

## **BAB IV**

### **HASIL DAN PEMBAHASAN**

#### **A. Hasil**

##### **1. Gambar Umum Rumah Sakit**

RSU PKU Muhammadiyah Bantul adalah rumah sakit umum tipe C yang berlokasi di Jalan Jenderal Sudirman No.124, Bantul, Daerah Istimewa Yogyakarta. Berdiri sejak 1 Maret 1966, rumah sakit ini telah lama memberikan pelayanan kesehatan kepada masyarakat. Seiring perkembangannya, rumah sakit ini terus meningkatkan kapasitas layanan serta sistem pengelolaan informasinya, khususnya dalam manajemen data rekam medis pasien. Sebagai salah satu rumah sakit yang sudah menggunakan Rekam Medis Elektronik (RME), RSU PKU Muhammadiyah Bantul memiliki sistem informasi kesehatan yang terintegrasi. Dengan adanya sistem tersebut, proses pencatatan, dokumentasi, dan pelaporan data pasien dapat dilakukan dengan lebih cepat, efisien, dan akurat. Data digital yang tersimpan juga memudahkan pengambilan keputusan medis berbasis informasi serta mendukung analisis epidemiologi maupun penelitian ilmiah.

Proses pengolahan data di RSU PKU Muhammadiyah Bantul umumnya dimulai dari pengumpulan data pasien yang berasal dari berbagai unit pelayanan, seperti rawat jalan, rawat inap, laboratorium, dan instalasi lainnya. Semua data dicatat secara elektronik melalui sistem Rekam Medis Elektronik (RME), sehingga informasi pasien dapat terdokumentasi dengan cepat, akurat, dan terstruktur. Setelah terkumpul, data kemudian diolah melalui beberapa tahap, antara lain verifikasi, klasifikasi, dan rekapitulasi untuk mendukung kebutuhan operasional rumah sakit. Tahap ini juga mencakup analisis data yang bermanfaat untuk evaluasi mutu pelayanan, penyusunan laporan internal maupun eksternal, serta proses klaim biaya ke pihak terkait, seperti BPJS Kesehatan. Pada tahap akhir, data yang sudah diolah dijadikan dasar dalam pengambilan keputusan manajerial, pembuatan laporan statistik, dan evaluasi

layanan. Dengan sistem pengolahan data yang terintegrasi, rumah sakit tidak hanya dapat meningkatkan efisiensi kerja, tetapi juga menjamin ketepatan informasi serta kualitas pelayanan kesehatan secara keseluruhan.

Penelitian ini menggunakan data sebanyak 1.122 kasus pasien rawat inap dengan diagnosis Diabetes Melitus Tipe II selama periode Januari hingga Desember 2024. Pengolahan data dilakukan dengan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) dan dibantu oleh perangkat lunak *RapidMiner*, sehingga proses transformasi data dapat berlangsung secara sistematis. Pemanfaatan rekam medis elektronik di RSUD Muhammadiyah Bantul sebagai sumber utama penelitian menunjukkan kesiapan rumah sakit dalam mengaplikasikan data mining untuk kepentingan klinis maupun manajerial. Selain itu, penelitian ini juga menegaskan peran penting petugas manajemen informasi kesehatan (PMIK) dalam mengelola data strategis yang berguna untuk mendukung pencegahan penyakit serta peningkatan mutu pelayanan kesehatan.

## 2. Proses Pengolahan Data dengan CRISP-DM

Penelitian ini menggunakan algoritma *decision tree* untuk melakukan prediksi kasus pasien Diabetes Melitus Tipe II berdasarkan data rekam medis dari RSUD Muhammadiyah Bantul. Data yang digunakan merupakan data sekunder berupa data pasien rawat inap Diabetes Melitus Tipe II periode Januari hingga Desember 2024. Data dibagi menjadi dua bagian, yaitu 80% data latih (800 data) dan 20% data uji (200 data). Pengolahan dan analisis data dilakukan menggunakan aplikasi *RapidMiner Version 10.3*. Penelitian ini menggunakan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri atas lima tahapan utama dalam penelitian ini sebagai berikut:

### a. *Business Understanding*

Pada penelitian ini di dapatkan data rekam medis pasien Diabetes Melitus Tipe II di RSUD Muhammadiyah Bantul sebagai berikut:

No Reg	No Ctn	Nama	Usia	JK	Tensi Atas	Tensi Bawah	Tinggi Badan	Berat Badan	GDS
3081774	10*****	A001	46	L	188	112	1	0	314
3244200	10*****	A002	72	L	223	104	165	83	216
3245749	10*****	A003	58	L	177	95	165	70	288
3311327	10*****	A004	33	P	125	80	155	0	158
3297037	10*****	A005	76	L	116	50	165	62	null
3363598	10*****	A006	75	L	102	86	155	50	203
3379051	10*****	A007	41	L	201	104	165	90	754
3192099	10*****	A008	44	L	143	88	1	1	510
3274149	10*****	A009	50	L	143	83	167	67	427
3341015	10*****	A010	70	L	155	78	1	0	304
3087592	10*****	A011	49	L	189	89	165	75	222
3213972	10*****	A012	68	L	174	84	160	65	212
3246024	10*****	A013	63	L	144	72	1	1	156
3223262	10*****	A014	65	L	147	98	160	60	170
3252211	10*****	A015	30	L	125	80	163	35	264
3334900	10*****	A016	71	P	141	85	150	55	711
3148249	10*****	A017	60	P	null	null	0	0	null
3091260	10*****	A018	46	L	133	78	160	60	311
3146107	10*****	A019	48	P	177	84	168	97	null

**Gambar 4.1 Data Rekam Medis Elektronik Pasien DM**

Tahapan ini bertujuan untuk memahami secara mendalam tujuan dari penelitian. RSUD Muhammadiyah Bantul mengalami peningkatan jumlah kasus Diabetes Melitus Tipe II secara signifikan, yaitu sebanyak 1122 kasus periode Januari - Desember 2024, menjadikannya penyakit keempat terbanyak di rumah sakit tersebut. Tujuan utama dari penelitian ini adalah untuk membantu rumah sakit dalam melakukan prediksi potensi Diabetes Melitus Tipe II menggunakan algoritma *Decision Tree*. Dengan hasil prediksi ini, rumah sakit diharapkan dapat melakukan deteksi dini terhadap pasien berisiko tinggi, mengembangkan intervensi yang lebih terarah, menunjang pengambilan keputusan medis berbasis data, serta meningkatkan efisiensi dalam pelayanan kesehatan dan penanganan penyakit kronis.

b. *Data Understanding*

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	No Reg	No Cmi	Nama	Usia	JK	Tensi Atas	Tensi Bawah	Tinggi Badan	Berat Badan	GDS							
2	3081774	10*****	A001	46	L	188	112	1	0	314							
3	3244200	10*****	A002	72	L	223	104	165	83	216							
4	3245749	10*****	A003	58	L	177	95	165	70	288							
5	3311327	10*****	A004	33	P	125	80	155	0	158							
6	3297037	10*****	A005	76	L	116	50	165	62	null							
7	3363598	10*****	A006	75	L	102	86	155	50	203							
8	3379051	10*****	A007	41	L	201	104	165	90	754							
9	3192099	10*****	A008	44	L	143	88	1	1	510							
10	3274149	10*****	A009	50	L	143	83	167	67	427							
11	3341015	10*****	A010	70	L	155	78	1	0	304							
12	3087592	10*****	A011	49	L	189	89	165	75	222							
13	3213972	10*****	A012	68	L	174	84	160	65	212							
14	3246024	10*****	A013	63	L	144	72	1	1	156							
15	3222622	10*****	A014	65	L	147	98	160	60	170							
16	3252211	10*****	A015	30	L	125	80	163	35	264							
17	3334900	10*****	A016	71	P	141	85	150	55	711							
18	3148249	10*****	A017	60	P	null	null	0	0	null							
19	3091260	10*****	A018	46	L	133	78	160	60	311							
20	3146107	10*****	A019	48	P	177	84	168	97	null							

**Gambar 4.2** Atribut-atribut yang akan digunakan

Keterangan: data yang diberikan tanda berwarna hijau merupakan atribut yang akan digunakan dalam membangun pemodelan *decision tree*.

Pada tahap ini dilakukan eksplorasi terhadap data yang akan digunakan. Data bersumber dari rekam medis elektronik pasien rawat inap yang terdiagnosis Diabetes Melitus Tipe II pada RSUD Muhammadiyah Bantul selama periode Januari - Desember 2024, sebanyak 1122 data pasien. Data yang tersedia mencakup atribut-atribut berikut:

- 1) Nomor registrasi pasien
- 2) Nomor rekam medis pasien
- 3) Nama pasien
- 4) Usia (remaja, dewasa, dan lansia)
- 5) Jenis kelamin (laki-laki dan perempuan)
- 6) Tekanan darah sistolik dan diastolik
- 7) Indeks Massa Tubuh (IMT) (*underweight*, normal, *overweight*, obesitas I dan obesitas II)
- 8) Gula Darah Sewaktu (GDS) (normal, cukup tinggi, dan tinggi)

Setiap atribut ditinjau dan dianalisis distribusinya, *outlier*, serta nilai kosong (*missing values*) untuk memahami kualitas dan karakteristik data. Namun, pada penelitian ini atribut yang digunakan untuk prediksi kasus

Diabetes Melitus Tipe II hanya atribut usia, jenis kelamin, tekanan darah (sistolik dan diastolik), IMT, dan GDS.

c. *Data Preparation*

Tahapan ini meliputi pembersihan, transformasi, dan pengkodean data agar siap untuk dimodelkan. Proses yang dilakukan mencakup:

- 1) *Data Cleaning*: Menghapus data yang tidak dipakai, data duplikat dan tidak lengkap, seperti kolom tekanan darah atau IMT yang kosong, dan kolom atribut yang tidak dipakai. Pada tahap ini dilakukan pembersihan data yang tidak sesuai, data awal berjumlah 1122 sampel, kemudian setelah dilakukan *cleaning* data didapatkan 1000 data yang siap digunakan dalam pemodelan decision tree. Berikut merupakan gambar dari hasil proses *cleaning* data:

	A	B	C	D	E	F	G	H	I	J
1	No Reg	No Cia	Nama	Usia	JK	Tensi Atas	Tensi Bawah	Tinggi Badan	Berat Badan	GDS
2	3081774	10*****	A001	46	L	186	112	71	6	314
3	3244200	10*****	A002	72	L	223	104	165	83	216
4	3245760	10*****	A003	58	L	179	95	163	70	288
5	3311327	10*****	A004	53	P	165	86	155	6	158
6	3297027	10*****	A005	76	L	116	50	165	62	null
7	3363398	10*****	A006	75	L	102	86	155	50	203
8	3379053	10*****	A007	41	L	201	104	165	90	754
9	3192099	10*****	A008	44	L	145	86	165	1	300
10	3274149	10*****	A009	50	L	143	83	167	67	427
11	3341013	10*****	A010	70	L	165	76	165	74	204
12	3087902	10*****	A011	49	L	189	89	165	75	222
13	3213975	10*****	A012	68	L	174	84	160	65	212
14	3246024	10*****	A013	63	L	184	72	1	1	156
15	3223262	10*****	A014	65	L	147	98	160	60	170
16	3252213	10*****	A015	30	L	125	80	163	35	264
17	3348990	10*****	A016	71	P	141	85	150	55	711
18	3148249	10*****	A017	60	P	null	null	0	0	null
19	3091260	10*****	A018	46	L	133	78	160	60	311
20	3146160	10*****	A019	48	P	109	84	168	97	null

Gambar 4.3 Data Cleaning

Keterangan: kolom yang berwarna merah adalah kolom yang akan dihapus atau tidak terpakai pada tahap *cleaning* data.

- 2) *Data Transformation*: Mengubah satuan dan format data menjadi konsisten.

No Reg	No Ctn	Nama	Usia	Tensi Atas	Tensi Bawah	Tinggi Badan (cm)	Berat Badan (Kg)	IMT	GDS
3081774	A001	A001	46	188	112	0.01	0	mll	314
3244200	A002	A002	72	223	104	1.65	65	83	30.49
3243749	A003	A003	58	177	95	1.65	163	70	25.71
3311327	A004	A004	33	125	80	1.55	155	0	0.00
3297037	A005	A005	76	116	50	1.65	165	62	22.77
3363598	A006	A006	75	102	86	1.55	155	50	20.81
3379051	A007	A007	41	201	104	1.65	165	90	33.06
3192099	A008	A008	44	143	88	0.01	1	1	mll
3274149	A009	A009	50	143	83	1.67	167	67	24.02
3340105	A010	A010	70	155	78	0.01	1	0	0.00
3087592	A011	A011	49	189	89	1.65	165	75	27.55
3213972	A012	A012	68	174	84	1.6	160	65	25.39
3246024	A013	A013	63	144	72	0.01	1	1	mll
322302	A014	A014	65	147	98	1.6	160	60	23.44
3252211	A015	A015	30	125	80	1.63	163	35	13.17
3334900	A016	A016	71	141	85	1.5	150	55	24.44
3148249	A017	A017	60	mll	mll	0	0	0	mll
3091260	A018	A018	46	133	78	1.6	160	60	23.44

Gambar 4.4 Transformasi Data

Pada tahap ini dilakukan transformasi data dengan mengubah satuan tinggi badan yang semula centimeter (cm), menjadi meter (m). Hal tersebut dilakukan untuk menghitung Indeks Massa Tubuh (IMT).

- 3) *Data Encoding*: Mengubah data kategorikal menjadi numerik untuk proses analisis, contohnya:

No Reg	No Ctn	Nama	Usia	Kategori Usia	JK	Tensi Atas	Tensi Bawah	Tinggi Badan (cm)	Tinggi Badan (m)	Berat Badan (Kg)	IMT	Kategori GDS	Kategori GDS
3081774	A001	A001	46	2	1	188	112	0.01	1	0	mll	314	3
3244200	A002	A002	72	3	1	223	104	1.65	1.65	83	30.49	5	216
3243749	A003	A003	58	2	1	177	95	1.65	1.65	70	25.71	4	288
3311327	A004	A004	33	1	1	125	80	1.55	1.55	0	0.00	1	158
3297037	A005	A005	76	3	1	116	50	1.65	1.65	62	22.77	2	203
3363598	A006	A006	75	3	1	102	86	1.55	1.55	50	20.81	2	203
3379051	A007	A007	41	2	1	201	104	1.65	1.65	90	33.06	5	754
3192099	A008	A008	44	2	1	143	88	0.01	1	1	mll	mll	510
3274149	A009	A009	50	2	1	143	83	1.67	1.67	67	24.02	3	427
3340105	A010	A010	70	3	1	155	78	0.01	1	0	0.00	1	304
3087592	A011	A011	49	2	1	189	89	1.65	1.65	75	27.55	4	222
3213972	A012	A012	68	3	1	174	84	1.6	1.60	65	25.39	4	212
3246024	A013	A013	63	3	1	144	72	0.01	1	1	mll	mll	156
322302	A014	A014	65	3	1	147	98	1.6	1.60	60	23.44	3	170
3252211	A015	A015	30	2	1	125	80	1.63	1.63	35	13.17	1	264
3334900	A016	A016	71	3	2	141	85	1.5	1.50	55	24.44	3	711
3148249	A017	A017	60	3	2	mll	mll	0	0	0	mll	mll	3
3091260	A018	A018	46	2	1	133	78	1.6	1.60	60	23.44	3	311

Gambar 4.5 Data Encoding

- Usia: Remaja (1), Dewasa (2), Lansia (3)
- Jenis Kelamin: Laki-laki (1), Perempuan (2)
- Tekanan Darah: Normal (1), Meningkat (2), Hipertensi Tingkat I (3), Hipertensi Tingkat II (4)

- d) IMT: *Underweight* (1), *Normal* (2), *Overweight* (3), *Obesitas I* (4), *Obesitas II* (5)
- e) GDS: *Normal* (1), *Cukup Tinggi* (2), *Tinggi* (3)
- 4) *Data Splitting*

Data yang sudah diproses selanjutnya dibagi menjadi dua bagian, yaitu data latih yang akan digunakan untuk membangun model klasifikasi, dan data uji yang digunakan untuk mengevaluasi kinerja model.

- a) Data latih: 80% (800 data)

Gambar 4.6 Data Latih 80%

- b) Data uji: 20% (200 data)

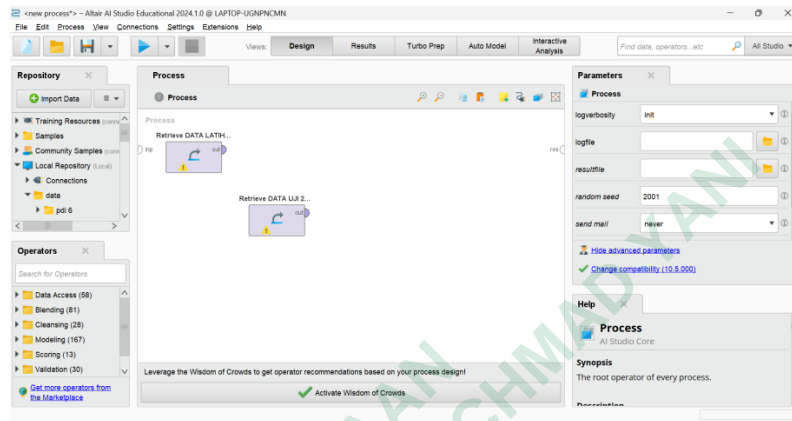
Gambar 4.7 Data Uji 20%

- d. *Modeling*

Model dibangun menggunakan algoritma *Decision Tree* dengan bantuan perangkat lunak *RapidMiner Version 10.3*. Seluruh variabel input

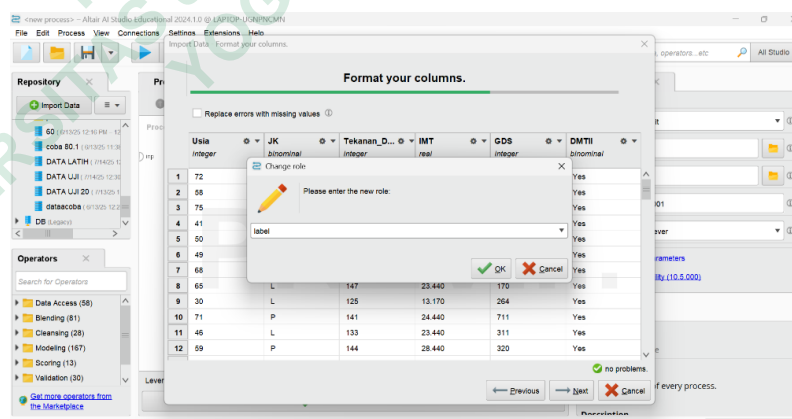
digunakan sebagai fitur prediktif untuk menentukan klasifikasi risiko Diabetes Melitus Tipe II. Proses *modeling* meliputi:

- 1) Memasukkan data latih (80%) dan data uji (20%) pada RapidMiner



**Gambar 4.8 Proses Modelling Input Data Latih dan Data Uji**

Pada tahap ini, data latih dan data uji dimasukkan ke dalam *workspace* RapidMiner dengan memanfaatkan operator *Read Excel* yang terdapat pada menu Operators. Operator ini berfungsi untuk mengimpor *file Excel* yang berisi dataset yang telah disiapkan sebelumnya, baik untuk keperluan pelatihan model maupun untuk proses pengujian.

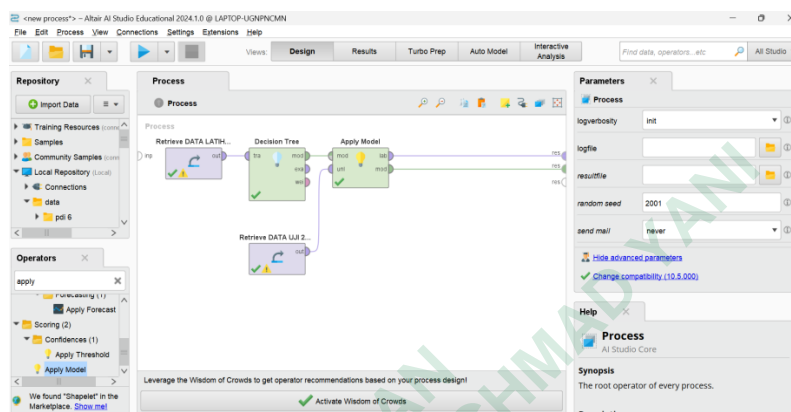


**Gambar 4.9 Penentuan Label**

Langkah berikutnya adalah mengatur peran variabel Diabetes Melitus Tipe II (DMTII) sebagai label menggunakan menu *change role*, agar

variabel tersebut dapat dijadikan target klasifikasi dalam proses pelatihan model.

2) Memasukkan *decision tree* dan *apply model*



**Gambar 4.10 Penerapan Pemodelan Decision Tree dan Apply Model**

Langkah berikutnya dalam proses klasifikasi adalah menempatkan operator *decision tree* dan *apply model* ke dalam *workspace*. Sebelum proses dijalankan, parameter algoritma *decision tree* disesuaikan dengan mengganti metode pemilihan atribut (*criterion*) dari *gain ratio* menjadi *information gain*. Perubahan ini dilakukan agar model lebih memprioritaskan atribut yang mampu mengurangi *entropi* secara signifikan dalam pembentukan pohon keputusan. Setelah konfigurasi selesai dilakukan, proses dijalankan untuk memprediksi pasien yang terindikasi Diabetes Melitus Tipe II pada data uji.

### 3) Hasil klasifikasi

Row No.	DMTI	prediction(DMTI)	confidence(Yes)	confidence(No)	Usia	JK	Tekanan_D...	IMT	GDS
1	?	Yes	1	0	56	L	95	27.470	185
2	?	Yes	1	0	56	L	140	24.800	335
3	?	Yes	1	0	56	L	139	27.430	413
4	?	Yes	0.972	0.028	59	P	118	27.340	632
5	?	Yes	0.972	0.028	56	L	125	23.310	422
6	?	Yes	1	0	74	P	172	20.810	504
7	?	Yes	1	0	73	P	159	17.780	289
8	?	Yes	1	0	66	P	152	65.780	122
9	?	Yes	1	0	63	P	194	20.810	242
10	?	Yes	1	0	72	P	158	21.640	274
11	?	Yes	0.755	0.245	62	P	126	25.390	189
12	?	Yes	1	0	60	P	156	15.020	434
13	?	Yes	1	0	65	P	145	18.610	369

**Gambar 4.11 Hasil Klasifikasi**

Setelah kriteria pemisahan pada algoritma *Decision Tree* diubah menjadi *Information Gain*, model digunakan untuk menguji 200 data pasien. Hasil prediksi ditampilkan dalam bentuk label prediksi (DMTII) serta tingkat keyakinan (*confidence*) terhadap masing-masing kelas. Sebagian besar hasil menunjukkan tingkat keyakinan yang tinggi, seperti pada prediksi kelas 'Yes' dengan nilai *confidence* mencapai 1.000. Meski demikian, terdapat beberapa kasus dengan tingkat keyakinan yang lebih rendah, contohnya pada baris ke-4 yang menunjukkan *confidence* 0.972 untuk kelas 'Yes' dan 0.028 untuk 'No', serta baris ke-9 dengan *confidence* 0.755 untuk 'Yes' dan 0.245 untuk 'No'. Hal ini mengindikasikan bahwa meskipun model umumnya memberikan prediksi yang cukup meyakinkan, terdapat beberapa data yang mendekati batas ambang klasifikasi.

#### e. *Evaluation*

Tujuan dari evaluasi model adalah untuk mengukur sejauh mana tingkat keakuratan model klasifikasi yang telah dibangun dalam memprediksi pasien Diabetes Melitus Tipe II berdasarkan data rekam medis di RSUD Muhammadiyah Bantul. Proses evaluasi dilakukan menggunakan metode *confusion matrix* sebagai alat penilaian, dengan menghitung nilai

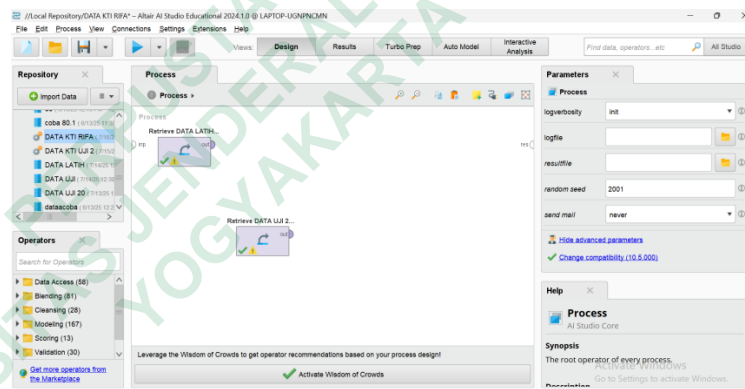
akurasi (*accuracy*), presisi (*precision*), sensitivitas (*recall*), dan *F1-score* pada masing-masing kategori prediksi.

Model klasifikasi yang digunakan adalah algoritma *Decision Tree*, dan evaluasi dilakukan menggunakan data uji sebanyak 200 data pasien. Proses evaluasi ini dilaksanakan melalui aplikasi *RapidMiner versi 10.3*. Hasil evaluasi yang diperoleh yaitu sebagai berikut:

### 1) *Accuracy*

Evaluasi akurasi dilakukan untuk menilai sejauh mana ketepatan model dalam memprediksi kasus Diabetes Melitus Tipe II. Nilai akurasi diperoleh dengan menghitung persentase prediksi yang sesuai dibandingkan dengan keseluruhan data uji. Adapun tahapan yang dilakukan dalam proses evaluasi akurasi meliputi:

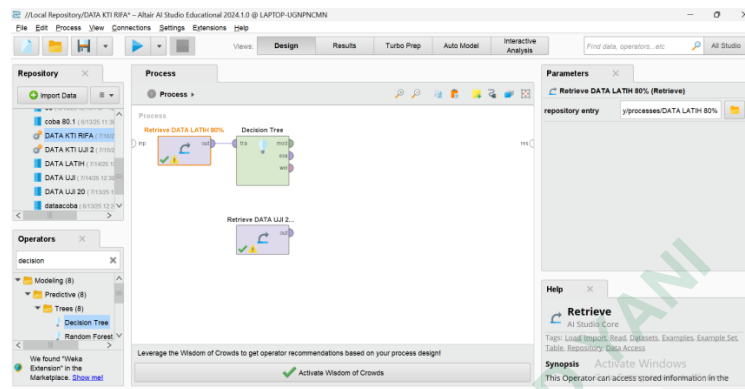
#### a) *Input Data*



**Gambar 4.12** *Input Data Latih dan Data Uji untuk Evaluasi*

Pembagian data menjadi data latih (80%) dan data uji (20%) dilakukan dengan tujuan untuk membangun model klasifikasi serta mengukur tingkat akurasinya, sehingga model dapat mengenali pola dengan baik dan memberikan prediksi yang optimal terhadap pasien DMTII pada data yang belum dikenali sebelumnya.

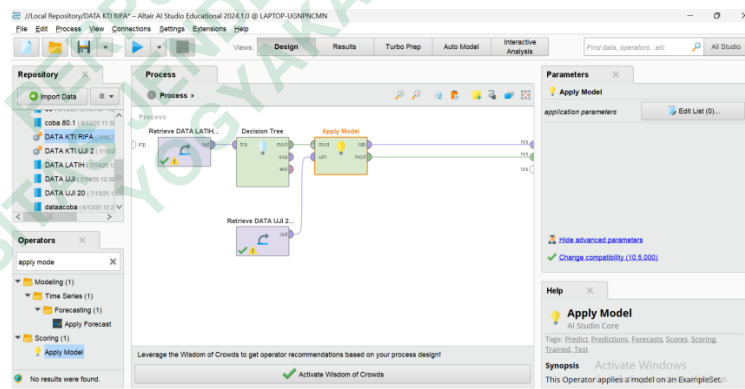
b) Pemodelan *Decision Tree*



**Gambar 4.13** Pemodelan *Decision Tree*

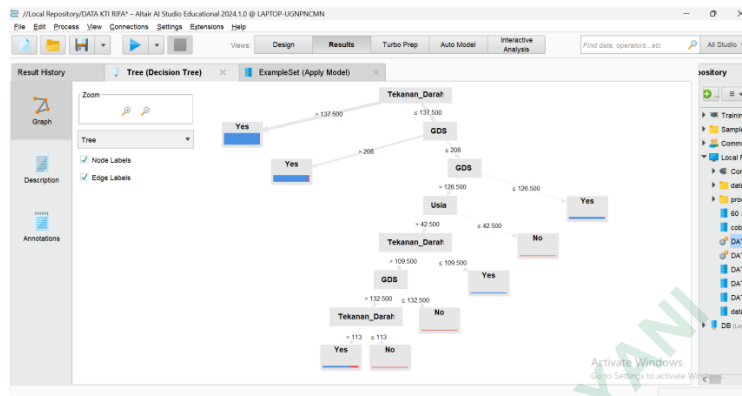
Algoritma *Decision Tree* ditambahkan untuk membangun model klasifikasi berdasarkan pola yang ditemukan dalam data latih, sehingga dapat menghasilkan aturan-aturan keputusan yang berguna dalam memprediksi pasien Diabetes Melitus Tipe II.

c) *Apply Model*



**Gambar 4.14** Penerapan *Apply Model*

Langkah berikutnya adalah menambahkan algoritma *Apply Model* dan menjalankannya, yang berfungsi untuk menerapkan model klasifikasi *Decision Tree* yang telah dibuat pada data uji. Dengan demikian, model dapat menghasilkan pohon keputusan prediksi pasien DMTII berdasarkan pola yang telah dipelajari dari data latih.



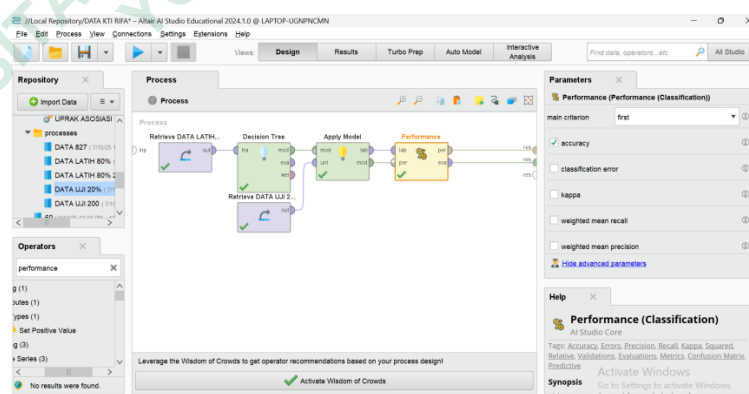
**Gambar 4.15 Pohon Keputusan**

Pohon keputusan pada gambar di atas digunakan untuk membantu menentukan status kondisi seseorang berdasarkan beberapa atribut, seperti tekanan darah, kadar gula darah sewaktu (GDS), dan usia. Model ini bekerja dengan mengikuti alur dari atas ke bawah, dimulai dari akar pohon hingga mencapai daun, yaitu hasil akhir klasifikasi berupa “Yes” atau “No”. Setiap percabangan menunjukkan syarat atau kondisi nilai dari suatu atribut. Proses klasifikasi pertama-tama dimulai dari atribut Tekanan Darah. Jika nilainya lebih dari 137,5, maka model langsung memutuskan hasil akhir “Yes”. Namun jika Tekanan Darah kurang dari atau sama dengan 137,5, maka sistem melanjutkan pemeriksaan ke atribut GDS. Jika GDS kurang dari atau sama dengan 206, model akan kembali memeriksa atribut GDS kedua dengan batas nilai 208. Selanjutnya, jika GDS masih memenuhi kondisi  $\leq 208$ , maka atribut Usia akan diperiksa. Jika Usia  $\leq 52,5$  tahun, sistem akan mengevaluasi kembali atribut Tekanan Darah pada batas 100,5. Jika nilainya lebih tinggi, maka hasil akhir adalah “Yes”; jika lebih rendah, maka atribut GDS akan diperiksa kembali dengan batas 101,5, dan begitu seterusnya hingga keputusan akhir “Yes” atau “No” ditemukan di bagian bawah pohon.

Sebagai contoh, jika seorang pasien memiliki tekanan darah 165 mmHg, maka karena nilainya lebih dari 137,5, sistem langsung mengklasifikasikannya sebagai “Yes”, yaitu berisiko atau mengalami kondisi yang dimaksud. Namun jika pasien memiliki tekanan darah 120 mmHg, GDS 190, usia 45 tahun, dan nilai-nilai berikutnya sesuai alur yang dilalui dalam pohon, maka keputusan akhirnya juga bisa berujung pada “Yes”. Sebaliknya, jika pasien memiliki tekanan darah 120 mmHg, GDS 190, usia 55 tahun, dan nilai GDS selanjutnya tinggi, maka model akan mengarahkan pada keputusan “No”. Dari penjabaran ini dapat disimpulkan bahwa pohon keputusan sangat membantu dalam proses klasifikasi karena memberikan urutan logis dan sistematis yang bisa ditelusuri dengan mudah, sehingga hasil akhirnya dapat dipahami oleh pengguna tanpa memerlukan perhitungan kompleks.

d) *Performance*

Evaluasi kinerja model dilakukan dengan menggunakan operator *Performance*, dengan akurasi (*accuracy*) sebagai parameter penilaiannya.



**Gambar 4.16 Penerapan Operator Performance**

Penambahan algoritma *Performance* bertujuan untuk menilai kinerja model klasifikasi yang telah dikembangkan dengan menghitung persentase prediksi yang tepat dari seluruh data uji,

sehingga dapat diketahui sejauh mana model mampu memprediksi kasus DMFTII secara akurat dan konsisten.

accuracy: 93.50% +/- 3.37% (micro average: 93.50%)

	true Yes	true No	class precision
pred. Yes	159	8	95.21%
pred. No	5	28	84.85%
class recall	96.95%	77.78%	

**Gambar 4.17 Hasil Nilai Accuracy**

Berdasarkan hasil evaluasi, model klasifikasi *Decision Tree* menghasilkan nilai akurasi sebesar 93,50%, yang berarti bahwa 93,50% dari seluruh data uji berhasil diklasifikasikan dengan benar.

accuracy: 93.50% +/- 3.37% (micro average: 93.50%)

	true Yes	true No	class precision
pred. Yes	159	8	95.21%
pred. No	5	28	84.85%
class recall	96.95%	77.78%	

**Gambar 4.18 Hasil Prediksi**

*Confusion matrix* di atas menunjukkan distribusi hasil klasifikasi model terhadap data uji. Angka pada diagonal menunjukkan jumlah prediksi yang benar sesuai kelas sebenarnya, sedangkan angka di luar diagonal merepresentasikan prediksi yang salah. Misalnya, sebanyak 159 pasien yang sebenarnya menderita Diabetes Melitus Tipe II (kelas “Yes”) berhasil diklasifikasikan dengan benar, sedangkan 8 pasien yang seharusnya termasuk kelas “No” salah diprediksi sebagai “Yes”. Untuk kelas “No”, sebanyak 28 pasien berhasil diklasifikasikan dengan benar, namun 5 pasien yang sebenarnya “Yes” salah diprediksi sebagai “No”. Hasil ini menunjukkan bahwa model memiliki performa klasifikasi yang cukup baik.

2) *Precision*

accuracy: 93.50% +/- 3.37% (micro average: 93.50%)

	true Yes	true No	class precision
pred. Yes	159	8	95.21%
pred. No	5	28	84.85%
class recall	96.95%	77.78%	

**Gambar 4.19 Hasil Precision**

Model menghasilkan nilai *precision* sebesar 95,21% untuk kelas 'Yes' (penderita DMTII) dan 84,85% untuk kelas 'No' (bukan penderita DMTII), yang menandakan bahwa model memiliki tingkat akurasi yang baik dalam menentukan status Diabetes Melitus Tipe II pada pasien. Tingginya *precision* pada kelas 'Yes' menunjukkan bahwa mayoritas pasien yang diprediksi menderita DMTII benar-benar positif, sedangkan *precision* yang cukup tinggi pada kelas 'No' mencerminkan kemampuan model dalam meminimalkan kesalahan klasifikasi pada pasien yang tidak mengidap DMTII. Secara keseluruhan, hasil ini memperlihatkan bahwa model mampu memberikan prediksi yang cukup akurat untuk kedua kelompok tersebut.

3) *Recall*

accuracy: 93.50% +/- 3.37% (micro average: 93.50%)

	true Yes	true No	class precision
pred. Yes	159	8	95.21%
pred. No	5	28	84.85%
class recall	96.95%	77.78%	

**Gambar 4.20 Hasil Recall**

*Recall* yang dihasilkan mencapai 96,95% untuk kelas "Yes" (penderita DMTII) dan 77,78% untuk kelas "No" (bukan penderita DMTII), yang mengindikasikan bahwa model memiliki kinerja yang sangat baik dalam mendeteksi pasien yang benar-benar terdiagnosis Diabetes Melitus Tipe II. Tingginya nilai *recall* pada kelas "Yes" menunjukkan bahwa sebagian besar kasus positif berhasil teridentifikasi secara tepat. Sementara itu, meskipun *recall* untuk kelas "No" tidak setinggi kelas "Yes", model masih mampu mengenali

sebagian besar pasien *non*-DMTII dengan cukup baik, walau terdapat sejumlah kesalahan klasifikasi. Secara umum, hasil ini menggambarkan kemampuan model dalam mengenali data sebenarnya dari masing-masing kategori dengan akurasi yang cukup baik.

#### 4) F1-Score

F1-Score dihitung untuk mengukur keseimbangan antara *precision* dan *recall* dari masing-masing kelas, dengan menggunakan rumus:

$$F1\ Score = 2 \times \frac{Recall \times Precision}{Recall + Precision}$$

Berikut hasil perhitungan dari masing-masing kelas:

##### a) Kelas positif DMTII

$$F1\ Score = 2 \times \frac{95.21 \times 96.95}{95.21 + 96.95}$$

$$F1\ Score = 2 \times \frac{9230.61}{192.16}$$

$$F1\ Score = 2 \times 48.03606$$

$$F1\ Score = 96.07$$

##### b) Kelas negatif DMTII

$$F1\ Score = 2 \times \frac{84.85 \times 77.78}{84.85 + 77.78}$$

$$F1\ Score = 2 \times \frac{6599.63}{162.63}$$

$$F1\ Score = 2 \times 40.58066$$

$$F1\ Score = 81.16$$

Berdasarkan hasil evaluasi, nilai F1-Score yang diperoleh menunjukkan bahwa model klasifikasi memiliki performa yang sangat baik dalam membedakan antara pasien yang menderita dan tidak menderita Diabetes Melitus Tipe II (DMTII). F1-Score tertinggi diperoleh pada kelas positif DMTII sebesar 96,07%, menunjukkan bahwa model sangat efektif dalam mengidentifikasi pasien yang benar-benar positif DMTII dengan keseimbangan yang tinggi antara presisi dan sensitivitas.

Sementara itu, pada kelas negatif DMTII, *F1-Score* mencapai 81,16%, yang juga mencerminkan kinerja model yang cukup baik dalam mengenali pasien yang tidak menderita DMTII. Kedua nilai *F1-Score* tersebut menunjukkan bahwa model tidak hanya akurat, tetapi juga konsisten dalam melakukan klasifikasi, sehingga layak digunakan sebagai alat bantu dalam sistem pengambilan keputusan klinis di rumah sakit, khususnya dalam mendukung upaya deteksi dini dan pencegahan komplikasi DMTII.

## B. Pembahasan

Penelitian ini bertujuan untuk membangun model klasifikasi prediksi Diabetes Melitus Tipe II menggunakan algoritma *Decision Tree* berdasarkan data rekam medis pasien rawat inap di RSUD Muhammadiyah Bantul. Hasil analisis menunjukkan bahwa:

### 1. Membangun Model Prediksi

Model klasifikasi pada penelitian ini dibangun menggunakan algoritma *Decision Tree* yang diimplementasikan melalui aplikasi RapidMiner, dengan pembagian data 80% untuk proses pelatihan (*training*) dan 20% untuk pengujian (*testing*). Variabel yang digunakan sebagai prediktor meliputi tekanan darah, kadar gula darah sewaktu (GDS), usia, indeks massa tubuh (IMT), dan jenis kelamin. Hasil pemodelan menghasilkan pohon keputusan yang mudah dipahami karena setiap percabangan menggambarkan kondisi logis dari nilai atribut. Sebagai contoh, jika tekanan darah seseorang melebihi 137,5 mmHg, maka sistem secara otomatis mengklasifikasikannya ke kategori “Yes”, yaitu berisiko mengalami Diabetes Melitus Tipe II.

Penelitian oleh Nafi'ah & Fatah (2024) menyatakan bahwa *Decision Tree* efektif dalam memetakan keputusan berbasis atribut medis, karena mampu membentuk aturan yang jelas dan mudah diinterpretasi, terutama oleh praktisi medis. Temuan ini selaras dengan penelitian yang dilakukan oleh Safitri & Fatah (2023), dimana mereka mengimplementasikan *Decision Tree* menggunakan aplikasi *RapidMiner* dengan parameter seperti *gain ratio*,

*maximal depth*, dan teknik *pruning/pre-pruning*. Walaupun akurasi yang diperoleh sekitar 65 %, penelitian tersebut tetap menegaskan bahwa model dengan aturan yang jelas sangat membantu dalam mendukung pengambilan keputusan medis berbasis data. Penelitian lain oleh Haryadi et al. (2024) juga memperkuat bukti bahwa Decision Tree yang dibangun dengan RapidMiner dapat mencapai akurasi hingga 77,34%, presisi 75,08%, recall 97,60%, dan F1-score 0,85. Ini menunjukkan bahwa meskipun hasil bervariasi, model tetap mampu menghasilkan performa klasifikasi yang memadai untuk prediksi kasus diabetes.

Temuan-temuan tersebut menunjukkan bahwa penggunaan *Decision Tree* melalui *RapidMiner* memiliki dua keunggulan utama, performa klasifikasi yang cukup baik serta kemudahan interpretasi. Model ini mampu menghasilkan aturan eksplisit, seperti batas tekanan darah atau IMT, yang sangat membantu tenaga medis karena mudah dipahami dan dapat dijadikan dasar dalam panduan klinis. Dengan demikian, tahap pemodelan dalam penelitian ini membuktikan bahwa model yang dikembangkan tidak hanya akurat, tetapi juga transparan dan relevan untuk digunakan di RSUD Muhammadiyah Bantul.

## 2. Mengukur Akurasi Model Prediksi

Tahap evaluasi dalam penelitian ini bertujuan untuk mengukur performa model klasifikasi yang dibangun dengan algoritma *Decision Tree* dalam memprediksi kasus Diabetes Melitus Tipe II berdasarkan data rekam medis di RSUD Muhammadiyah Bantul. Evaluasi dilakukan dengan menggunakan metode *confusion matrix*, yang mencakup metrik akurasi, presisi, sensitivitas (*recall*), dan *F1-score*. Model diterapkan pada data uji sebanyak 200 data pasien, setelah sebelumnya data dibagi menjadi 80% data latih dan 20% data uji. Hasil evaluasi menunjukkan bahwa model *Decision Tree* mampu mengklasifikasikan data dengan tingkat akurasi sebesar 93.50%. Nilai ini menunjukkan bahwa sebagian besar prediksi model sesuai dengan data aktual, sehingga model dapat dianggap cukup andal dalam mengenali pola-pola yang berhubungan dengan risiko Diabetes Melitus Tipe II. Selain akurasi, metrik lainnya seperti *precision* dan *recall* juga memberikan hasil yang baik, yang

mencerminkan bahwa model tidak hanya akurat dalam membuat prediksi positif, tetapi juga mampu menangkap sebagian besar kasus yang benar-benar positif. Hasil ini menunjukkan bahwa model memiliki performa yang seimbang dan tidak terlalu condong ke satu sisi prediksi.

Temuan ini sejalan dengan penelitian oleh Hendriyansyah et al. (2024) yang menunjukkan bahwa algoritma *Decision Tree* memiliki akurasi sebesar 85,81% dalam mengklasifikasikan penyakit stroke otak. Selain itu, penelitian Jalali et al. (2024) juga mendukung temuan ini, di mana algoritma *Decision Tree* berhasil memprediksi risiko Diabetes Melitus Tipe II dengan akurasi 81,84%. Kedua studi ini memperkuat bahwa algoritma *Decision Tree* merupakan metode yang efektif untuk klasifikasi penyakit berbasis data klinis. Dengan demikian, model yang dikembangkan dalam penelitian ini dinilai layak dan relevan untuk digunakan dalam mendukung pengambilan keputusan medis, terutama dalam proses deteksi dini dan pencegahan Diabetes Melitus Tipe II di lingkungan rumah sakit. Evaluasi yang dilakukan tidak hanya menunjukkan keberhasilan teknis, tetapi juga memberikan kontribusi praktis dalam penerapan machine learning pada sistem informasi kesehatan.

### C. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan dalam interpretasi hasil dan penerapannya di masa mendatang, yaitu sebagai berikut:

1. Data yang digunakan merupakan data sekunder dari rekam medis elektronik pasien rawat inap di RSUD Muhammadiyah Bantul tahun 2024. Hal ini membatasi variasi populasi karena tidak mencakup pasien rawat jalan atau data dari fasilitas kesehatan lainnya, sehingga generalisasi hasil masih terbatas pada konteks rumah sakit tersebut.
2. Variabel yang digunakan dalam penelitian ini hanya mencakup usia, jenis kelamin, tekanan darah, indeks massa tubuh (IMT), dan gula darah sewaktu (GDS). Variabel penting lain seperti riwayat keluarga, pola makan, aktivitas fisik, dan kadar HbA1c tidak disertakan karena keterbatasan ketersediaan data pada sistem rekam medis elektronik. Hal tersebut membuat hasil penelitian ini menjadi bias.

3. Selain itu, pembagian data latih dan data uji hanya menggunakan metode train/test split tanpa validasi silang (*cross-validation*), yang berpotensi menyebabkan ketidaktepatan dalam mengukur performa model secara umum. Keterbatasan-keterbatasan ini diharapkan dapat menjadi bahan pertimbangan dalam pengembangan penelitian selanjutnya agar hasil yang diperoleh lebih akurat, komprehensif, dan aplikatif di berbagai konteks pelayanan kesehatan.

PERPUSTAKAAN  
UNIVERSITAS JENDERAL ACHMAD YANI  
YOGYAKARTA