit memerlukan waktu kurang dari 20 detik untuk menghasilkan diagnosis. Cepatnya proses diagnosis dalam membantu pengguna dalam membuat suatu keputusan yang bersifat medis, seperti apakah ia harus pergi ke klinik atau rumah sakit untuk diperiksa, atau pergi ke apotik untuk membeli obat. Sistem pakar juga telah menggunakan algoritma yang memiliki nilai evaluasi yang baik sehingga hasil diagnosis dapat dipercaya. Melalui penggunaan sistem pakar, pasien dapat mengetahui penyakit yang mungkin dideritanya dan menyampaikannya kepada petugas medis sehingga pemeriksaan mengenai potensi penyakit yang diderita oleh pasien dapat dilakukan dengan lebih cepat dan terarah.

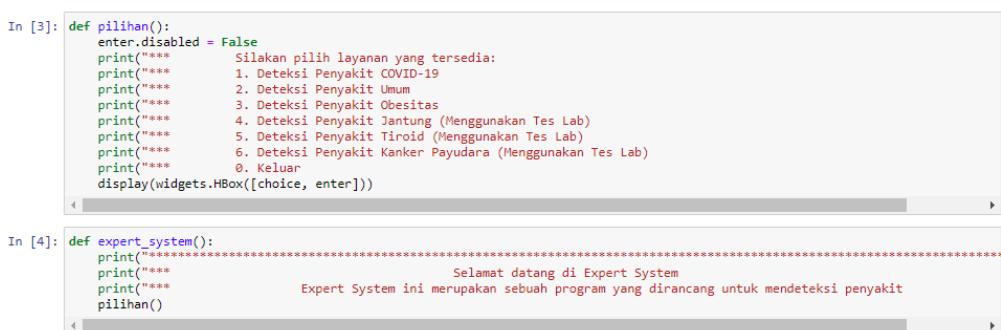
## 4.9 Tampilan Sistem Pakar (*Expert System*)

Penelitian ini membuat sebuah model sistem pakar yang dapat mengolah input gejala dari pengguna hingga menjadi sebuah diagnosis penyakit. Model sistem pakar dibuat dalam bentuk file *Python* yang dijalankan dalam *Jupyter Notebook*. Model dibuat berdasarkan gambar 3.6 pada subab 3.6 tentang alur model sistem pakar. Pengguna akan diberikan tampilan menu dimana pengguna dapat memilih 1 diantara enam penyakit dalam sistem pakar. Pengguna lalu memasukkan data gejala mereka, dan mengklik *button* untuk menghasilkan diagnosis. Setelah *button* diklik, pengguna akan diminta untuk menunggu sementara sistem pakar memberitahu progres dari pengolahan data. Diagnosis akan diberikan saat seluruh proses pengolahan telah dilakukan. Pengguna dapat memilih untuk menggunakan kembali sistem pakar dengan memilih kembali penyakit pada menu atau menghentikan program sistem pakar. Masing – masing tipe penyakit memiliki file untuk menginisialisasi tampilan dan mengolah data gejala untuk menjadi sebuah diagnosis. Terdapat satu file yang berfungsi sebagai pengolah tampilan utama sistem pakar dan satu file sebagai *launcher* atau program yang bertugas untuk menjalankan sistem pakar dari awal hingga akhir. Gambar 4.25 sampai 4.29 menunjukkan baris kode dan tampilan awal dari sistem pakar.



```
In [1]: 1 # some handy functions to use along widgets
2 from ipywidgets import Layout, Output
3 from IPython.display import display, Markdown, clear_output
4 # widget packages
5 import ipywidgets as widgets
6 import pandas as pd
```

Gambar 4.25 Baris Kode *Import Ipywidgets* untuk Tampilan Awal Sistem Pakar



```
In [3]: def pilihan():
    enter.disabled = False
    print("**** Silakan pilih layanan yang tersedia:")
    print("**** 1. Deteksi Penyakit COVID-19")
    print("**** 2. Deteksi Penyakit Umum")
    print("**** 3. Deteksi Penyakit Obesitas")
    print("**** 4. Deteksi Penyakit Jantung (Menggunakan Tes Lab)")
    print("**** 5. Deteksi Penyakit Tiroid (Menggunakan Tes Lab)")
    print("**** 6. Deteksi Penyakit Kanker Payudara (Menggunakan Tes Lab)")
    print("**** 8. Keluar")
    display(widgets.HBox([choice, enter]))
```

```
In [4]: def expert_system():
    print("*****")
    print("***** Selamat datang di Expert System")
    print("***** Expert System ini merupakan sebuah program yang dirancang untuk mendeteksi penyakit")
    print("*****")
    pilihan()
```

Gambar 4.26 Baris Kode Tampilan Awal Sistem Pakar

```

In [2]: 1 choice = widgets.IntText(
2     value=1,
3     description="Pilihan anda:",
4     disabled=False
5 )
6 enter = widgets.Button(
7     description="Masukan Pilihan",
8     disabled=False,
9     button_style='success', # 'success', 'info', 'warning', 'danger' or ''
10    tooltip='Masukan Pilihan',
11    icon='check', # (FontAwesome names without the 'fa-' prefix)
12    layout=Layout(width='30%', height='30px')
13 )
14 ya = widgets.Button(
15     description='Ya',
16     disabled=False,
17     button_style='success', # 'success', 'info', 'warning', 'danger' or ''
18     tooltip='Ya',
19     icon='check', # (FontAwesome names without the 'fa-' prefix)
20     layout=Layout(width='50%', height='40px')
21 )
22 tidak = widgets.Button(
23     description='Tidak',
24     disabled=False,
25     button_style='danger', # 'success', 'info', 'warning', 'danger' or ''
26     tooltip='Tidak',
27     icon='times', # (FontAwesome names without the 'fa-' prefix)
28     layout=Layout(width='50%', height='40px')
29 )
30 output = widgets.Output()
31 def on_button_clicked(b):
32     clear_output()
33     enter.disabled = True
34     if choice.value == 1:
35         #run ExpertSystem_COVID19.ipynb
36     elif choice.value == 2:
37         #run ExpertSystem_General.ipynb
38     elif choice.value == 3:
39         #run ExpertSystem_Obesity.ipynb
40     elif choice.value == 4:
41         #run ExpertSystem_HeartDisease.ipynb
42     elif choice.value == 5:
43         #run ExpertSystem_Thyroid.ipynb
44     elif choice.value == 6:
45         #run ExpertSystem_BreastCancer.ipynb
46     elif choice.value == 8:
47         print("**** Terima kasih dan sampai jumpa kembali!")
48     else:
49         with output:
50             print("**** Pilihan layanan yang anda pilih tidak tersedia, silakan memilih pilihan layanan lain.")
51             enter.disabled = False
52             pilahan()
53
54 enter.on_click(on_button_clicked)

```

Gambar 4.27 Baris Menginisialisasi *Button*, *Inttext*, dan Fungsi Saat *Button* Diklik dalam Menu Sistem Pakar

```
In [ ]: #run EXPERT_SYSTEM.ipynb
expert_system()
```

Gambar 4.28 Baris Kode untuk Menjalankan Sistem Pakar

```
*****
*** Selamat datang di Expert System
*** Expert System ini merupakan sebuah program yang dirancang untuk mendeteksi penyakit
*** Silakan pilih layanan yang tersedia:
*** 1. Deteksi Penyakit COVID-19
*** 2. Deteksi Penyakit Umum
*** 3. Deteksi Penyakit Obesitas
*** 4. Deteksi Penyakit Jantung (Menggunakan Tes Lab)
*** 5. Deteksi Penyakit Tiroid (Menggunakan Tes Lab)
*** 6. Deteksi Penyakit Kanker Payudara (Menggunakan Tes Lab)
*** 0. Keluar
***
```

Pilihan anda:  ✓ Masukan Pilihan

Gambar 4.29 Tampilan Awal Sistem Pakar

#### 4.9.1 Tampilan Sistem Pakar Penyakit Obesitas

Subbab ini menjelaskan tampilan sistem pakar saat pengguna memilih menu penyakit obesitas. Sistem pakar penyakit obesitas menggunakan *ipywidgets IntText*, *FloatText*, dan *RadioButton* dalam menerima data gejala dari pengguna. *Ipywidget Button* digunakan untuk menjalankan fungsi yang akan mengolah data gejala yang sudah diisi menjadi diagnosis penyakit obesitas. Selain itu juga *button* menjalankan fungsi untuk membuat animasi *loading* untuk memberikan informasi kepada pengguna progres dari pengolahan data. Subbab ini juga menampilkan baris kode penggunaan *classifier* atau model yang menjadi *knowledge base* dari sistem pakar untuk penyakit obesitas. Gambar 4.30 hingga 4.35 menunjukkan baris kode penggunaan *classifier* dan tampilan sistem pakar saat mendiagnosis penyakit obesitas.

```
In [14]: 1 def classifier_obesity(data):
2     import itertools
3     import threading
4     import time
5     import sys
6
7     done = False
8     #here is the animation
9     def animate():
10         for c in itertools.cycle(['|', '/', '-', '\\']):
11             if done:
12                 break
13             sys.stdout.write('\r*** Mohon Menunggu... ' + c + ' ')
14             sys.stdout.flush()
15             time.sleep(0.1)
16             sys.stdout.write('\r*** Berhasil!')
17
18
19     ##Feature scaling untuk menyetarakan kolom yang memiliki bermacam-macam range pengukuran
20     print("\n*** Menjalankan feature scaling...")
21     done = False
22     ob4 = threading.Thread(target=animate)
23     ob4.start()
24     from sklearn.preprocessing import StandardScaler
25     sc = StandardScaler()
26     data = sc.fit_transform(data)
27     done = True
28
29     time.sleep(3)
30     ##Buat classifier algoritma
31     print("\n*** Membuat model klasifikasi algoritma dan memasukkan data ke dalam model...")
32     done = False
33     ob5 = threading.Thread(target=animate)
34     ob5.start()
35     import joblib
36     classifier_obesity = joblib.load("classifier_obesity.pkl")
37     done = True
38
39     ##menghasilkan diagnosis
40     time.sleep(3)
41     print("\n*** Menghasilkan diagnosis...")
42     done = False
43     ob6 = threading.Thread(target=animate)
44     ob6.start()
45     prediksi = classifier_obesity.predict(data)
46     done = True
47     time.sleep(3)
48     return prediksi
```

Gambar 4.30 Baris Kode Penggunaan *Classifier* Sebagai *Knowledge Base* Sistem Pakar Penyakit Obesitas

\*\*\*\*\* Expert System Penyakit Obesitas \*\*\*\*\*

\*\*\* Ini adalah sebuah expert system yang dapat digunakan untuk mendeteksi potensi terjangkitnya penyakit obesitas \*\*\*

\*\*\* Harap diperhatikan bahwa diagnosis yang dihasilkan sistem ini hanya bersifat sebagai nasihat medis \*\*\*

\*\*\* Periksa diri anda ke klinik/puskesmas/rumah sakit terdekat berdasarkan hasil diagnosis yang anda dapatkan \*\*\*

\*\*\* Tolong berikan jawaban yang sejujur-jujurnya agar hasil diagnosis dari sistem dapat seakurat mungkin \*\*\*

\*\*\* Silakan isi umur anda dalam angka (tahun). \*\*\*

\*\*\* Umur

\*\*\* Silakan pilih kategori jenis kelamin anda. \*\*\*

\*\*\*  Laki-Laki  
 Perempuan

\*\*\* Silakan isi tinggi badan anda. \*\*\*

\*\*\* Tinggi Badan

\*\*\* Silakan isi berat badan anda. \*\*\*

\*\*\* Berat Badan

\*\*\* Apakah keluarga anda memiliki riwayat obesitas? \*\*\*

\*\*\*  ya  
 tidak

\*\*\* Apakah anda sering memakan makanan berkalori tinggi? \*\*\*

\*\*\*  ya  
 tidak

\*\*\*

Gambar 4.31 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Obesitas Bagian I

\*\*\* Seberapa sering anda memakan sayuran dalam hidangan anda? \*\*\*

\*\*\*  Tidak Pernah  
 Kadang-Kadang  
 Selalu

\*\*\* Berapa banyak hidangan utama (sarapan, makan siang, dll) yang anda makan dalam sehari? \*\*\*

\*\*\*  Satu kali  
 Dua kali  
 Tiga kali  
 Lebih dari tiga kali

\*\*\* Diantara waktu antar hidangan (sarapan ke makan siang, makan siang ke makan malam, dll), seberapa sering anda memakan sesuatu dalam jangka waktu tersebut? \*\*\*

\*\*\*  Selalu  
 Sering  
 Kadang-Kadang

\*\*\* Apakah anda merokok? \*\*\*

\*\*\*  ya  
 tidak

Gambar 4.32 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Obesitas Bagian II

\*\*\* Seberapa banyak air yang anda minum tiap harinya?

\*\*\*  Kurang dari satu liter  
 Antara satu sampai dua liter  
 Lebih dari dua liter

\*\*\* Apakah anda memperhatikan kalori yang anda makan?

\*\*\*  ya  
 tidak

\*\*\* Seberapa sering anda berolahraga atau melakukan aktivitas fisik?

\*\*\*  Tidak pernah  
 1 atau 2 hari  
 2 atau 4 hari  
 4 atau 5 hari

\*\*\* Seberapa sering anda meminum alkohol?

\*\*\*  Selalu  
 Sering  
 Kadang - Kadang  
 Tidak minum

\*\*\* Transportasi apa yang biasanya anda gunakan?

\*\*\*  Mobil  
 Sepeda  
 Motor  
 Kendaraan Umum  
 Jalan Kaki

\*\*\* Klik Tombol ini Untuk Melihat Hasil Diagnosis

Gambar 4.33 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Obesitas Bagian III

\*\*\* Kadang - Kadang

\*\*\* Transportasi apa yang biasanya anda gunakan?

\*\*\*  Mobil  
 Sepeda  
 Motor  
 Kendaraan Umum  
 Jalan Kaki

\*\*\* Klik Tombol ini Untuk Melihat Hasil Diagnosis

\*\*\* Baik, terima kasih atas kerjasama anda. Sekarang kami akan memproses data anda dan menampilkan hasil diagnosis anda. \*\*\*  
\*\*\* Memuka file data training...  
\*\*\* Berhasil!  
\*\*\* Mengubah kolom - kolom bernilai kategorial...  
\*\*\* Berhasil!  
\*\*\* Hemisahkan kolom gejala dan kolom diagnosis...  
\*\*\* Berhasil!

Gambar 4.34 Tampilan Saat Menunggu Diagnosis Penyakit Obesitas Dihasilkan

\*\*\* Terima kasih atas perhatian anda. Melalui gejala dan data yang dikumpulkan,  
\*\*\* anda berada pada kategori KEGEMUKAN TINGKAT I.  
\*\*\* Anda dapat berkonsultasi kepada ahli gizi untuk meningkatkan berat badan anda.  
\*\*\* Terima kasih anda telah menggunakan layanan expert system, apakah anda ingin menggunakan layanan ini kembali?

✓ Ya      ✗ Tidak

Gambar 4.35 Tampilan Saat Diagnosis Penyakit Obesitas Dihasilkan

## 4.9.2 Tampilan Sistem Pakar Penyakit Tiroid

Subbab ini menjelaskan tampilan sistem pakar saat pengguna memilih menu penyakit tiroid. *Expert system* ini membutuhkan data – data pengukuran pada laboratorium seperti pemeriksaan T3, TT4, T4U, TSH, FTI, dan TBG. *Expert system* penyakit tiroid menggunakan *ipywidgets IntText*, *FloatText*, dan *RadioButton* dalam menerima data gejala dari pengguna. *IntText* dan *FloatText* digunakan untuk data – data pengukuran laboratorium, sedangkan *RadioButton* berfungsi untuk mencatat data yang memiliki nilai ya dan tidak. *Ipywidget Button* digunakan untuk menjalankan fungsi yang akan mengolah data gejala yang sudah diisi menjadi diagnosis penyakit tiroid. Selain itu juga *button* menjalankan fungsi untuk membuat animasi *loading* untuk memberikan informasi kepada pengguna progres dari pengolahan data. Subbab ini juga menampilkan baris kode penggunaan *classifier* atau model yang menjadi *knowledge base* dari sistem pakar untuk penyakit tiroid. Gambar 4.36 hingga 4.42 menunjukkan baris kode penggunaan *classifier* dan tampilan sistem pakar saat mendiagnosis penyakit tiroid.

```
In [5]: 1 def classifier_thyroid(data):
2     import itertools
3     import threading
4     import time
5     import sys
6
7     done = False
8     #here is the animation
9     def animate():
10         for c in itertools.cycle(['|', '/', '-', '\\']):
11             if done:
12                 break
13             sys.stdout.write("\r*** Mohon Menunggu... " + c + " ")
14             sys.stdout.flush()
15             time.sleep(0.1)
16     sys.stdout.write("\r*** Berhasil!")
17
18     ##Feature scaling untuk menyetarakan kolom yang memiliki bermacam-macam range pengukuran
19     print("\n*** Menjalankan feature scaling...")
20     done = False
21     th4 = threading.Thread(target=animate)
22     th4.start()
23     from sklearn.preprocessing import StandardScaler
24     sc = StandardScaler()
25     data = sc.fit_transform(data)
26     done = True
27
28     time.sleep(3)
29     ##Buat classifier algoritma
30     print("\n*** Membuka model klasifikasi algoritma dan memasukkan data ke dalam model...")
31     done = False
32     th5 = threading.Thread(target=animate)
33     th5.start()
34     import joblib
35     classifier_thyroid = joblib.load("classifier_thyroid.pkl")
36     done = True
37
38     ##menghasilkan diagnosis
39     time.sleep(3)
40     print("\n*** Menghasilkan diagnosis...")
41     done = False
42     th6 = threading.Thread(target=animate)
43     th6.start()
44     prediksi = classifier_thyroid.predict(data)
45     done = True
46     time.sleep(3)
47     return prediksi
```

Gambar 4.36 Baris Kode Penggunaan *Classifier* Sebagai *Knowledge Base* Sistem Pakar Penyakit Tiroid

\*\*\*\*\*
 \*\*\*\* Expert System Penyakit Tiroid
 \*\*\*\* Ini adalah sebuah expert system yang dapat digunakan untuk mendeteksi potensi terjangkitnya penyakit tiroid
 \*\*\*\* Harap diperhatikan bahwa diagnosis yang dihasilkan sistem ini hanya bersifat sebagai nasihat medis
 \*\*\*\* Expert System ini lebih diperuntukan untuk tenaga medis dikarenakan banyak
 \*\*\*\* membutuhkan data yang berupa pengujian laboratorium dan tes terkait tiroid
 \*\*\*\* Tolong berikan jawaban yang sejajar-jujurnya agar hasil diagnosis dari sistem dapat seakurat mungkin
 \*\*\*\* Silakan isi umur pasien dalam angka (tahun).

\*\*\* Umur

\*\*\*-----  
\*\*\* Silakan pilih kategori jenis kelamin pasien.

\*\*\*    Laki-Laki  
\*\*\*    Perempuan

\*\*\*-----  
\*\*\* Apakah pasien mengkonsumsi obat thyroxine?.

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien mendapatkan preskripsi/resep obat thyroxine?

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien mengkonsumsi obat antitiroid?

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien merasa ingin muntah?

\*\*\*    ya  
\*\*\*    tidak

Gambar 4.37 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Tiroid Bagian I

\*\*\*-----  
\*\*\* Apakah pasien saat ini hamil? (\*khusus perempuan, untuk laki - laki silakan hiraukan pertanyaan ini)

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien pernah melakukan operasi terkait tiroid?

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien sedang menjalani terapi I-131 (Iodine-131)?

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien pernah menjalani pemeriksaan terkait hipotiroid?

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien pernah menjalani pemeriksaan terkait hipertiroid?

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien mengkonsumsi obat litium?

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien memiliki benjolan di leher?

\*\*\*    ya  
\*\*\*    tidak

\*\*\*-----  
\*\*\* Apakah pasien pernah melakukan pengecekan tumor?

\*\*\*    ya  
\*\*\*    tidak

Gambar 4.38 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Tiroid Bagian II

\*\*\*  
\*\*\* Apakah pasien pernah melakukan pengecekan kondisi hipopituitar? \*\*\*

\*\*\*    ya  
      tidak

\*\*\*  
\*\*\* Apakah pasien pernah melakukan pengecekan kondisi psikis? \*\*\*

\*\*\*    ya  
      tidak

\*\*\*  
\*\*\* Apakah pasien pernah melakukan pengecekan tingkat TSH? \*\*\*

\*\*\*    ya  
      tidak

\*\*\*  
\*\*\* Silakan isi tingkat TSH pasien. Jika pasien tidak pernah melakukan pengecekan, silakan isi dengan nilai 0. \*\*\*

\*\*\*   TSH

\*\*\*  
\*\*\* Apakah pasien pernah melakukan pengecekan tingkat T3? \*\*\*

\*\*\*    ya  
      tidak

\*\*\*  
\*\*\* Silakan isi tingkat T3 pasien. Jika pasien tidak pernah melakukan pengecekan, silakan isi dengan nilai 0. \*\*\*

\*\*\*   T3

\*\*\*  
\*\*\* Apakah pasien pernah melakukan pengecekan tingkat TT4? \*\*\*

\*\*\*    ya  
      tidak

\*\*\*  
\*\*\* Silakan isi tingkat TT4 pasien. Jika pasien tidak pernah melakukan pengecekan, silakan isi dengan nilai 0. \*\*\*

\*\*\*   TT4

Gambar 4.39 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Tiroid Bagian III

\*\*\*  
\*\*\* Apakah pasien pernah melakukan pengecekan tingkat T4U? \*\*\*

\*\*\*    ya  
      tidak

\*\*\*  
\*\*\* Silakan isi tingkat T4U pasien. Jika pasien tidak pernah melakukan pengecekan, silakan isi dengan nilai 0. \*\*\*

\*\*\*   T4U

\*\*\*  
\*\*\* Apakah pasien pernah melakukan pengecekan tingkat FTI? \*\*\*

\*\*\*    ya  
      tidak

\*\*\*  
\*\*\* Silakan isi tingkat FTI pasien. Jika pasien tidak pernah melakukan pengecekan, silakan isi dengan nilai 0. \*\*\*

\*\*\*   FTI

\*\*\*  
\*\*\* Apakah pasien pernah melakukan pengecekan tingkat TBG? \*\*\*

\*\*\*    ya  
      tidak

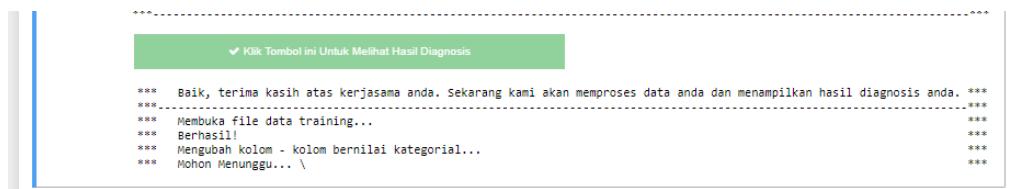
\*\*\*  
\*\*\* Silakan isi tingkat TBG pasien. Jika pasien tidak pernah melakukan pengecekan, silakan isi dengan nilai 0. \*\*\*

\*\*\*   TBG

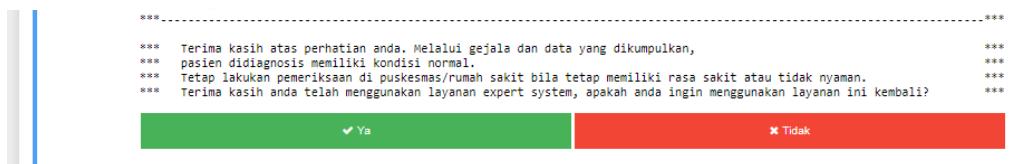
\*\*\*

**✓ Klik Tombol ini Untuk Melihat Hasil Diagnosis**

Gambar 4.40 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Tiroid Bagian IV



Gambar 4.41 Tampilan Saat Menunggu Diagnosis Penyakit Tiroid Dihasilkan



Gambar 4.42 Tampilan Saat Diagnosis Penyakit Tiroid Dihasilkan

#### 4.9.3 Tampilan Sistem Pakar Penyakit Jantung

Subbab ini menjelaskan tampilan sistem pakar saat pengguna memilih menu penyakit jantung. Sistem pakar (*expert system*) ini membutuhkan data – data pengukuran pada laboratorium seperti pemeriksaan elektrokardiografi, gula darah, detak jantung, dan pemeriksaan *flourescopy*. Sistem pakar penyakit jantung menggunakan *ipywidgets IntText*, *FloatText*, dan *RadioButton* dalam menerima data gejala dari pengguna. *IntText* dan *FloatText* digunakan untuk data – data tekanan darah, kolesterol, detak jantung, dan penurunan nilai segmen ST, sedangkan *RadioButton* berfungsi untuk menjawab pertanyaan yang memiliki nilai jawaban ya dan tidak, seperti contoh “Apakah pasien memiliki gula darah lebih dari 120 mmHg?”. *Ipywidget Button* digunakan untuk menjalankan fungsi yang akan mengolah data gejala yang sudah diisi menjadi diagnosis penyakit tiroid. Selain itu juga *button* menjalankan fungsi untuk membuat animasi *loading* untuk memberikan informasi kepada pengguna progres dari pengolahan data. Subbab ini juga menampilkan baris kode penggunaan *classifier* atau model yang menjadi *knowledge base* dari sistem pakar untuk penyakit jantung. Gambar 4.43 hingga 4.48 menunjukkan baris kode penggunaan *classifier* dan tampilan sistem pakar saat mendiagnosis penyakit jantung.

```

In [5]: 1 def classifier_heart_disease(data):
2     import itertools
3     import threading
4     import time
5     import sys
6
7     done = False
8     #here is the animation
9     def animate():
10         for c in itertools.cycle(['|', '/', '-', '\\']):
11             if done:
12                 break
13             sys.stdout.write('\r*** Mohon Menunggu... ' + c + ' ')
14             sys.stdout.flush()
15             time.sleep(0.1)
16         sys.stdout.write('\r*** Berhasil!')
17
18     ##Feature scaling untuk menyetarakan kolom yang memiliki bermacam-macam range pengukuran
19     print("\n*** Melakukan feature scaling...")
20     done = False
21     hd3 = threading.Thread(target=animate)
22     hd3.start()
23     from sklearn.preprocessing import StandardScaler
24     sc = StandardScaler()
25     data = sc.fit_transform(data)
26     done = True
27
28     time.sleep(3)
29     ##Buat classifier algoritma
30     print("\n*** Membuka model klasifikasi algoritma dan memasukkan data ke dalam model...")
31     done = False
32     hd4 = threading.Thread(target=animate)
33     hd4.start()
34     import joblib
35     classifier_heart_disease = joblib.load("classifier_heart_disease.pkl")
36     done = True
37
38     ##menghasilkan diagnosis
39     time.sleep(3)
40     print("\n*** Menghasilkan diagnosis...")
41     done = False
42     hd5 = threading.Thread(target=animate)
43     hd5.start()
44     prediksi = classifier_heart_disease.predict(data)
45     done = True
46     time.sleep(3)
47     return prediksi

```

Gambar 4.43 Baris Kode Penggunaan *Classifier* Sebagai *Knowledge Base* Sistem Pakar Penyakit Jantung

\*\*\*\*\*
 \* Expert System Penyakit Jantung
 \* Ini adalah sebuah expert system yang dapat digunakan untuk mendeteksi potensi terjangkitnya penyakit jantung
 \* Harap diperhatikan bahwa diagnosis yang dihasilkan sistem ini hanya bersifat sebagai nasihat medis
 \* Expert System ini lebih diperuntukan untuk tenaga medis dikarenakan banyak
 \* membutuhkan data yang berupa pengujian laboratorium dan tes terkait jantung
 \* Tolong berikan jawaban yang sejujur-jujurnya agar hasil diagnosis dari sistem dapat seakurat mungkin
 \* Silakan isi umur pasien dalam angka (tahun).

Umur

\*\*\*-----  
\*\*\* Silakan pilih kategori jenis kelamin pasien.

Laki-Laki  
 Perempuan

\*\*\*-----  
\*\*\* Nyeri dada adalah (1) rasa sakit atau ketidaknyamanan pada bagian diatas tulang dada yang (2) disebabkan oleh stres \*\*\*  
\*\*\* dan kelelahan dan (3) dapat diatasi dengan istirahat dan nitroglycerin atau keduanya  
\*\*\* Silakan pilih kategori nyeri dada pasien.

typical (memenuhi semua kriteria diatas)  
 atypical (memenuhi 2 dari 3 kriteria diatas)  
 non-anginal (hanya memenuhi 1 atau kurang kriteria)  
 asymptomatic (nyeri dada terjadi secara tiba-tiba)

\*\*\*-----  
\*\*\* Silakan isi hasil pengukuran tekanan darah pasien (dalam satuan mmHg, angka bulat).

Tekanan Darah

\*\*\*-----  
\*\*\* Silakan isi hasil pengukuran kolesterol pasien (dalam satuan mg/dl, angka bulat).

Kolesterol

Gambar 4.44 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Jantung Bagian I

\*\*\*  
\*\*\* Lakukan pengukuran tingkat gula darah setelah pasien berpuasa 8 jam.  
\*\*\* Apakah tingkat gula darah pasien melebihi 120 mg/dl?  
\*\*\*

ya  
 tidak

\*\*\*  
\*\*\* Silakan isi detak jantung maksimal pasien.  
\*\*\*

Detak Jantung Maksimal

\*\*\*  
\*\*\* Apakah pasien mengalami nyeri dada saat berolahraga?  
\*\*\*

ya  
 tidak

\*\*\*  
\*\*\* Lakukan tes Elektrokardiografi (EKG/ECG). Pilih hasil yang paling sesuai dengan keadaan pasien.  
\*\*\*

Normal  
 Gelombang ST-T abnormal  
 Menunjukkan kemungkinan/sejaka pasti kondisi hipertropi ventrikular kiri menurut kriteria Estes

\*\*\*  
\*\*\* Silakan isi nilai penurunan segment ST akibat berolahraga. (tanda koma menggunakan tanda baca titik .)  
\*\*\*

Penurunan Segmen ST akibat olahraga

\*\*\*  
\*\*\* Silakan pilih kategori yang sesuai untuk titik naik segment ST.  
\*\*\*

Upsloping/naik  
 Flat  
 Downsloping/turun

\*\*\*  
\*\*\* Lakukan tes flourosopy pada pembuluh darah pasien.  
\*\*\* Silakan pilih jumlah pembuluh darah yang menjadi berwarna berdasarkan hasil tes.  
\*\*\*

0  
 1  
 2  
 3

Gambar 4.45 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Jantung Bagian II

\*\*\*  
\*\*\* Lakukan thallium stress test atau nuclear stress test pada pasien. Silakan pilih kategori hasil yang didapatkan.  
\*\*\*

Normal  
 Fixed Defect (Sel miokardia rusak sehingga tidak menerima injeksi tracer)  
 Reversible Defect (Terdeteksi penyumbatan saat sel miokardia menyerap tracer saat beristirahat)

\*\*\*  
\*\*\*

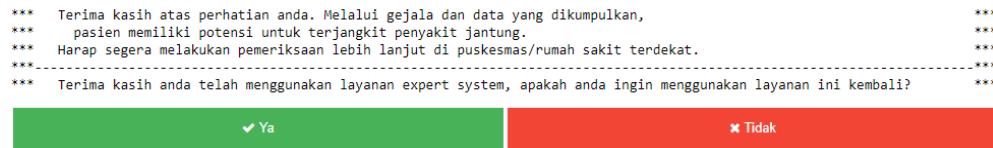


Gambar 4.46 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Jantung Bagian III

\*\*\*  
\*\*\*  
\*\*\* Klik Tombol ini Untuk Melihat Hasil Diagnosis

\*\*\*  
\*\*\* Baik, terima kasih atas kerjasama anda. Sekarang kami akan memproses data anda dan menampilkan hasil diagnosis anda. \*\*\*  
\*\*\*  
\*\*\* Membuka file data training...  
\*\*\* Berhasil!  
\*\*\* Memisahkan kolom gejala dan kolom diagnosis...  
\*\*\* Berhasil!  
\*\*\* Menjalankan feature scaling...  
\*\*\* Berhasil!  
\*\*\* Membuat model klasifikasi algoritma dan memasukkan data ke dalam model...  
\*\*\* Mohon Menunggu... |

Gambar 4.47 Tampilan Saat Menunggu Diagnosis Penyakit Jantung Dihasilkan



Gambar 4.48 Tampilan Saat Diagnosis Penyakit Jantung Dihasilkan

#### 4.9.4 Tampilan Sistem Pakar Penyakit Kanker Payudara

```
In [21]: 1 def classifier_breast_cancer(data):
2     import itertools
3     import threading
4     import time
5     import sys
6
7     done = False
8     #here is the animation
9     def animate():
10         for c in itertools.cycle(['|', '/', '-', '\\']):
11             if done:
12                 break
13             sys.stdout.write('\r*** Mohon Menunggu... ' + c + ' ')
14             sys.stdout.flush()
15             time.sleep(0.1)
16         sys.stdout.write('\r*** Berhasil')
17
18     ##Feature scaling untuk menyetarakan kolom yang memiliki bermacam-macam range pengukuran
19     print("\n*** Menjalankan feature scaling...")
20     done = False
21     bc4 = threading.Thread(target=animate)
22     bc4.start()
23     from sklearn.preprocessing import StandardScaler
24     sc = StandardScaler()
25     data = sc.fit_transform(data)
26     done = True
27
28     time.sleep(3)
29     ##Buat classifier algoritma
30     print("\n*** Melakukan load model klasifikasi algoritma dan memasukkan data ke dalam model...")
31     done = False
32     bc5 = threading.Thread(target=animate)
33     bc5.start()
34     import joblib
35     classifier_breast_cancer = joblib.load("classifier_breast_cancer.pkl")
36     done = True
37
38     ##menghasilkan diagnosis
39     time.sleep(3)
40     print("\n*** Menghasilkan diagnosis...")
41     done = False
42     bc6 = threading.Thread(target=animate)
43     bc6.start()
44     prediksi = classifier_breast_cancer.predict(data)
45     done = True
46     time.sleep(3)
47     return prediksi
```

Gambar 4.49 Baris Kode Penggunaan *Classifier* Sebagai *Knowledge Base* Sistem Pakar Penyakit Kanker Payudara

Subbab ini menjelaskan tampilan sistem pakar saat pengguna memilih menu penyakit kanker payudara. Sistem pakar penyakit ini membutuhkan data – data pengukuran pada laboratorium untuk mengukur tinggi, luas, kecekungan, keliling dari sel payudara. *Expert system* penyakit kanker payudara menggunakan *ipywidgets FloatText* sebagai metode utama dalam menerima data gejala dari pengguna. Pada tampilan *expert system* ini juga dibuat *grid* yang bertujuan agar merapikan tampilan saat pengguna mengisi data. *Ipywidget Button* digunakan untuk menjalankan fungsi yang akan mengolah data gejala yang sudah diisi menjadi diagnosis penyakit tiroid. Selain itu juga *button* menjalankan fungsi untuk membuat

animasi *loading* untuk memberikan informasi kepada pengguna progres dari pengolahan data. Subbab ini juga menampilkan baris kode penggunaan *classifier* atau model yang menjadi *knowledge base* dari sistem pakar untuk penyakit kanker payudara. Gambar 4.49 hingga 4.52 menunjukkan baris kode penggunaan *classifier* dan tampilan sistem pakar saat mendiagnosis penyakit kanker payudara.

```
*****
*** Expert System Penyakit Kanker Payudara ***
*** Ini adalah sebuah expert system yang dapat digunakan untuk mendeteksi potensi terjangkitnya penyakit kanker payudara ***
*** Harap diperhatikan bahwa diagnosis yang dihasilkan sistem ini hanya bersifat sebagai nasihat medis ***
*** Expert System ini lebih diperuntukan untuk tenaga medis dikarenakan banyak ***
*** membutuhkan data yang berupa pengujian laboratorium dan tes terkait sel payudara ***
*** Tolong berikan jawaban yang sejujur-jujurnya agar hasil diagnosis dari sistem dapat seakurat mungkin ***
*** Silakan isi nilai rerata (mean) dari pengukuran sel. ***
*** Mean Radius [0] Mean Tekstur [0]
*** Mean Keliling [0] Mean Luas [0]
*** Mean Kehalusan [0] Mean Kepadatan [0]
*** Mean Kecekungan [0] Mean Jumlah Cekungan [0]
*** Mean Simetri [0] Mean Dimensi Fraktal [0]
-----
*** Silakan isi nilai error standar (standard error) dari pengukuran sel. ***
*** Error Standar Radius [0] MeanError Standar Tekstur [0]
*** Error Standar Keliling [0] Error Standar Luas [0]
*** Error Standar Kehalusan [0] Error Standar Kepadatan [0]
*** Error Standar Kecekungan [0] Error Standar Jumlah Cekungan [0]
*** Error Standar Simetri [0] Error Standar Dimensi Fraktal [0]
-----
*** Silakan isi nilai terburuk (worst) dari pengukuran sel. ***
*** Radius Terburuk [0] Mean Tekstur [0]
*** Mean Keliling [0] Terburuk Luas [0]
*** Kehalusan Terburuk [0] Kepadatan Terburuk [0]
*** Kecekungan Terburuk [0] Jumlah Cekungan Terburuk [0]
*** Simetri Terburuk [0] Dimensi Fraktal Terburuk [0]
-----
▼ Klik Tombol ini Untuk Melihat Hasil Diagnosis
```

Gambar 4.50 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Kanker Payudara

```
*** Baik, terima kasih atas kerjasama anda. Sekarang kami akan memproses data anda dan menampilkan hasil diagnosis anda. ***
*** Membuka file data training...
*** Berhasil!
*** Memisahkan kolom gejala dan kolom diagnosis...
*** Berhasil!
*** Mengubah kolom - kolom bernilai kategorial...
*** Berhasil!
```

Gambar 4.51 Tampilan Saat Menunggu Diagnosis Penyakit Kanker Payudara Dihasilkan

```
*** Terima kasih atas perhatian anda. Melalui gejala dan data yang dikumpulkan,
*** pasien memiliki tumor tidak ganas pada payudaranya.
*** Lakukan pemeriksaan di puskesmas/rumah sakit bila pasien tetap merasa sakit atau tidak nyaman.
*** Terima kasih anda telah menggunakan layanan expert system, apakah anda ingin menggunakan layanan ini kembali?
```

▼ Ya	✖ Tidak
------	---------

Gambar 4.52 Tampilan Saat Diagnosis Penyakit Kanker Payudara Dihasilkan

#### 4.9.5 Tampilan Sistem Pakar Penyakit Umum

Subbab ini menjelaskan tampilan sistem pakar saat pengguna memilih menu penyakit umum. Sistem pakar penyakit umum membutuhkan data – data gejala yang cukup banyak. Sistem pakar penyakit umum menggunakan *ipywidgets checkbox*, untuk mengakomodasi jumlah data gejala yang banyak dari pengguna. Pada sistem pakar ini juga dibuat sebuah *grid* berdasarkan kategori gejala, seperti gejala yang muncul di daerah kepala dan wajah, indera, leher, otot, perut, dan lain – lain. Pembuatan *grid* ini bertujuan agar pengguna tidak merasa lelah saat mengisi data gejala. Selain itu, dengan adanya *grid* diharapkan pengguna dapat lebih teliti dan membaca setiap gejala yang diberikan sehingga mengurangi kesalahan diagnosis. *Ipywidget Button* digunakan untuk menjalankan fungsi yang akan mengolah data gejala yang sudah diisi menjadi diagnosis penyakit umum. Selain itu juga *button* menjalankan fungsi untuk membuat animasi *loading* untuk memberikan informasi kepada pengguna progres dari pengolahan data. Subbab ini juga menampilkan baris kode penggunaan *classifier* atau model yang menjadi *knowledge base* dari sistem pakar untuk penyakit umum. Gambar 4.53 hingga 4.59 menunjukkan baris kode penggunaan *classifier* dan tampilan sistem pakar saat mendiagnosis penyakit umum.

```
In [35]: 1 def classifier_general(data):
2     import itertools
3     import threading
4     import time
5     import sys
6
7     done = False
8     #here is the animation
9     def animate():
10         for c in itertools.cycle(['|', '/', '-', '\\']):
11             if done:
12                 break
13             sys.stdout.write('\r*** Mohon Menunggu... ' + c + ' ')
14             sys.stdout.flush()
15             time.sleep(0.1)
16         sys.stdout.write('\r*** Berhasil!')
17
18     time.sleep(3)
19     ##Buat classifier algoritma
20     print("\n*** Membuat model klasifikasi algoritma dan memasukkan data ke dalam model...")
21     done = False
22     gen3 = threading.Thread(target=animate)
23     gen3.start()
24     import joblib
25     classifier_general = joblib.load("classifier_general.pkl")
26     done = True
27
28     ##menghasilkan diagnosis
29     time.sleep(3)
30     print("\n*** Menghasilkan diagnosis...")
31     done = False
32     gen4 = threading.Thread(target=animate)
33     gen4.start()
34     prediksi = classifier_general.predict(data)
35     done = True
36     time.sleep(3)
37     return prediksi
```

Gambar 4.53 Baris Kode Penggunaan *Classifier* Sebagai *Knowledge Base* Sistem Pakar Penyakit Umum

\*\*\*\*\*
 \*\*\* Expert System Penyakit Umum
 \*\*\* Ini adalah sebuah expert system yang dapat digunakan untuk mendeteksi potensi terjangkitnya penyakit umum
 \*\*\* Harap diperhatikan bahwa diagnosis yang dihasilkan sistem ini hanya bersifat sebagai nasihat medis
 \*\*\* Periksakan diri anda ke klinik/puskesmas/rumah sakit terdekat berdasarkan hasil diagnosis yang anda dapatkan
 \*\*\*\* Tolong berikan jawaban yang sejujur-jujurnya agar hasil diagnosis dari sistem dapat seakurat mungkin
 \*\*\* Silakan klik pada gejala jika anda mengalami gejala yang disebutkan.
 \*\*\* Bagian Kepala dan Wajah

\*\*\*  Sakit kepala  Wajah dan mata mengembung  
 Bibir kering  Wajah pucat  
 Memiliki jerawat dengan nanah  Komedo  
 Merasa pusing

\*\*\* Bagian Indera (Mata, Hidung, Lidah, Telinga, Kulit)

\*\*\*  Gatal-gatal  Ruam pada kulit  
 Kulit menguning  Berkeringat  
 Kulit mengelupas  Bercak coklat pada kulit  
 Terdapat lebam  Adanya benjolan berisi cairan  
 Rasa gatal di dalam kulit  Titik merah di seluruh tubuh  
 Bersin terus menerus  Sulit bernafas  
 Mata memerah  Mata menguning  
 Adanya kantong mata  Hidung terasa tertekan  
 Pilek  Hidung tersumbat  
 Nanah  Adanya benjolan merah pada hidung  
 Mata berair  Penglihatan terganggu  
 Penglihatan tidak jelas  Rasa nyeri dibelakang mata  
 Indra penciuman hilang  Berbicara tidak lancar  
 Sariawan

Gambar 4.54 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Umum Bagian I

\*\*\* Bagian Leher dan Tenggorokan

\*\*\*  Titik putih di tenggorokan  Benjolan di leher  
 Dahak pada tenggorokan  Rasa gatal pada tenggorokan  
 Dahak berwarna jernih  Dahak berwarna kuning  
 Dahak berdarah  Rasa nyeri di leher  
 Leher terasa kaku  Batuk

\*\*\* Bagian Otot, Sendi, Kaki dan Tangan

\*\*\*  Otot mati rasa  Rasa sakit saat berjalan  
 Kuku rapuh dan kering  Otot melemah  
 Kaki terasa lemas  Bengkak pada kaki dan tangan  
 Kaki membengkak  Kaki dan tangan dingin  
 Adanya lekukan pada kuku  Kuku membengkak  
 Keram  Rasa nyeri pada otot  
 Rasa kaku saat bergerak  Nyeri pada persendian  
 Rasa nyeri pada lutut  Sendi membengkak

\*\*\* Bagian Dada dan Perut

\*\*\*  Nyeri punggung  Rasa tidak nyaman pada perut bagian atas  
 Nyeri pada perut  Perut membengkak  
 Rasa sakit di bagian dada  Jantung berdegup cepat  
 Jantung berdebar - debar  Rasa nyeri pada pinggul

Gambar 4.55 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Umum Bagian II

\*\*\*  
\*\*\* Bagian Sekresi

\*\*\*    Nyeri saat buang air kecil    Nyeri saat buang air besar  
 Rasa nyeri pada saluran kemih    Air seni berbau tidak sedap  
 Nyeri di sekitar bokong    Feses berdarah  
 Irritasi pada anus    Sering buang air kecil  
 Titik darah di air seni    Air seni berwarna gelap  
 Air seni berwarna kuning    Selalu ingin buang air kecil  
 Sembelit    Diare

\*\*\*  
\*\*\* Bagian Mental/Psikis

\*\*\*    Kelelahan    Cemas  
 Depresi    Perubahan suasana hati (mood swing)

\*\*\*  
\*\*\* Bagian Kondisi Tubuh

\*\*\*    Menggigil    Panas dingin  
 Asam Lambung Naik    Berat badan naik  
 Berat badan turun    Gula darah tidak normal  
 Demam tinggi    Demam ringan  
 Pusing seperti berputar    Kehilangan keseimbangan  
 Dehidrasi    Rasa tidak enak badan  
 Gagal ginjal    Kurang konsentrasi  
 Memiliki perasaan ingin muntah    Kehilangan nafsu makan  
 Nafsu makan meningkat    Rasa lapar berlebihan  
 Muntah - muntah    Berat badan berlebihan

Gambar 4.56 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Umum Bagian III

\*\*\*  
\*\*\* Bagian Lain - Lain

\*\*\*    Menstruasi tidak wajar    Mempunyai riwayat penyakit keluarga  
 Menerima transfusi darah    Memiliki riwayat mengkonsumsi alkohol  
 Pembuluh darah pada betis terlihat    Pembengkakan ekstrim (diseluruh tubuh)  
 Pembuluh darah membengkak    Berhubungan seks diluar pernikahan

\*\*\*  
✓ Klik Tombol ini Untuk Melihat Hasil Diagnosis

Gambar 4.57 Tampilan Sistem Pakar untuk Mendeteksi Penyakit Umum Bagian IV

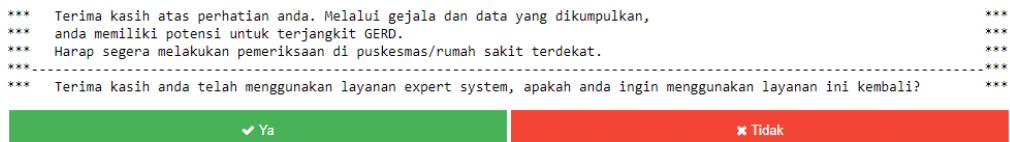
\*\*\*  
\*\*\* Bagian Lain - Lain

\*\*\*    Menstruasi tidak wajar    Mempunyai riwayat penyakit keluarga  
 Menerima transfusi darah    Memiliki riwayat mengkonsumsi alkohol  
 Pembuluh darah pada betis terlihat    Pembengkakan ekstrim (diseluruh tubuh)  
 Pembuluh darah membengkak    Berhubungan seks diluar pernikahan

\*\*\*  
✓ Klik Tombol ini Untuk Melihat Hasil Diagnosis

\*\*\* Baik, terima kasih atas kerjasama anda. Sekarang kami akan memproses data anda dan menampilkan hasil diagnosis anda. \*\*\*  
\*\*\* Membuka file data training...  
\*\*\* Berhasil!  
\*\*\* Mengubah kolom - kolom bernilai kategorial...  
\*\*\* Berhasil!  
\*\*\* Memisahkan kolom gejala dan kolom diagnosis...  
\*\*\* Berhasil!

Gambar 4.58 Tampilan Saat Menunggu Diagnosis Penyakit Umum Dihasilkan



Gambar 4.59 Tampilan Saat Diagnosis Penyakit Umum Dihasilkan

#### 4.9.6 Tampilan Sistem Pakar Penyakit COVID-19

Subbab ini menjelaskan tampilan sistem pakar (*expert system*) saat pengguna memilih menu penyakit COVID-19. Sistem pakar ini menggunakan *ipywidgets RadioButton* dalam menerima data gejala dari pengguna. *Ipywidget Button* digunakan untuk menjalankan fungsi yang akan mengolah data gejala yang sudah diisi menjadi diagnosis penyakit tiroid. Selain itu juga *button* menjalankan fungsi untuk membuat animasi *loading* untuk memberikan informasi kepada pengguna progres dari pengolahan data. Subbab ini juga menampilkan baris kode penggunaan *classifier* atau model yang menjadi *knowledge base* dari *expert system* untuk penyakit COVID-19. Gambar 4.60 hingga 4.64 menunjukkan baris kode penggunaan *classifier* dan tampilan sistem pakar saat mendiagnosis penyakit COVID-19.

```
In [6]: 1 def classifier_covid19(data):
2     import itertools
3     import threading
4     import time
5     import sys
6     done = False
7     #here is the animation
8     def animate():
9         for c in itertools.cycle(['|', '/', '-', '\\']):
10            if done:
11                break
12            sys.stdout.write('\r*** Mohon Menunggu... ' + c + ' ')
13            sys.stdout.flush()
14            time.sleep(0.1)
15    sys.stdout.write('\r*** Berhasil!')
16
17 ##Buat classifier algoritma
18 print("\n*** Membuat model klasifikasi algoritma dan memasukkan data ke dalam model...")
19 done = False
20 cov4 = threading.Thread(target=animate)
21 cov4.start()
22 import joblib
23 classifier_covid19 = joblib.load("classifier_COVID19.pkl")
24 done = True
25
26 time.sleep(3)
27 ##menghasilkan diagnosis
28 print("\n*** Menghasilkan diagnosis...")
29 done = False
30 cov5 = threading.Thread(target=animate)
31 cov5.start()
32 prediksi = classifier_covid19.predict(data)
33 done = True
34 time.sleep(3)
35 return prediksi
```

Gambar 4.60 Baris Kode Penggunaan *Classifier* Sebagai *Knowledge Base* Sistem Pakar Penyakit COVID-19

\*\*\*\*\*
 \*\*\*\* Expert System Penyakit COVID-19
 \*\*\*\*

\*\*\* Inil adalah sebuah expert system yang dapat digunakan untuk mendeteksi potensi terjangkitnya penyakit COVID-19
 \*\*\*

\*\*\* Harap diperhatikan bahwa diagnosis yang dihasilkan sistem ini hanya bersifat sebagai nasihat medis
 \*\*\*

\*\*\* Periksa diri anda ke klinik/puskesmas/rumah sakit terdekat berdasarkan hasil diagnosis yang anda dapatkan
 \*\*\*

\*\*\* Tolong berikan jawaban yang sejujur-jujurnya agar hasil diagnosis dari sistem dapat seakurat mungkin
 \*\*\*

\*\*\* Silakan pilih kategori jenis kelamin anda.
 \*\*\*

\*\*\* Laki-Laki
 \*\*\*

\*\*\* Perempuan
 \*\*\*

\*\*\* Transgender
 \*\*\*

\*\*\*\*\*
 \*\*\* Silakan pilih kategori umur anda.
 \*\*\*

\*\*\* Umur 0-9 tahun
 \*\*\*

\*\*\* Umur 10-19 tahun
 \*\*\*

\*\*\* Umur 20-24 tahun
 \*\*\*

\*\*\* Umur 25-59 tahun
 \*\*\*

\*\*\* Umur diatas 60 tahun
 \*\*\*

\*\*\*\*\*
 \*\*\* Apakah anda mengalami demam?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

\*\*\*\*\*
 \*\*\* Apakah anda mengalami kelelahan?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

\*\*\*\*\*
 \*\*\* Apakah anda mengalami batuk kering?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

\*\*\*\*\*
 \*\*\* Apakah anda mengalami kesulitan bernafas?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

Gambar 4.61 Tampilan Sistem Pakar untuk Mendeteksi Penyakit COVID-19 Bagian I

\*\*\* Apakah anda mengalami radang tenggorokan?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

\*\*\*\*\*
 \*\*\* Apakah disekitar tubuh anda mengalami nyeri?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

\*\*\*\*\*
 \*\*\* Apakah anda mengalami hidung tersumbat?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

\*\*\*\*\*
 \*\*\* Apakah anda mengalami pilek?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

\*\*\*\*\*
 \*\*\* Apakah anda mengalami diare?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

\*\*\*\*\*
 \*\*\* Apakah anda pernah melakukan hubungan kontak dengan orang yang terjangkit COVID-19 ?
 \*\*\*

\*\*\* ya
 \*\*\*

\*\*\* tidak
 \*\*\*

\*\*\* tidak tahu
 \*\*\*

\*\*\*\*\*
 \*\*\* Negara asing yang terakhir anda kunjungi 1 tahun terakhir?
 \*\*\*

\*\*\* Cina
 \*\*\*

\*\*\* Italia
 \*\*\*

\*\*\* Iran
 \*\*\*

\*\*\* Korea Selatan
 \*\*\*

\*\*\* Perancis
 \*\*\*

\*\*\* Spanyol
 \*\*\*

\*\*\* Jerman
 \*\*\*

\*\*\* Uni Emirat Arab
 \*\*\*

\*\*\* Negara lain di Eropa
 \*\*\*

\*\*\* Lain-lain
 \*\*\*


 ✓ Klik Tombol ini Untuk Melihat Hasil Diagnosis

Gambar 4.62 Tampilan Sistem Pakar untuk Mendeteksi Penyakit COVID-19 Bagian II

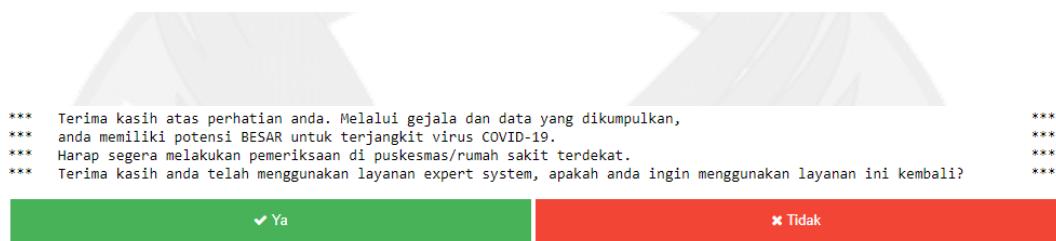
\*\*\*  
\*\*\* Negara asing yang terakhir anda kunjungi 1 tahun terakhir?  
\*\*\*

\*\*\*    Cina  
 Italia  
 Iran  
 Korea Selatan  
 Perancis  
 Spanyol  
 Jerman  
 Uni Emirat Arab  
 Negara lain di Eropa  
 Lain-lain

**✓ Klik Tombol ini Untuk Melihat Hasil Diagnosis**

\*\*\*   Baik, terima kasih atas kerjasama anda. Sekarang kami akan memproses data anda dan menampilkan hasil diagnosis anda. \*\*\*  
\*\*\*  
\*\*\* Membuka file data training...  
\*\*\* Mohon Menunggu... |

Gambar 4.63 Tampilan Saat Menunggu Diagnosis Penyakit COVID-19 Dihasilkan



Gambar 4.64 Tampilan Saat Diagnosis Penyakit COVID-19 Dihasilkan

