

BAB. IV

IMPLEMENTASI DAN PEMBAHASAN

4.1. Data Penelitian

Penelitian ini menggunakan data sekunder yang diperoleh dari rekam medis pasien di Puskesmas Kota Rantauprapat, yang telah disiapkan dalam format Excel dengan nama file “data_malaria.xlsx”. Dataset ini berisi 500 entri (baris) data pasien dengan 8 atribut (kolom) yang menggambarkan karakteristik pasien serta hasil diagnosa terhadap penyakit malaria. Tetapi pada penelitian ini, data yang saya paparkan sebanyak 150 data. Dataset pada penelitian ini diperlihatkan pada Tabel 4.1 berikut.

Tabel 4. 1. Data Sampel

No	Usia	Jenis Kelamin	Suhu Tubuh (°C)	Demam	Menggigil	Mual/ Muntah	Sakit Kepala	Hasil Diagnosa
1	76	Perempuan	34,1	Tidak	Ya	Tidak	Tidak	Negatif
2	27	Laki-laki	37,5	Ya	Ya	Tidak	Tidak	Negatif
3	65	Perempuan	36,9	Ya	Ya	Tidak	Ya	Negatif
4	51	Perempuan	37,9	Ya	Ya	Tidak	Ya	Negatif
5	43	Perempuan	36,2	Ya	Tidak	Ya	Ya	Negatif
6	35	Laki-laki	39,2	Tidak	Ya	Tidak	Tidak	Negatif
7	62	Perempuan	38,5	Ya	Tidak	Tidak	Tidak	Negatif
8	67	Perempuan	37	Tidak	Ya	Ya	Tidak	Negatif
9	20	Perempuan	35	Tidak	Ya	Ya	Tidak	Negatif
10	44	Perempuan	38,5	Tidak	Tidak	Tidak	Tidak	Negatif
11	62	Perempuan	34,7	Ya	Ya	Ya	Tidak	Negatif
12	10	Perempuan	36,8	Tidak	Ya	Ya	Tidak	Negatif
13	8	Laki-laki	38,6	Tidak	Tidak	Tidak	Tidak	Negatif
14	54	Perempuan	36,4	Tidak	Tidak	Tidak	Tidak	Negatif
15	2	Laki-laki	36,5	Ya	Tidak	Tidak	Ya	Negatif
16	17	Laki-laki	36,3	Ya	Ya	Tidak	Ya	Negatif
17	38	Laki-laki	39,4	Tidak	Tidak	Ya	Ya	Negatif
18	12	Perempuan	35,6	Tidak	Tidak	Tidak	Tidak	Negatif

19	57	Laki-laki	37,2	Ya	Ya	Ya	Tidak	Negatif
20	48	Perempuan	36,2	Ya	Ya	Ya	Tidak	Negatif
21	43	Laki-laki	35,7	Tidak	Tidak	Tidak	Tidak	Negatif
22	45	Laki-laki	38,3	Ya	Ya	Ya	Tidak	Positif
23	60	Perempuan	35,6	Tidak	Tidak	Tidak	Ya	Negatif
24	28	Laki-laki	38,7	Ya	Ya	Ya	Ya	Positif
25	33	Perempuan	38,6	Tidak	Tidak	Tidak	Ya	Negatif
26	13	Laki-laki	37,7	Tidak	Tidak	Ya	Ya	Negatif
27	13	Laki-laki	34,8	Tidak	Ya	Tidak	Ya	Negatif
28	63	Laki-laki	36,6	Ya	Tidak	Ya	Ya	Negatif
29	70	Perempuan	37,1	Tidak	Tidak	Ya	Tidak	Negatif
30	4	Laki-laki	36,5	Ya	Tidak	Tidak	Ya	Negatif
31	66	Perempuan	36	Ya	Tidak	Tidak	Tidak	Negatif
32	19	Laki-laki	37,3	Ya	Ya	Ya	Ya	Negatif
33	59	Laki-laki	36,2	Ya	Tidak	Ya	Tidak	Negatif
34	54	Laki-laki	35,4	Tidak	Ya	Tidak	Ya	Negatif
35	39	Laki-laki	37,9	Tidak	Tidak	Ya	Ya	Positif
36	55	Perempuan	35,1	Ya	Ya	Tidak	Tidak	Negatif
37	19	Perempuan	37,5	Tidak	Tidak	Ya	Ya	Negatif
38	59	Laki-laki	37,4	Ya	Tidak	Ya	Ya	Negatif
39	38	Laki-laki	39,6	Tidak	Tidak	Ya	Tidak	Negatif
40	78	Perempuan	35,3	Ya	Tidak	Ya	Ya	Negatif
41	73	Laki-laki	36,6	Ya	Tidak	Tidak	Ya	Negatif
42	59	Perempuan	35,9	Tidak	Ya	Tidak	Tidak	Negatif
43	55	Perempuan	38,6	Tidak	Ya	Tidak	Ya	Negatif
44	18	Laki-laki	37,2	Tidak	Ya	Tidak	Ya	Negatif
45	22	Perempuan	39,2	Tidak	Tidak	Ya	Ya	Positif
46	22	Perempuan	39,7	Tidak	Tidak	Tidak	Ya	Negatif
47	17	Laki-laki	35,5	Ya	Ya	Tidak	Ya	Negatif
48	78	Perempuan	37,8	Tidak	Tidak	Tidak	Tidak	Negatif
49	26	Perempuan	36,9	Ya	Tidak	Ya	Ya	Negatif
50	6	Laki-laki	38,1	Ya	Ya	Tidak	Ya	Positif
51	61	Perempuan	39,1	Ya	Ya	Tidak	Ya	Positif
52	3	Perempuan	36,1	Ya	Tidak	Tidak	Tidak	Negatif
53	60	Perempuan	36,3	Ya	Tidak	Tidak	Ya	Negatif
54	11	Laki-laki	37,5	Ya	Tidak	Tidak	Tidak	Negatif
55	46	Laki-laki	36,5	Tidak	Tidak	Ya	Ya	Negatif
56	11	Laki-laki	36	Tidak	Ya	Ya	Ya	Negatif
57	51	Perempuan	37,3	Tidak	Ya	Ya	Tidak	Negatif
58	21	Laki-laki	35,4	Tidak	Ya	Ya	Ya	Negatif
59	50	Perempuan	34,4	Ya	Tidak	Tidak	Ya	Negatif

60	36	Perempuan	38,7	Ya	Tidak	Tidak	Ya	Positif
61	1	Laki-laki	35,3	Ya	Tidak	Ya	Ya	Negatif
62	16	Perempuan	38,1	Ya	Tidak	Ya	Ya	Positif
63	22	Perempuan	39,8	Ya	Ya	Tidak	Tidak	Positif
64	61	Perempuan	36,2	Ya	Ya	Ya	Tidak	Negatif
65	41	Perempuan	36,3	Ya	Tidak	Tidak	Ya	Negatif
66	54	Laki-laki	34,5	Ya	Tidak	Tidak	Tidak	Negatif
67	9	Perempuan	39,1	Tidak	Tidak	Tidak	Ya	Negatif
68	41	Laki-laki	39,1	Tidak	Ya	Ya	Tidak	Negatif
69	25	Laki-laki	37,3	Ya	Ya	Ya	Ya	Negatif
70	64	Perempuan	36,7	Ya	Tidak	Ya	Tidak	Negatif
71	62	Perempuan	36,1	Ya	Tidak	Tidak	Tidak	Negatif
72	62	Laki-laki	38,1	Ya	Tidak	Ya	Ya	Positif
73	7	Perempuan	36,4	Ya	Tidak	Ya	Tidak	Negatif
74	48	Perempuan	32,6	Ya	Tidak	Tidak	Tidak	Negatif
75	74	Laki-laki	36,4	Ya	Tidak	Ya	Tidak	Negatif
76	29	Perempuan	37,4	Ya	Ya	Ya	Tidak	Negatif
77	62	Laki-laki	35,7	Tidak	Ya	Tidak	Ya	Negatif
78	54	Laki-laki	36,8	Tidak	Ya	Ya	Tidak	Negatif
79	47	Laki-laki	39,5	Ya	Tidak	Ya	Tidak	Positif
80	73	Laki-laki	39,8	Tidak	Tidak	Tidak	Ya	Negatif
81	47	Laki-laki	35,3	Tidak	Tidak	Tidak	Ya	Negatif
82	28	Laki-laki	39,9	Tidak	Ya	Tidak	Tidak	Negatif
83	33	Perempuan	35,9	Ya	Tidak	Ya	Ya	Negatif
84	51	Perempuan	34,6	Ya	Tidak	Ya	Tidak	Negatif
85	38	Laki-laki	37,8	Ya	Ya	Tidak	Ya	Positif
86	45	Laki-laki	37,1	Ya	Tidak	Ya	Tidak	Negatif
87	32	Laki-laki	35,3	Tidak	Tidak	Tidak	Ya	Negatif
88	69	Laki-laki	37	Tidak	Tidak	Ya	Tidak	Negatif
89	74	Laki-laki	36,3	Ya	Ya	Tidak	Ya	Negatif
90	52	Laki-laki	38,4	Tidak	Ya	Ya	Tidak	Negatif
91	29	Laki-laki	35,2	Ya	Ya	Ya	Tidak	Negatif
92	77	Laki-laki	36,6	Ya	Ya	Tidak	Tidak	Negatif
93	2	Laki-laki	36,9	Ya	Ya	Ya	Tidak	Negatif
94	67	Laki-laki	35,4	Ya	Tidak	Ya	Ya	Negatif
95	9	Laki-laki	34,9	Tidak	Ya	Ya	Tidak	Negatif
96	65	Laki-laki	37,2	Ya	Ya	Ya	Ya	Negatif
97	6	Perempuan	36,3	Ya	Tidak	Tidak	Tidak	Negatif
98	59	Perempuan	38,6	Tidak	Ya	Tidak	Ya	Positif
99	79	Laki-laki	37,5	Tidak	Tidak	Ya	Ya	Negatif
100	39	Perempuan	34,5	Ya	Ya	Tidak	Tidak	Negatif

101	15	Perempuan	35,4	Ya	Tidak	Ya	Tidak	Negatif
102	22	Perempuan	35,4	Tidak	Ya	Tidak	Tidak	Negatif
103	62	Perempuan	36,3	Ya	Tidak	Ya	Tidak	Negatif
104	5	Perempuan	38	Tidak	Tidak	Tidak	Ya	Negatif
105	26	Laki-laki	37,5	Tidak	Tidak	Ya	Ya	Positif
106	52	Perempuan	39,2	Ya	Tidak	Ya	Tidak	Negatif
107	6	Perempuan	35,9	Ya	Tidak	Tidak	Tidak	Negatif
108	54	Laki-laki	36,3	Tidak	Tidak	Ya	Ya	Negatif
109	12	Perempuan	38,1	Tidak	Tidak	Ya	Ya	Negatif
110	36	Perempuan	35,3	Ya	Ya	Ya	Ya	Negatif
111	56	Perempuan	35,1	Tidak	Ya	Tidak	Tidak	Negatif
112	56	Laki-laki	39,8	Tidak	Tidak	Ya	Tidak	Negatif
113	13	Perempuan	37,4	Ya	Tidak	Tidak	Ya	Negatif
114	23	Laki-laki	35,8	Ya	Tidak	Tidak	Ya	Negatif
115	45	Perempuan	34,7	Tidak	Tidak	Tidak	Tidak	Negatif
116	39	Laki-laki	35,6	Tidak	Ya	Ya	Ya	Negatif
117	17	Perempuan	39,5	Ya	Tidak	Ya	Ya	Positif
118	49	Perempuan	36,8	Ya	Ya	Tidak	Tidak	Negatif
119	76	Perempuan	36,8	Tidak	Ya	Tidak	Tidak	Negatif
120	72	Perempuan	36,4	Ya	Ya	Ya	Tidak	Negatif
121	62	Laki-laki	38,5	Tidak	Tidak	Ya	Tidak	Negatif
122	35	Laki-laki	37	Ya	Ya	Ya	Tidak	Negatif
123	11	Laki-laki	37,6	Tidak	Tidak	Ya	Tidak	Negatif
124	19	Laki-laki	35,1	Tidak	Ya	Tidak	Tidak	Negatif
125	62	Perempuan	36,1	Tidak	Ya	Tidak	Tidak	Negatif
126	38	Laki-laki	37,7	Ya	Tidak	Ya	Ya	Positif
127	40	Perempuan	36,7	Tidak	Ya	Tidak	Tidak	Negatif
128	56	Perempuan	36,8	Ya	Tidak	Tidak	Ya	Negatif
129	58	Perempuan	36,6	Tidak	Ya	Ya	Ya	Negatif
130	73	Perempuan	35,1	Ya	Tidak	Tidak	Tidak	Negatif
131	6	Laki-laki	38,3	Ya	Ya	Tidak	Ya	Positif
132	67	Perempuan	36,7	Tidak	Tidak	Ya	Tidak	Negatif
133	19	Perempuan	35,1	Tidak	Ya	Tidak	Tidak	Negatif
134	76	Perempuan	35,5	Ya	Ya	Ya	Ya	Negatif
135	21	Perempuan	37,2	Ya	Tidak	Tidak	Ya	Negatif
136	68	Laki-laki	35,5	Tidak	Ya	Tidak	Ya	Negatif
137	42	Perempuan	36,9	Tidak	Tidak	Ya	Tidak	Negatif
138	64	Laki-laki	35,9	Ya	Tidak	Ya	Tidak	Negatif
139	32	Perempuan	37,4	Ya	Ya	Ya	Tidak	Negatif
140	79	Laki-laki	38,8	Ya	Tidak	Ya	Tidak	Negatif
141	59	Perempuan	38,7	Ya	Ya	Tidak	Tidak	Negatif

142	64	Perempuan	37,1	Ya	Tidak	Tidak	Ya	Negatif
143	23	Laki-laki	39,1	Ya	Tidak	Tidak	Tidak	Negatif
144	63	Laki-laki	37,3	Ya	Tidak	Ya	Ya	Negatif
145	57	Perempuan	37,6	Ya	Tidak	Ya	Ya	Positif
146	62	Laki-laki	35,3	Tidak	Tidak	Tidak	Tidak	Negatif
147	59	Perempuan	38,5	Tidak	Tidak	Tidak	Tidak	Negatif
148	42	Perempuan	38,6	Ya	Tidak	Tidak	Tidak	Negatif
149	21	Perempuan	37,1	Tidak	Ya	Ya	Tidak	Negatif
150	58	Laki-laki	37,3	Ya	Tidak	Tidak	Ya	Negatif

Dataset pada Tabel 4.1 terdiri dari beberapa atribut yang merepresentasikan gejala dan kondisi medis pasien yang diperiksa di Puskesmas Kota Rantauuprapt, dengan tujuan akhir untuk mengetahui apakah pasien tersebut terdiagnosis malaria (Positif) atau tidak (Negatif). Adapun atribut-atribut dalam dataset tersebut dijelaskan sebagai berikut:

1. Usia (numerik)

Atribut ini menunjukkan umur pasien dalam satuan tahun. Rentang usia dalam dataset berkisar dari 1 tahun hingga 79 tahun, dengan usia rata-rata pasien adalah 41,21 tahun. Atribut ini termasuk tipe numerik (integer) dan memiliki kontribusi penting dalam melihat kecenderungan usia terhadap kemungkinan menderita malaria.

2. Jenis Kelamin (kategori)

Terdapat dua kategori pada atribut ini, yaitu “Laki-laki” dan “Perempuan”. Distribusi data menunjukkan bahwa jumlah pasien perempuan sedikit lebih banyak, yaitu 257 orang (51,4%), sementara pasien laki-laki sebanyak 243 orang (48,6%).

3. Suhu Tubuh (numerik)

Atribut ini menyatakan suhu tubuh pasien dalam satuan derajat Celcius. Nilai suhu tubuh berkisar antara 32,1°C hingga 40,7°C, dengan rata-rata sebesar 36,91°C dan standar deviasi sebesar 1,46. Suhu tubuh yang tinggi dapat menjadi indikator gejala malaria.

4. Demam (kategori)

Atribut ini mengindikasikan apakah pasien mengalami demam atau tidak, dengan dua nilai yaitu “Ya” dan “Tidak”. Sebanyak 262 pasien (52,4%) dalam dataset mengalami demam.

5. Menggigil (kategori)

Atribut ini menunjukkan keberadaan gejala menggigil pada pasien. Sama seperti atribut demam, atribut ini juga terdiri dari nilai “Ya” dan “Tidak”. Sebanyak 251 pasien (50,2%) diketahui mengalami menggigil.

6. Mual/Muntah (kategori)

Gejala mual dan/atau muntah juga dicatat sebagai indikator tambahan. Atribut ini memiliki dua nilai “Ya” dan “Tidak”, dengan 245 pasien (49%) mengalami mual atau muntah, sementara sisanya tidak.

7. Sakit Kepala (kategori)

Atribut ini menyatakan apakah pasien mengalami sakit kepala, dengan distribusi terbagi antara “Ya” dan “Tidak”. Sebanyak 229 pasien (45,8%) mengalami sakit kepala, sedangkan 271 pasien (54,2%) tidak mengalaminya.

8. Hasil Diagnosa (target class)

Ini merupakan atribut target (label) yang akan diprediksi dalam penelitian. Nilainya terdiri dari dua kelas, yaitu “Positif” (menderita malaria) dan “Negatif” (tidak menderita malaria). Berdasarkan distribusi data, sebanyak 59 pasien (11,8%) dinyatakan positif malaria, sedangkan 441 pasien (88,2%) dinyatakan negatif.

4.2. Tahapan Implementasi dengan RapidMiner

Implementasi sistem prediksi dilakukan menggunakan platform RapidMiner Studio. Proses ini mencakup semua tahapan data mining yang telah dijelaskan pada BAB III, mulai dari import data hingga evaluasi model. RapidMiner dipilih karena kemampuannya untuk membangun model machine learning secara visual tanpa memerlukan penulisan kode program, sehingga lebih efisien dan mudah direplikasi

Data pasien penyakit malaria pada Puskesmas Kota Rantauprapat akan diimplementasikan ke dalam aplikasi RapidMiner 9.10 dengan menerapkan algoritma Naïve Bayes. Proses implementasi dilakukan dengan menjadikan variabel Hasil Diagnosa sebagai atribut Label (target) untuk prediksi, dengan atribut-atribut lainnya sebagai atribut independen.

4.2.1. Import Data

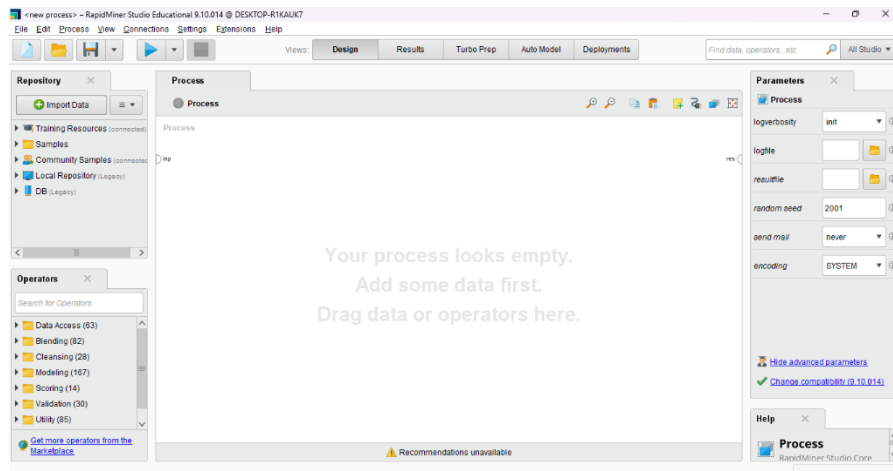
Tahap import data adalah proses memasukkan data dari sumber eksternal ke dalam lingkungan perangkat lunak atau sistem yang akan digunakan untuk analisis atau pemrosesan lebih lanjut.



Gambar 4. 1. Proses Membuka Aplikasi RapidMiner

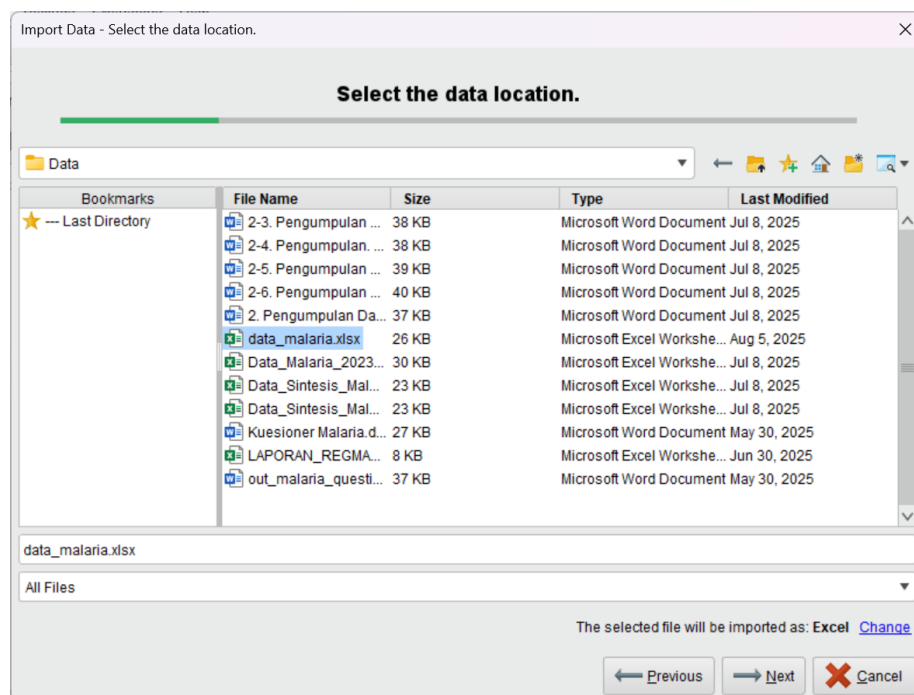
Gambar 4.1 menunjukkan tampilan awal (*splash screen*) saat membuka aplikasi RapidMiner Studio versi 9.10. Berikut penjelasan detailnya:

1. Di bagian kiri atas terdapat logo RapidMiner yang merupakan identitas visual dari perangkat lunak ini.
2. Terdapat tulisan besar "Studio" yang menandakan nama produk, yaitu RapidMiner Studio.
3. Tertera versi aplikasi yaitu Version 9.10, menunjukkan bahwa pengguna menggunakan versi terbaru atau spesifik dari RapidMiner Studio.
4. Tulisan "Reading Configuration Files" menandakan bahwa aplikasi sedang memuat file konfigurasi yang diperlukan untuk menjalankan program dengan benar.



Gambar 4. 2. Lembar Kerja Utama Aplikasi RapidMiner

Gambar 4.2 menampilkan tampilan utama (main workspace) dari aplikasi RapidMiner Studio setelah aplikasi berhasil dibuka dan siap digunakan untuk melakukan proses analisis data.

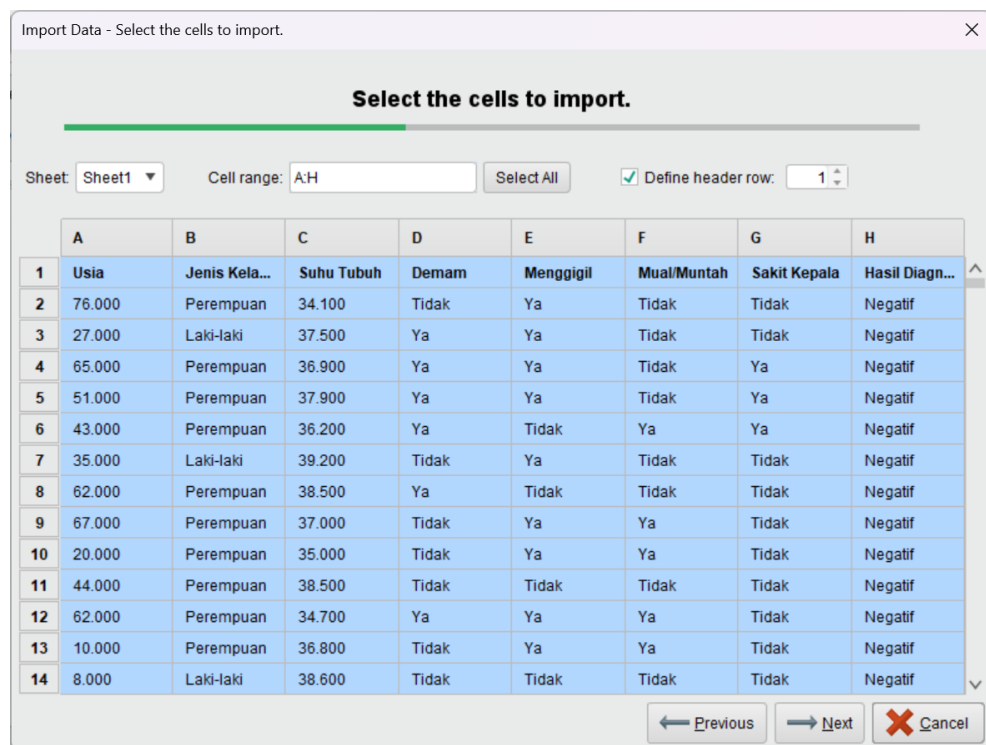


Gambar 4. 3. Direktori File Import Data

Gambar 4.3 menunjukkan tampilan jendela “Import Data - Select the data location” pada RapidMiner Studio. Tampilan ini muncul saat pengguna melakukan proses awal untuk mengimpor dataset ke dalam RapidMiner, khususnya dataset

yang akan digunakan sebagai input untuk proses machine learning. Dalam gambar ini, terlihat bahwa sedang memilih file data_malaria.xlsx, yang merupakan file Excel berisi data pasien dari Puskesmas Kota Rantauprapat yang digunakan untuk membangun model prediksi penyakit malaria. File ini tersimpan dalam folder lokal yang ditandai dengan label "Last Directory" di bagian kiri tampilan. File data_malaria.xlsx berukuran 37 KB, menunjukkan bahwa data yang dimuat cukup ringan dan tidak terlalu besar.

Proses pemilihan file ini merupakan tahap awal dari seluruh rangkaian data preprocessing di RapidMiner. File data_malaria.xlsx yang berisi atribut seperti usia, jenis kelamin, suhu tubuh, gejala-gejala (demam, menggigil, mual/muntah, sakit kepala), serta label diagnosis malaria (positif/negatif), akan diproses lebih lanjut sebelum dilatih menggunakan algoritma Naïve Bayes.



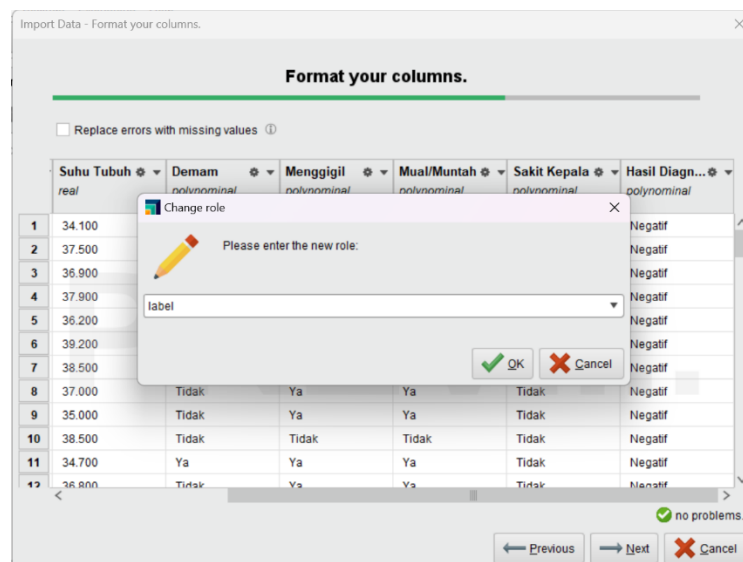
Gambar 4. 4. Select the Cells to Import

Gambar 4.4 menampilkan tahap kedua dari proses pengimporan data ke dalam RapidMiner Studio, yaitu langkah untuk memilih range data dari worksheet yang akan dimasukkan ke dalam proses analisis. Pada tampilan ini, pengguna dihadapkan pada jendela antarmuka dengan judul “Select the cells to import”, yang menunjukkan isi tabel dari Sheet1 file Excel data_malaria.xlsx.

Tabel 4. 2. Atribut Data

No	Atribut	Deskripsi
1	Usia	Umur pasien (dalam tahun)
2	Jenis Kelamin	Jenis kelamin pasien (Laki-laki / Perempuan)
3	Suhu Tubuh	Suhu tubuh saat kunjungan ke puskesmas (dalam °C)
4	Demam	Gejala demam (Ya / Tidak)
5	Menggigil	Gejala menggigil (Ya / Tidak)
6	Mual/Muntah	Gejala mual atau muntah (Ya / Tidak)
7	Sakit Kepala	Gejala sakit kepala (Ya / Tidak)
8	Hasil Diagnosa	Label klasifikasi: hasil diagnosis malaria (Positif / Negatif)

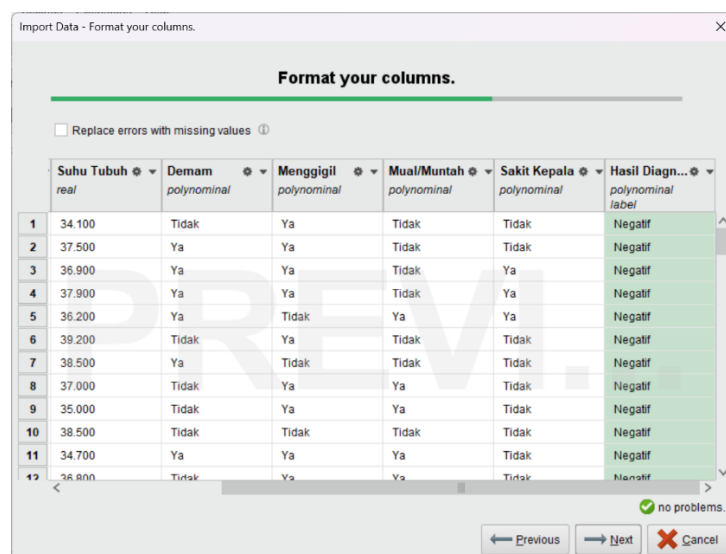
Tabel 4.2 yang ditampilkan terdiri dari 8 atribut (kolom) dan beberapa baris data pasien, yang masing-masing mewakili entri rekam medis. Setiap kolom mewakili fitur (feature) atau variabel yang akan digunakan dalam proses klasifikasi.



Gambar 4. 5. Set Role Variabel Target

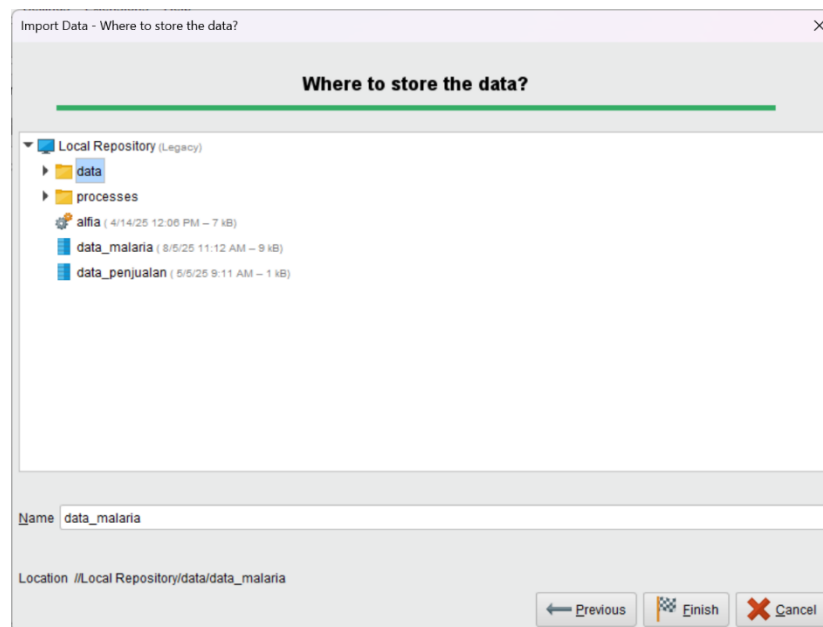
Gambar 4.5 menampilkan tahap formatting kolom dalam proses impor data menggunakan RapidMiner Studio. Pada tahap ini, pengguna diminta untuk menentukan peran (*role*) dari masing-masing atribut atau kolom yang telah dipilih sebelumnya dalam dataset. Fokus utama gambar ini adalah penetapan variabel target (label) untuk proses klasifikasi. Variabel target dalam penelitian ini adalah “Hasil Diagnosa”, yang akan digunakan oleh algoritma Naïve Bayes untuk membedakan antara pasien yang positif dan negatif malaria berdasarkan atribut gejala dan kondisi lainnya.

Penetapan peran atribut sebagai label merupakan langkah sangat penting dalam proses pembelajaran mesin (machine learning). Tanpa penetapan ini, RapidMiner tidak akan mengetahui atribut mana yang menjadi target prediksi. Dalam konteks penelitian ini: Label: Hasil Diagnosa (Positif / Negatif). Fitur input: Usia, Suhu Tubuh, Demam, Menggigil, Mual/Muntah, Sakit Kepala, dan Jenis Kelamin. Dengan set role yang tepat, algoritma Naïve Bayes dapat menghitung probabilitas kemunculan label berdasarkan kombinasi atribut gejala pasien.



Gambar 4. 6. Format dan Tipe Data

Gambar 4.6 memperlihatkan tampilan akhir dari proses penyesuaian format dan tipe data sebelum data digunakan dalam proses machine learning di RapidMiner Studio. Tahapan ini merupakan bagian dari proses impor data yang dikenal dengan jendela “Format your columns”. Tujuan utama dari langkah ini adalah untuk memastikan bahwa setiap atribut (kolom) dalam dataset memiliki tipe data dan peran (role) yang sesuai dengan karakteristik datanya dan kebutuhan algoritma klasifikasi yang akan digunakan.



Gambar 4. 7. Direktori Penyimpanan File

Gambar 4.7 memperlihatkan tampilan jendela RapidMiner pada tahap akhir proses import dataset, tepatnya di bagian “Where to store the data?”. Pada tahap ini, pengguna menentukan lokasi penyimpanan internal untuk menyimpan dataset yang telah berhasil diimpor, agar dapat digunakan kembali dalam proses analisis selanjutnya tanpa perlu mengimpor ulang.

ExampleSet (500 / 500 examples): all

Row No.	Hasil Diagnosa	Usia	Jenis Kelamin	Suhu Tubuh	Demam	Menggigil	Mual/Muntah
1	Negatif	76	Perempuan	34.100	Tidak	Ya	Tidak
2	Negatif	27	Laki-laki	37.500	Ya	Ya	Tidak
3	Negatif	65	Perempuan	36.900	Ya	Ya	Tidak
4	Negatif	51	Perempuan	37.900	Ya	Ya	Tidak
5	Negatif	43	Perempuan	36.200	Ya	Tidak	Ya
6	Negatif	35	Laki-laki	39.200	Tidak	Ya	Tidak
7	Negatif	62	Perempuan	38.500	Ya	Tidak	Tidak
8	Negatif	67	Perempuan	37	Tidak	Ya	Ya
9	Negatif	20	Perempuan	35	Tidak	Ya	Ya
10	Negatif	44	Perempuan	38.500	Tidak	Tidak	Tidak
11	Negatif	62	Perempuan	34.700	Ya	Ya	Ya
12	Negatif	10	Perempuan	36.800	Tidak	Ya	Ya

ExampleSet (500 examples, 1 special attribute, 7 regular attributes)

Gambar 4. 8. Preview Dataset

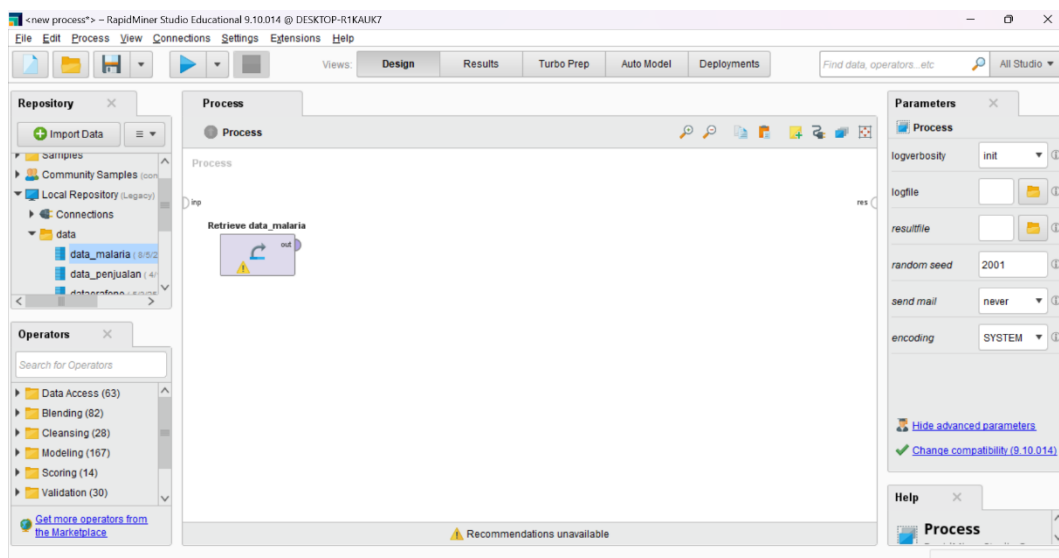
Gambar 4.8 menampilkan tampilan jendela RapidMiner Studio saat pengguna sedang melakukan preview dataset data_malaria yang telah berhasil diimpor ke dalam sistem. Dataset ini merupakan kumpulan data pasien Puskesmas Kota Rantauprapat yang digunakan dalam penelitian untuk memprediksi penyakit malaria dengan algoritma Naïve Bayes. Gambar ini menunjukkan tampilan aktif pada tab Results, yaitu bagian dari RapidMiner yang menampilkan hasil eksekusi dari operator seperti Read Excel atau Retrieve. Tab ini memungkinkan pengguna melihat isi dataset sebelum melakukan proses lebih lanjut seperti preprocessing, training model, dan evaluasi.

Nama Dataset ditampilkan pada tab “ExampleSet (Local Repository/data/data_malaria)”, menandakan bahwa dataset yang sedang ditinjau adalah data_malaria yang tersimpan di repositori lokal dalam folder data. Jumlah Data menyebutkan: “ExampleSet (500 examples, 1 special attribute, 7 regular attributes)”. Artinya: Total 500 baris data (instances). Terdapat 1 atribut khusus (special attribute), kemungkinan besar adalah Hasil Diagnosa yang telah diset

sebagai label (target). Dan terdapat 7 atribut regular (fitur prediktor) seperti: Usia, Jenis Kelamin, Suhu Tubuh, Demam, Menggigil, Mual/Muntah, dll.

4.2.2. Pra-pemrosesan Data

Setelah data berhasil diimpor ke dalam RapidMiner, langkah selanjutnya adalah melakukan pra-pemrosesan data. Tahap ini sangat penting untuk memastikan kualitas dan konsistensi data sebelum digunakan dalam proses analisis atau pemodelan machine learning.



Gambar 4. 9. Retrieve Data

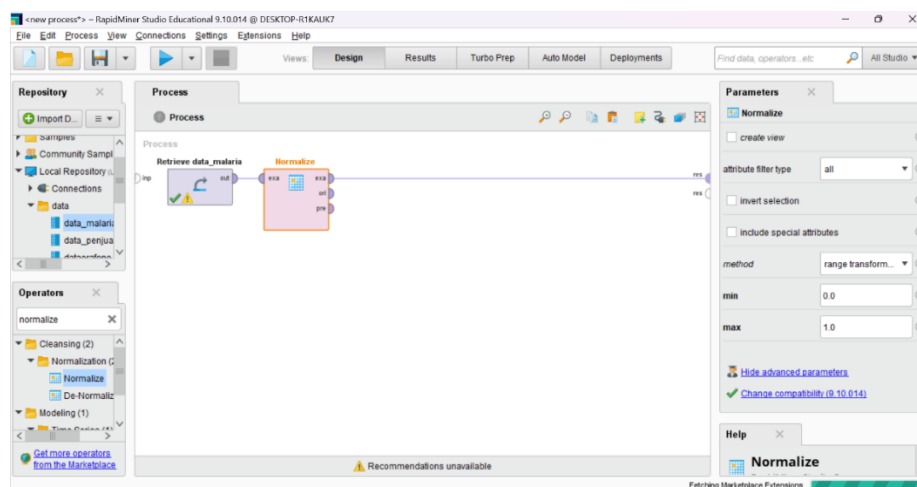
Gambar 4.9 menunjukkan tampilan proses awal dalam RapidMiner Studio, yaitu langkah Retrieve “data_malaria” yang berfungsi untuk mengambil dataset yang sudah disimpan di repository lokal agar dapat digunakan dalam pra-pemrosesan data.

Name	Type	Missing	Least	Most
Label	Polynomial	0	Positif (59)	Negatif (441)
Usia	Integer	0	1	79
Jenis Kelamin	Polynomial	0	Laki-laki (243)	Perempuan (257)
Suhu Tubuh	Real	0	32.100	40.700
Demam	Polynomial	0	Tidak (238)	Ya (262)
Menggigil	Polynomial	0	Tidak (249)	Ya (251)
Mual/Muntah	Polynomial	0	Ya (245)	Tidak (255)

Showing attributes 1 - 8
Examples: 500 Special Attributes: 1 Regular Attributes: 7

Gambar 4. 10. Handle Missing Value

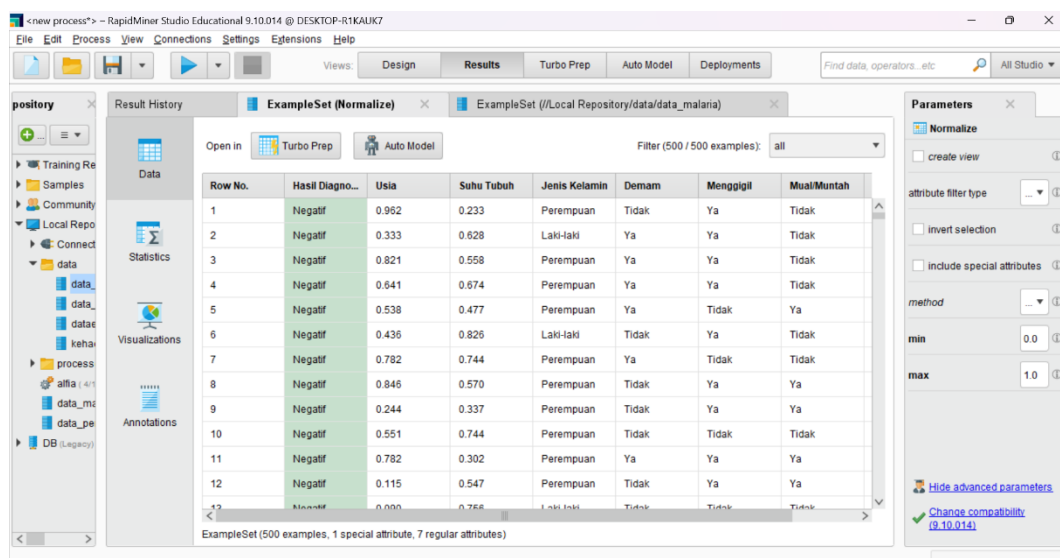
Gambar 4.10 ini menampilkan antarmuka RapidMiner Studio saat pengguna meninjau statistik ringkasan dari dataset `data_malaria` setelah proses import selesai. Tampilan ini muncul di tab Results, bagian dari operator `Read Excel` atau `Retrieve`, yang digunakan untuk meninjau kualitas data sebelum memasuki tahap preprocessing atau pemodelan. Kolom Missing menunjukkan nilai 0 untuk semua atribut, yang berarti tidak ada data yang hilang dalam dataset ini. Ini menunjukkan bahwa dataset telah siap digunakan langsung tanpa perlu proses imputasi atau pembersihan data lanjutan.



Gambar 4. 11. Normalisasi Data

Gambar 4.11 merupakan tahapan penting dalam preprocessing sebelum dilakukan pemodelan di RapidMiner yaitu tahap Normalisasi Data. Gambar ini menunjukkan tampilan antarmuka RapidMiner Studio pada tab Design, ketika pengguna sedang membangun alur proses untuk proyek klasifikasi prediksi malaria. Fokus proses ini adalah pada penerapan proses normalisasi data menggunakan operator `Normalize`. Operator ini berfungsi untuk melakukan normalisasi skala nilai pada atribut numerik, dalam hal ini adalah kolom seperti `Usia` dan `Suhu Tubuh`.

Normalisasi berguna untuk menstandarkan rentang nilai sehingga metode klasifikasi seperti Naive Bayes dapat bekerja secara optimal dan tidak berat sebelah. Normalisasi memastikan bahwa semua atribut numerik memberikan kontribusi yang seimbang. Dengan nilai input dalam rentang standar (0–1), model dapat belajar pola data dengan lebih efisien. Normalisasi membantu mempercepat proses pelatihan karena menghindari nilai ekstrim atau outlier yang besar.



The screenshot shows the RapidMiner Studio interface with the 'Normalize' operator selected. The 'Parameters' panel on the right shows the 'min' value set to 0.0 and the 'max' value set to 1.0. The main workspace displays a table of 12 rows of data, with the first 11 rows highlighted in green. The table columns are: Row No., Hasil Diagno..., Usia, Suhu Tubuh, Jenis Kelamin, Demam, Menggigil, and Mual/Muntah.

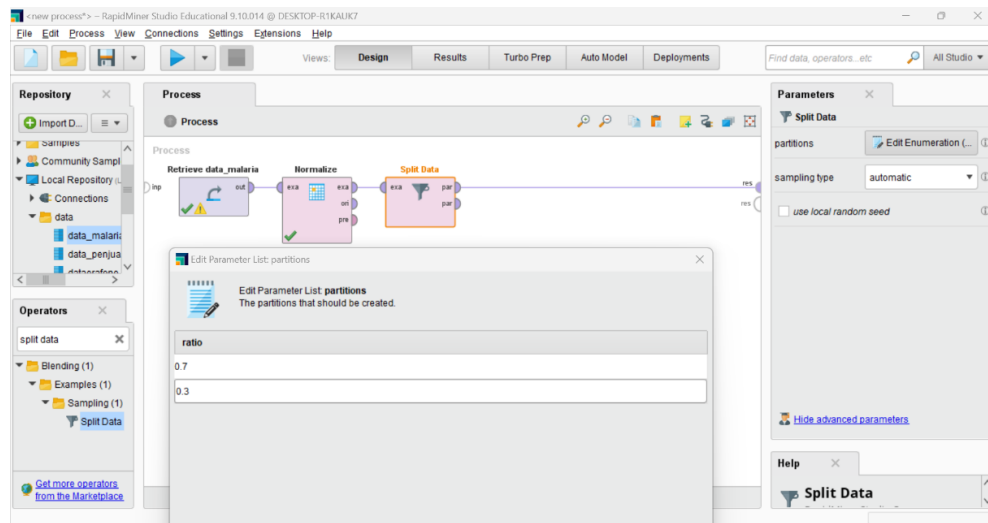
Row No.	Hasil Diagno...	Usia	Suhu Tubuh	Jenis Kelamin	Demam	Menggigil	Mual/Muntah
1	Negatif	0.962	0.233	Perempuan	Tidak	Ya	Tidak
2	Negatif	0.333	0.628	Laki-laki	Ya	Ya	Tidak
3	Negatif	0.821	0.558	Perempuan	Ya	Ya	Tidak
4	Negatif	0.641	0.674	Perempuan	Ya	Ya	Tidak
5	Negatif	0.538	0.477	Perempuan	Ya	Tidak	Ya
6	Negatif	0.436	0.826	Laki-laki	Tidak	Ya	Tidak
7	Negatif	0.782	0.744	Perempuan	Ya	Tidak	Tidak
8	Negatif	0.846	0.570	Perempuan	Tidak	Ya	Ya
9	Negatif	0.244	0.337	Perempuan	Tidak	Ya	Ya
10	Negatif	0.551	0.744	Perempuan	Tidak	Tidak	Tidak
11	Negatif	0.782	0.302	Perempuan	Ya	Ya	Ya
12	Negatif	0.115	0.547	Perempuan	Tidak	Ya	Ya

Gambar 4. 12. Hasil Normalisasi Data

Gambar 4.12 ini menampilkan tampilan pada tab Results dari RapidMiner setelah proses normalisasi data berhasil dijalankan. Proses ini merupakan lanjutan dari alur pada Gambar 4.11, dan menunjukkan hasil akhir dari transformasi nilai numerik dataset data_malaria ke dalam rentang nilai yang telah ditentukan, yaitu antara 0 dan 1. Kolom Usia – Telah dinormalisasi (misalnya dari 76 menjadi 0.963). Kolom Suhu Tubuh – Telah dinormalisasi (misalnya dari 34.1 menjadi 0.233). Ini sesuai dengan metode normalisasi range transformation, dengan rentang minimum 0.0 dan maksimum 1.0. Sedangkan atribut Kategorikal Tetap Dipertahankan. Atribut seperti Jenis Kelamin, Demam, Menggigil, dan Mual/Muntah tetap dalam format polynominal, karena tidak relevan untuk dinormalisasi secara numerik. Jumlah baris tetap 500 contoh (examples), sesuai dengan jumlah data asli. Kolom Hasil Diagnosa tetap ditandai sebagai label, dan tidak ikut dinormalisasi.

4.2.3. Pembagian Data

Setelah data berhasil diimpor dan dilakukan pra-pemrosesan seperti normalisasi, langkah selanjutnya adalah pembagian data (data splitting). Proses ini sangat penting dalam machine learning untuk memastikan model yang dibangun dapat diuji secara objektif.



Gambar 4. 13. Pembagian Data

Gambar 4.13 di atas menunjukkan proses pembagian data (*data splitting*) dalam RapidMiner. Dalam proses ini, data dibagi menggunakan operator Split Data di RapidMiner, dengan parameter rasio pembagian sebesar 70% (0.7) untuk data pelatihan dan 30% (0.3) untuk data pengujian, yang diterapkan pada dataset *data_malaria*. Ini merupakan praktik umum dalam machine learning agar model dapat dilatih dengan cukup data sambil tetap memiliki set data terpisah untuk evaluasi kinerja model.

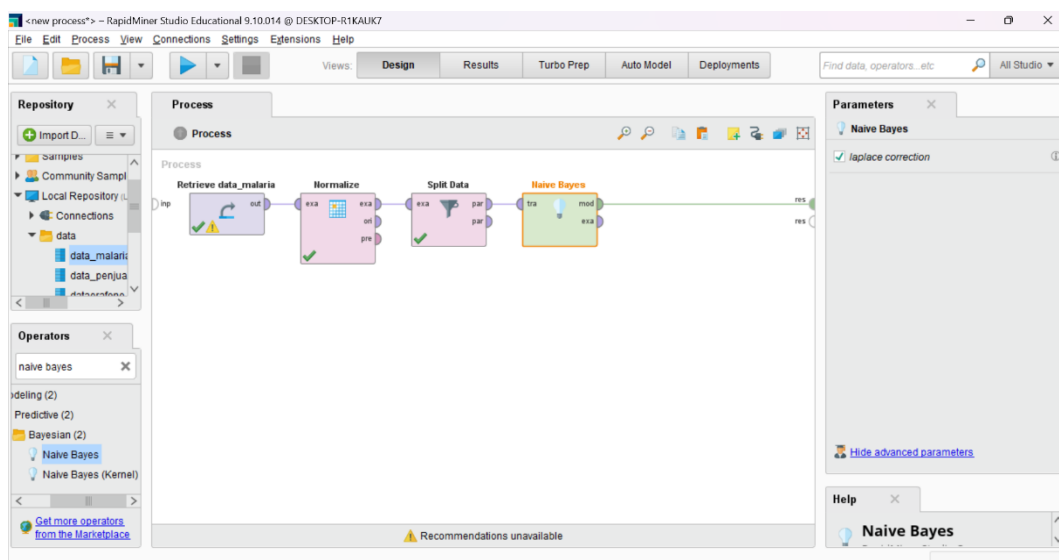
Row No.	Hasil Diagno...	Usia	Suhu Tubuh	Jenis Kelamin	Demam	Menggigil	Mual/Muntah
1	Negatif	0.333	0.628	Laki-laki	Ya	Ya	Tidak
2	Negatif	0.821	0.558	Perempuan	Ya	Ya	Tidak
3	Negatif	0.641	0.674	Perempuan	Ya	Ya	Tidak
4	Negatif	0.436	0.826	Laki-laki	Tidak	Ya	Tidak
5	Negatif	0.782	0.744	Perempuan	Ya	Tidak	Tidak
6	Negatif	0.846	0.570	Perempuan	Tidak	Ya	Ya
7	Negatif	0.244	0.337	Perempuan	Tidak	Ya	Ya
8	Negatif	0.551	0.744	Perempuan	Tidak	Tidak	Tidak
9	Negatif	0.090	0.756	Laki-laki	Tidak	Tidak	Tidak
10	Negatif	0.013	0.512	Laki-laki	Ya	Tidak	Tidak
11	Negatif	0.141	0.407	Perempuan	Tidak	Tidak	Tidak
12	Negatif	0.718	0.593	Laki-laki	Ya	Ya	Ya
13	Negatif	0.602	0.477	Perempuan	Ya	Ya	Ya

Gambar 4. 14. Hasil Pembagian Data

Gambar 4.14 di atas menunjukkan hasil dari proses pembagian data menggunakan RapidMiner. Data sebanyak 350 baris data dari total 500 digunakan untuk proses pelatihan model (training). Data ini akan digunakan oleh algoritma Naive Bayes untuk mempelajari pola hubungan antara atribut input (gejala, suhu tubuh, usia, dll.) dan output (diagnosa malaria). Data Pengujian, Sisanya, 150 baris data, digunakan untuk menguji performa model terhadap data yang belum pernah dilihat sebelumnya. Tahapan ini bertujuan untuk mengetahui kemampuan generalisasi model, serta menghitung akurasi dan metrik evaluasi lainnya.

4.2.4. Implementasi Metode Naïve Bayes

Bagian ini menjelaskan proses Implementasi Metode Naïve Bayes. Ini merupakan tahap di mana algoritma Naïve Bayes diterapkan pada data pelatihan yang telah disiapkan sebelumnya.



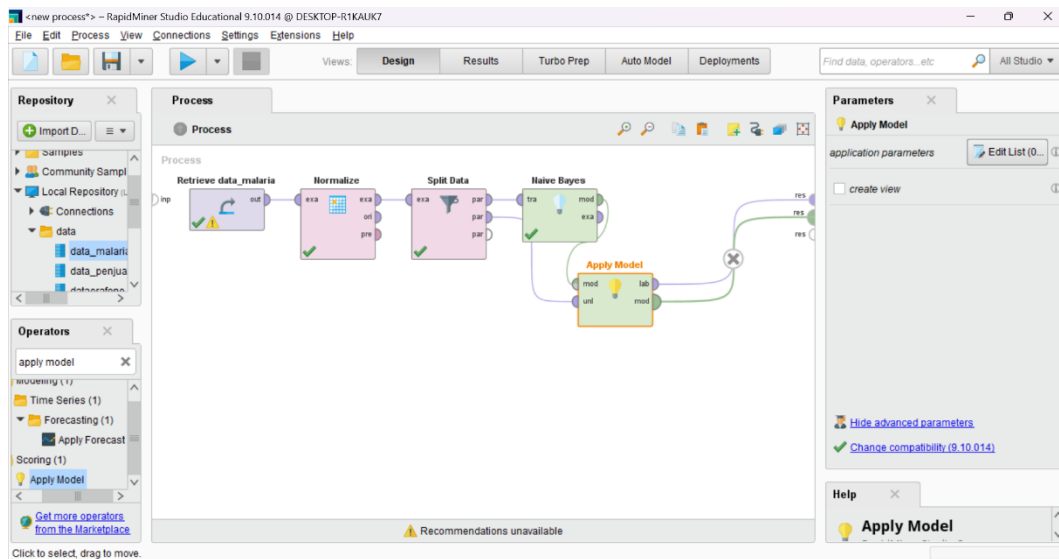
Gambar 4.15. Implementasi Naïve Bayes

Gambar 4.15 menunjukkan proses implementasi algoritma Naïve Bayes dalam RapidMiner Studio yang digunakan dalam penelitian ini. Proses ini

merupakan salah satu tahapan utama dalam pembangunan model klasifikasi untuk prediksi penyakit malaria berdasarkan data pasien Puskesmas Kota Rantauprapat. Implementasi ini dilakukan dengan pendekatan visual workflow yang memanfaatkan serangkaian operator yang saling terhubung dalam tab Design.

Operator utama dalam implementasi ini adalah Naïve Bayes. Operator ini menerima input data latih dari Split Data dan menghasilkan model klasifikasi berbasis probabilistik. Parameter Laplace correction diaktifkan untuk menghindari probabilitas nol pada atribut kategori yang tidak muncul dalam data latih. Model yang dihasilkan selanjutnya akan digunakan untuk melakukan klasifikasi terhadap data uji pada tahap berikutnya.

Gambar ini menunjukkan tahapan penting dalam membangun model klasifikasi menggunakan Naïve Bayes dalam RapidMiner. Dimulai dari pengambilan data, normalisasi, pembagian data, hingga pelatihan model. Implementasi ini mencerminkan pendekatan sistematis dan terstruktur dalam membangun pipeline machine learning, khususnya dalam penelitian yang berkaitan dengan klasifikasi data penyakit malaria atau studi serupa.

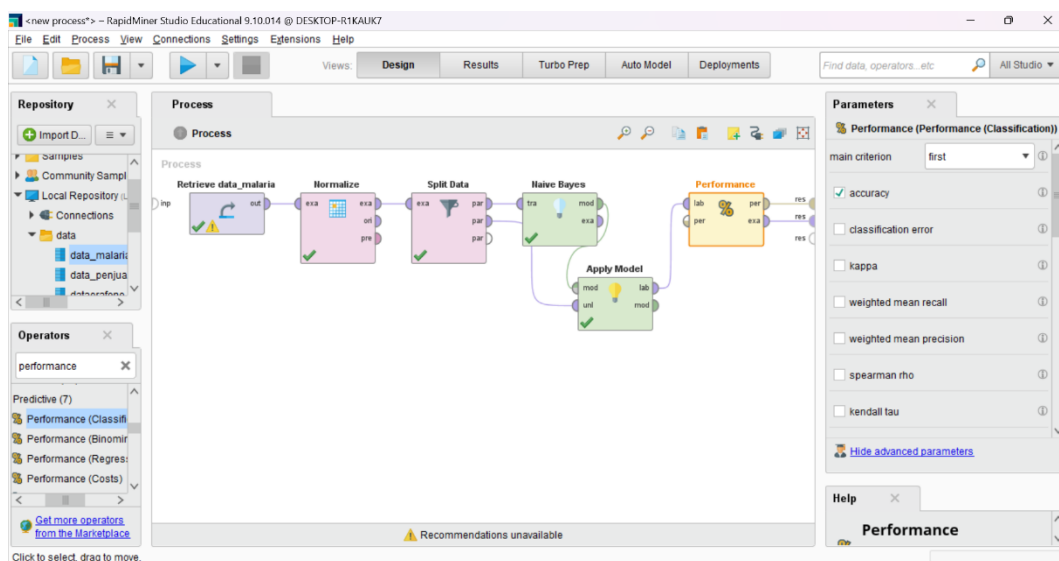


Gambar 4. 16. Apply Model

Gambar 4.16 menunjukkan proses penerapan model (Apply Model) terhadap data pengujian yang telah dibagi sebelumnya. Proses ini merupakan lanjutan dari hasil pelatihan model Naïve Bayes pada tahap sebelumnya. Tujuan utama dari langkah ini adalah untuk menguji kemampuan model dalam melakukan prediksi terhadap data baru yang belum pernah dilihat oleh model selama proses pelatihan.

Adapun operator utama yang digunakan dalam proses ini adalah: Apply Model dan Output Apply Model. Operator Apply Model berfungsi untuk menerapkan model klasifikasi Naïve Bayes yang telah dibangun terhadap subset data uji. Operator ini menerima dua input, yaitu: mod (model) dari output operator Naïve Bayes, dan unl (unlabeled data) yang berasal dari output “testing data” operator Split Data. Hasil dari operator ini adalah data yang telah diprediksi kelasnya oleh model Naïve Bayes berdasarkan atribut-atribut input seperti demam, nyeri otot, suhu tubuh, dan lainnya. Sedangkan Output dari operator Apply Model menghasilkan sebuah dataset baru yang terdiri dari data uji disertai kolom label

hasil prediksi (prediction) dari model. Kolom ini nantinya akan dibandingkan dengan label aktual untuk mengetahui seberapa akurat model dalam memprediksi penyakit malaria. Gambar 4.16 dengan demikian merepresentasikan proses pengujian prediktif yang krusial dalam proses data mining. Tanpa tahap ini, tidak akan diketahui apakah model mampu melakukan klasifikasi dengan baik atau tidak.



Gambar 4. 17. Pengukuran Performance

Gambar 4.17 memperlihatkan proses evaluasi performa model klasifikasi Naïve Bayes. Evaluasi dilakukan dengan menggunakan operator Performance yang mengukur seberapa baik hasil prediksi yang dihasilkan oleh model terhadap data uji, berdasarkan metrik evaluasi klasifikasi. Operator Performance digunakan untuk mengevaluasi hasil prediksi yang telah dihasilkan oleh operator Apply Model. Input dari operator ini adalah hasil prediksi (example set) yang telah dilengkapi dengan kolom label aktual dan kolom prediksi. Operator ini menghitung metrik-metrik penting yang mencakup: Akurasi (Accuracy): Persentase prediksi yang benar dari keseluruhan data uji. Precision: Kemampuan model dalam mengklasifikasikan pasien yang benar-benar terkena malaria secara tepat. Recall (Sensitivity): Seberapa

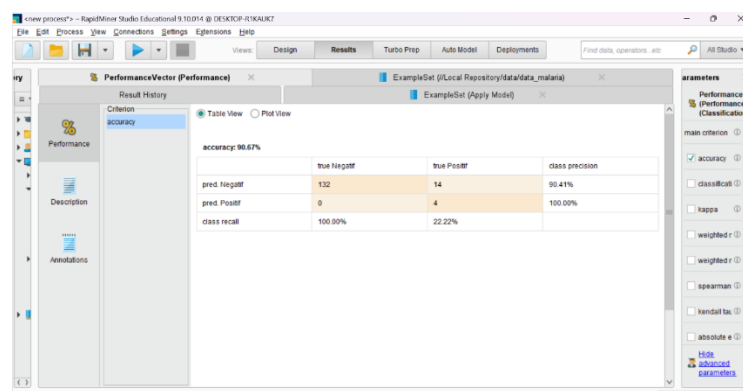
baik model mendeteksi seluruh pasien yang benar-benar menderita malaria. F1-Score: Rata-rata harmonis dari precision dan recall.

Hasil evaluasi akan ditampilkan dalam bentuk tabel atau ringkasan angka performa yang ditampilkan secara otomatis oleh RapidMiner di tab Results. Nilai-nilai metrik ini akan dijadikan dasar dalam menilai apakah model layak digunakan untuk proses prediksi di lingkungan nyata, khususnya untuk mendukung diagnosa awal penyakit malaria di puskesmas.

Dengan adanya tahapan Apply Model dan Performance seperti ditunjukkan dalam Gambar 4.16 dan Gambar 4.17, maka keseluruhan proses klasifikasi Naïve Bayes telah dijalankan secara lengkap. Selanjutnya, hasil evaluasi yang dihasilkan akan dibahas secara lebih rinci dalam subbab Pembahasan, yang akan mencakup interpretasi nilai akurasi dan metrik lainnya serta relevansinya dalam konteks penelitian ini.

4.3. Pembahasan

Hasil dari prediksi yang sudah dilakukan berdasarkan fitur-fitur yang telah ditentukan mendapatkan hasil evaluasi yang cukup bagus yaitu sebagai berikut.



	true Negatf	true Positf	class precision
accuracy	98.62%		
pred Negatf	132	14	90.41%
pred Positf	0	4	100.00%
class recall	100.00%	22.22%	

Gambar 4. 18. Hasil Performance Naïve Bayes

Gambar 4.18 menunjukkan tampilan hasil performa klasifikasi model Naïve Bayes yang telah diterapkan pada data uji melalui platform RapidMiner. Hasil ini disajikan dalam bentuk confusion matrix beserta metrik evaluasi utama, yaitu akurasi, precision, recall, dan f1-score untuk masing-masing kelas (positif dan negatif). Adapun rincian hasil klasifikasi adalah sebagai berikut:

1. Total data uji: 150 record
2. Jumlah True Negative (TN): 132
3. Jumlah False Positive (FP): 0
4. Jumlah False Negative (FN): 14
5. Jumlah True Positive (TP): 4

Berdasarkan data di atas, dapat disusun confusion matrix seperti diperlihatkan pada tabel 4.3.

Tabel 4. 3. Confusion Matrix

	Aktual Negatif	Aktual Positif
Prediksi Negatif	132 (TN)	14 (FN)
Prediksi Positif	0 (FP)	4 (TP)

1. Dari total 150 data uji, model berhasil memprediksi 136 data dengan benar, terdiri dari:
 1. 132 data negatif yang diklasifikasikan secara benar (True Negative).
 2. 4 data positif yang diklasifikasikan secara benar (True Positive).
2. Terdapat 14 data aktual positif yang diprediksi sebagai negatif (False Negative).

3. Tidak terdapat kesalahan klasifikasi positif terhadap data negatif (False Positive = 0).

Evaluasi lebih dalam terhadap confusion matrix menunjukkan bahwa model memiliki tingkat akurasi yang tinggi, namun recall untuk kelas positif sangat rendah. Hal ini berarti model cenderung mengklasifikasikan data ke kelas negatif (yaitu tidak terkena malaria), dan kesulitan mengenali data positif (terkena malaria). Walaupun precision untuk kelas positif mencapai 100% (tidak ada False Positive), recall-nya hanya 22,22%, karena hanya 4 dari total 18 kasus positif yang berhasil dikenali model.

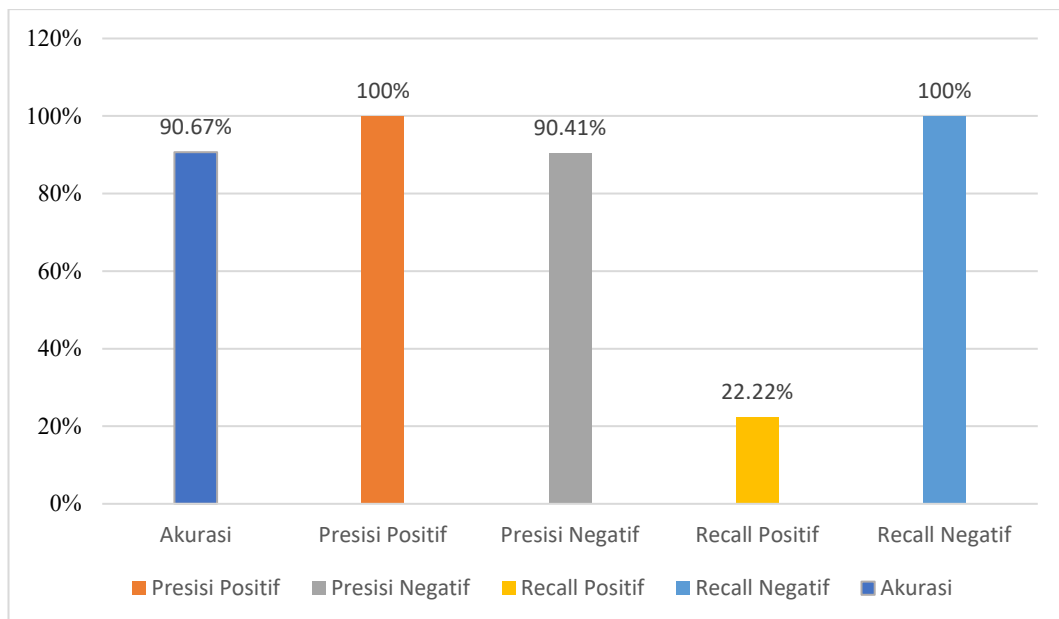
Selanjutnya, beberapa metrik performa dapat dihitung dan diinterpretasikan sebagai berikut:

1. Akurasi: menunjukkan proporsi prediksi yang benar terhadap seluruh prediksi. Berdasarkan hasil pengujian, diperoleh nilai akurasi sebesar 90,67%, yang artinya model berhasil mengklasifikasikan 90,67% data uji dengan tepat.
2. Presisi: untuk kelas positif (penderita malaria) adalah 100%, yang berarti seluruh prediksi positif benar-benar sesuai dengan label aktual positif. Artinya, tidak ada kesalahan dalam mengklasifikasikan pasien yang diprediksi mengidap malaria.
3. Recall (Sensitivity): untuk kelas positif sebesar 22,22%, menunjukkan bahwa hanya sekitar 22% dari seluruh pasien yang benar-benar mengidap malaria berhasil dikenali oleh model. Hal ini mengindikasikan adanya kelemahan model dalam mendeteksi seluruh kasus positif (terjadi 14 FN).

4. F1-Score (tidak ditampilkan secara eksplisit dalam gambar): dengan precision 100% dan recall 22.22%, nilai F1-score dapat dihitung sebagai berikut:

$$\begin{aligned}\text{F1-score} &= 2 \times (\text{Precision} \times \text{Recall}) / (\text{Precision} + \text{Recall}) \\ &= 2 \times (1 \times 0.2222) / (1 + 0.2222) \\ &\approx 0.3636 \text{ atau } 36.36\%\end{aligned}$$

Hal ini menunjukkan bahwa walaupun precision tinggi, namun karena recall rendah, maka F1-score sebagai metrik harmonis berada di angka sedang.



Gambar 4. 19. Metrik Evaluasi Naïve Bayes

Gambar 4.19 di atas merupakan visualisasi hasil evaluasi model klasifikasi penyakit malaria menggunakan algoritma Naïve Bayes dalam bentuk diagram batang. Grafik ini menampilkan empat metrik utama untuk menilai performa model, yaitu Akurasi, Presisi, Recall, dan F1-Score.

Akurasi model tercatat sebesar 90,67%. Angka ini menunjukkan bahwa dari seluruh data uji, sebanyak 90,67% prediksi yang dihasilkan oleh model adalah benar (baik prediksi positif maupun negatif). Presisi mencapai 100%, artinya seluruh kasus yang diprediksi sebagai positif (pasien terinfeksi malaria) oleh model memang benar-benar positif. Ini menunjukkan bahwa tidak ada kesalahan dalam mengklasifikasikan pasien yang tidak terinfeksi sebagai terinfeksi. Recall hanya sebesar 22,22%, yang berarti dari seluruh pasien yang benar-benar terinfeksi malaria, hanya 22,22% yang berhasil dikenali oleh model. Nilai ini cukup rendah dan menunjukkan bahwa sebagian besar pasien terinfeksi tidak berhasil terdeteksi (False Negative tinggi). F1-Score, yaitu rata-rata harmonis dari presisi dan recall, tercatat sebesar 36,36%. Nilai ini menunjukkan bahwa meskipun model sangat presisi, tetapi karena recall-nya rendah, maka secara keseluruhan kemampuan model dalam mendeteksi kelas positif masih kurang optimal.

Secara keseluruhan, grafik ini memperjelas bahwa meskipun Naïve Bayes mampu menghasilkan prediksi yang sangat tepat ketika memutuskan kasus positif, namun kelemahan model terletak pada sensitivitasnya dalam mengenali seluruh kasus positif yang sebenarnya. Oleh karena itu, diperlukan upaya peningkatan recall melalui perbaikan data atau penggunaan metode balancing agar model dapat lebih efektif dalam mendeteksi seluruh pasien yang terinfeksi malaria.