

BAB IV HASIL DAN PEMBAHASAN

A. Hasil

1. Gambaran Umum Puskesmas Godean 1

a. Letak geografis

Puskesmas Godean 1 merupakan salah satu fasilitas pelayanan kesehatan tingkat pertama yang berada di wilayah Kabupaten Sleman, tepatnya beralamat di Dusun Pandean VII, Kalurahan Sidoluhur, Kapanewon Godean. Puskesmas ini mulai beroperasi sejak tahun 1983 dan telah beberapa kali mengalami perpindahan lokasi guna menyesuaikan kebutuhan serta peningkatan pelayanan. Pada tahun 1995 Puskesmas Godean 1 dibangun di lokasi yang saat ini digunakan dan pada tahun 2007 dilakukan rehabilitasi total bangunan agar sesuai dengan standar puskesmas yang telah ditetapkan oleh Pemerintah Kabupaten Sleman. Puskesmas Godean 1 menempati area dengan luas halaman sebesar 524.815 meter persegi dan luas bangunan 736 meter persegi. Wilayah kerja puskesmas ini cukup luas yaitu mencapai 13,19 kilometer persegi, mencakup beberapa dusun dan wilayah padat penduduk yang membutuhkan akses layanan kesehatan yang merata dan berkelanjutan.

Puskesmas Godean 1 menyediakan berbagai layanan kesehatan dasar dan penunjang yang cukup lengkap. Di antaranya adalah Poliklinik BPU (Balai Pengobatan Umum) untuk pengobatan penyakit umum, Poliklinik KIA (Kesehatan Ibu dan Anak) yang memberikan layanan kehamilan, persalinan, imunisasi, dan tumbuh kembang anak serta Poliklinik BPG (Balai Pengobatan Gigi) untuk perawatan dan pemeriksaan kesehatan gigi dan mulut. Selain itu tersedia pula Unit Laboratorium untuk pemeriksaan penunjang, layanan Psikologi untuk mendukung kesehatan mental, layanan Gizi untuk pemantauan dan edukasi status gizi masyarakat, layanan Sanitasi untuk pengawasan kesehatan lingkungan serta layanan Fisioterapi bagi pasien dengan

kebutuhan terapi fisik. Dengan fasilitas dan layanan yang dimiliki, Puskesmas Godean 1 terus berupaya meningkatkan mutu pelayanan serta menjangkau seluruh lapisan masyarakat di wilayah kerjanya. Puskesmas ini juga menjadi mitra penting dalam program kesehatan masyarakat yang dicanangkan pemerintah, baik dalam kegiatan promotif, preventif, kuratif, maupun rehabilitatif.

b. Pengolahan dan Pemanfaatan Data di Puskesmas Godean 1

Berdasarkan hasil pengamatan, Puskesmas Godean 1 sudah menggunakan sistem informasi kesehatan seperti SIMPUS dan SATUSEHAT untuk mencatat dan mengelola data rekam medis pasien. Data dari berbagai layanan seperti poli umum, KIA, laboratorium dan gizi telah dicatat secara digital. Hal ini menunjukkan bahwa proses pencatatan data sudah dilakukan secara terstruktur dan berbasis sistem elektronik.

Meskipun begitu, dari hasil analisis diketahui bahwa pemanfaatan data rekam medis masih terbatas, yaitu lebih banyak digunakan untuk pelaporan program dan keperluan administratif. Belum banyak dilakukan pengolahan data lebih lanjut yang bisa mendukung analisis klinis, misalnya untuk membantu deteksi dini penyakit atau pengambilan keputusan medis. Oleh karena itu penelitian ini dilakukan untuk melihat bagaimana data yang sudah tersedia dapat dimanfaatkan lebih lanjut, salah satunya dengan membangun model klasifikasi untuk membantu mendeteksi risiko hipertensi.

c. Visi dan Misi Puskesmas Godean 1

1) Visi

Terwujudnya pelayannya puskesmas yang berkualitas, terjangkau dan berdaya saing menuju masyarakat sleman yang berbudaya hidup sehat.

2) Misi

a) Menerapkan tata kelola puskesmas yang baik dengan dukungan teknologi untuk meningkatkan kualitas pelayanan pada masyarakat

- b) Meningkatkan pelayanan kesehatan yang bermutu, profesional dan terjangkau bagi semua masyarakat melalui penerapan sistem manajemen mutu secara konsisten dan berkesinambungan
- c) Meningkatkan kemandirian masyarakat untuk berbudaya hidup sehat melalui kerjasama lintas program dan lintas sektor.

2. Pengolahan data menggunakan pendekatan CRISP-DM

Penelitian ini menggunakan algoritma *Decision Tree* untuk melakukan klasifikasi status hipertensi pasien berdasarkan data rekam medis dari Puskesmas Godean 1. Data yang digunakan merupakan data sekunder berupa data pasien rawat jalan hipertensi selama periode Januari-Desember 2024. Data dibagi menjadi dua bagian yaitu 60% data latih (1.252 data) dan 40% data uji (835 data). Pengolahan dan analisis data dilakukan menggunakan aplikasi *RapidMiner* versi 10,3. Penelitian ini menggunakan pendekatan CRISP-DM (*Cross Industry Standard Process for Data Mining*) yang terdiri dari lima tahapan utama dalam penelitian ini seperti berikut :

a. *Business Understanding*

Data yang diperoleh dari Puskesmas Godean 1 menunjukkan adanya peningkatan jumlah kasus hipertensi dari tahun 2023 ke tahun 2024. Pada tahun 2023, tercatat sebanyak 11.774 pasien yang didiagnosis hipertensi sedangkan pada tahun 2024 meningkat menjadi 19.730 pasien. Data ini menunjukkan bahwa kasus hipertensi cenderung mengalami peningkatan dari tahun ke tahun. Data rekam medis yang digunakan dalam penelitian ini meliputi beberapa variabel yaitu tekanan darah sistolik, tekanan darah diastolik, usia, jenis kelamin dan IMT.

Selama ini data rekam medis di Puskesmas Godean 1 hanya dimanfaatkan untuk keperluan pelaporan program rutin dan belum digunakan secara maksimal dalam proses analisis untuk mendukung pelayanan klinis. Melalui penelitian ini data rekam medis dimanfaatkan untuk membangun model klasifikasi hipertensi menggunakan algoritma *Decision Tree*. Pengolahan data dilakukan dengan tahapan pembersihan data, pemilihan atribut yang relevan serta pengelompokan kategori

hipertensi berdasarkan pedoman *JNC 7*. Proses pemodelan dilakukan menggunakan aplikasi *RapidMiner 10,3* dengan tujuan agar hasil klasifikasi dapat membantu tenaga kesehatan dalam mengidentifikasi tingkat risiko hipertensi pada pasien secara lebih cepat dan tepat.

b. *Data Understanding*

Tahap ini bertujuan untuk memahami struktur dan karakteristik data yang akan digunakan dalam proses klasifikasi. Data yang digunakan dalam penelitian ini berasal dari file *Excel* yang *diekspor* melalui sistem informasi manajemen puskesmas yaitu SIMPUS. Data tersebut berisi informasi rekam medis pasien rawat jalan dengan diagnosa hipertensi yang berobat di Puskesmas Godean 1 selama periode Januari-Desember 2024 dengan total 8.058 data pasien. Atribut-atribut yang tersedia dalam data tersebut meliputi:

- 1) Tanggal
- 2) No Rm
- 3) NIK
- 4) Nama (disamarkan)
- 5) Nama Orangtua
- 6) Tanggal Lahir
- 7) Usia
- 8) Alamat
- 9) Desa
- 10) No Bpjs
- 11) Jenis Kunjungan
- 12) No Telephone
- 13) Jenis Pembayaran
- 14) Tinggi Badan
- 15) Berat Badan
- 16) Tekanan Darah
- 17) Tekanan Nadi/ Pernapasan
- 18) Diagnosa

- 19) Kasus
- 20) Klinik
- 21) Tindakan
- 22) Obat
- 23) Dokter
- 24) Perawat

c. *Data Preparation*

Tahapan ini mencakup proses pembersihan dan pengolahan data sebelum dianalisis:

1) Menggabungkan Data *Excel* Perbulan Menjadi 1 Tahun

1	Tanggal	No Rm	NIK	Nama	Nama Ora	Jenis Kela	Tgl Lahir	Usia	Alamat	Desa	No Bpjs	Jenis Kunj
2	2024-01-31	081	340	Si	pi	Perempu	1957	67 tahun		MARGOLU		Baru
3	2024-01-31	040	340	IV	Fi	Perempu	1958	66 tahun		SIDOAGU	000	Lama
4	2024-01-31	070	340	N	Pi	Perempu	1942	82 tahun		SIDOMOY	000	Baru Kaler
5	2024-01-31	021	340	IV	M	Perempu	1963	61 tahun		SIDOLUHU	000	Baru Kaler
6	2024-01-31	030	340	Yi	H	Perempu	1957	67 tahun		SIDOMULY	000	Baru Kaler
7	2024-01-31	030	340	Yi	H	Perempu	1960	65 tahun		SIDOMULY	000	Lama
8	2024-01-31	040	317	Fi	A	Perempu	1964	61 tahun		SIDOAGU	000	Baru Kaler
9	2024-01-31	020	340	R	Bi	Perempu	1956	68 tahun		SIDOLUHU	000	Lama
10	2024-01-31	040	340	IV	Si	Perempu	1954	70 tahun		SIDOAGU	000	Lama
11	2024-01-31	021	340	Si	Si	Perempu	1972	53 tahun		MARGOA	000	Lama
12	2024-01-31	020	340	K	Ki	Perempu	1954	71 tahun		SIDOLUHU	000	Lama
13	2024-01-31	020	340	Si	Si	Perempu	1973	51 tahun		SIDOLUHU	000	Baru Kaler
14	2024-01-31	020	340	N	N	Laki - Laki	1957	67 tahun		SIDOLUHU	000	Lama
15	2024-01-31	020	340	Si	Si	Perempu	1956	68 tahun		SIDOLUHU	000	Baru Kaler
16	2024-01-31	000	340	Ti	Ri	Perempu	1960	65 tahun		SUMBER A	000	Lama
17	2024-01-31	080	340	A	A	Laki - Laki	1961	63 tahun		MARGOLU	000	Baru Kaler
18	2024-01-31	070	340	Li	Si	Laki - Laki	1982	42 tahun		SIDOMOY	000	Baru Kaler
19	2024-01-31	020	340	V	Si	Perempu	1971	53 tahun		SIDOLUHU	000	Lama
20	2024-01-31	000	340	R	Pi	Perempu	1968	57 tahun		SENDANG	000	Baru Kaler
21	2024-01-31	070	340	Ei	Si	Perempu	1970	54 tahun		SIDOMOY	000	Baru Kaler
22	2024-01-31	040	340	V	Si	Perempu	1942	83 tahun		SIDOAGU	000	Baru Kaler

Gambar 4.1 Gabungan data *Excel* 1 Tahun

Berdasarkan data *Excel* yang diperoleh dari aplikasi SIMPUS, dilakukan proses penggabungan data kunjungan pasien hipertensi dari bulan Januari-Desember tahun 2024 menjadi satu file. Hasil penggabungan tersebut menghasilkan total sebanyak 8.058 kasus hipertensi yang tercatat selama periode tersebut.

- 2) Menghapus Data Duplikat, Data yang Tidak Lengkap (*Missing Values*) dan menghapus variabel yang tidak diperlukan
 - a) Menghapus Data Duplikat dan Data yang Tidak Lengkap (*Missing Values*)

	A	B	C	D	E	F
1	2024	08:53:52	030	340	Firmina Sumarn	Perempuar
2	2024	08:44:17	030	340	Firmina Sumarn	Perempuar
3	2024	10:05:53	040	340	RUSTAM HADJI	Laki - Laki
4	2024	07:20:33	040	340	RUSTAM HADJI	Laki - Laki
5	2024	07:31:18	040	340	RUSTAM HADJI	Laki - Laki
6	2024	07:52:45	040	340	RUSTAM HADJI	Laki - Laki
7	2024	09:19:13	040	340	RUSTAM HADJI	Laki - Laki
8	2024	09:40:09	040	340	RUSTAM HADJI	Laki - Laki
9	2024	07:37:38	040	340	RUSTAM HADJI	Laki - Laki
10	2024	07:06:48	040	340	RUSTAM HADJI	Laki - Laki
11	2024	08:02:56	040	340	RUSTAM HADJI	Laki - Laki
12	2024	07:34:28	040	340	RUSTAM HADJI	Laki - Laki
13	2024	08:55:52	040	340	RUSTAM HADJI	Laki - Laki
14	2024	09:27:08	040	340	RUSTAM HADJI	Laki - Laki
15	2024	07:27:39	040	340	RUSTAM HADJI	Laki - Laki
16	2024	07:59:21	040	340	RUSTAM HADJI	Laki - Laki
17	2024	07:22:45	040	340	RUSTAM HADJI	Laki - Laki
18	2024	07:18:40	040	340	RUSTAM HADJI	Laki - Laki
19	2024	07:11:57	040	340	RUSTAM HADJI	Laki - Laki
20	2024	08:36:41	021	350	Abdul Adim	Laki - Laki
21	2024	08:44:14	060	217	Abdul Rahman I	Laki - Laki

Gambar 4.2 Data Duplikat

Berdasarkan data diatas yaitu mengurutkan dan mengidentifikasi potensi data duplikat. Data yang menunjukkan kesamaan nama ditinjau ulang dengan mencocokkan atribut tambahan guna memastikan apakah terjadi duplikasi pencatatan.

C	D	E	F	G	H	I	J	K	L
340	Firmina Sumarni, Ny	Firmina Sumarni, Ny	Perempuar	1955-05-17	70 tahun	G		000	Baru Kalen
340	RUSTAM HADJI	R	Laki - Laki	1951	tahun	Ji		000	Lama
350	Abdul Adim	A	Laki - Laki	1960	tahun	D		000	Baru
217	Abdul Rahman Nurul	A	Laki - Laki	1979	tahun	K		000	Lama
340	Abdul Rohim Sidiq	A	Laki - Laki	1985	tahun	Ji		000	Baru Kalen
340	Abidayati, Ny	Kus						000	Lama
340	Abu Shaleh	Abu						000	Lama
340	Abu Tohir	Abu						000	Lama
340	Achmad Ali	Ach						000	Lama
340	Achmadi, Tn	Ach						000	Lama
340	Adi Suwito/slamet	Adi	Laki - Laki	1951	tahun	G		000	Baru Kalen
340	Adyatma Prawira Suc	A	Laki - Laki	1956	tahun	S		000	Baru Kalen
340	Afida Prabandari	R	Perempuar	1973	tahun	K		000	Lama
340	Agung Dwi Laksono	A	Laki - Laki	1993	tahun	B		000	Lama
340	Agus Haryadi	M	Laki - Laki	1977	tahun	G		000	Lama
340	Agus Paryono Tn	A	Laki - Laki	1988	tahun	T		000	Lama
340	Agus Supriatno	A	Laki - Laki	1977	tahun	K		000	Lama
340	Agus Supriyanto	A	Laki - Laki	1970	tahun	Ji		000	Baru Kalen
340	Ahmad Fakhrurozi Su	A	Laki - Laki	1966	tahun	P		000	Lama
340	AHMAD IRFAI	A	Laki - Laki	1986	tahun	P		000	Lama
340	Ahmad Sofyan/ Sagi	A	Laki - Laki	1967	tahun	K		000	Lama

Gambar 4.3 Penghapusan Data Duplikat

Berdasarkan data yang telah diperoleh hanya satu data yang diambil untuk setiap pasien, yaitu data dari kunjungan terakhir selama tahun 2024. Pemilihan kunjungan terakhir bertujuan untuk

merepresentasikan kondisi terkini pasien, sehingga model klasifikasi yang dibangun dapat mencerminkan status hipertensi secara lebih akurat dan relevan terhadap penanganan terakhir yang diterima pasien yaitu sebanyak 2.087 kasus hipertensi.

b) Menghapus Variabel yang Tidak Diperlukan

	F	G	H	I	J	K	L	M	N
1	Jenis Kelamin	Tgl Lahir	Usia	Alamat	Desa	No Bpjs	Jenis Kunju	Telp	Jenis Pe
2	Perempuan	1957-	67 tahun	M			Baru	62896877	UMUM
3	Perempuan	1958-	66 tahun	G		0000	Lama	620	BPJS PB
4	Perempuan	1942-	82 tahun	Ki		0000	Baru Kaler	620	BPJS PB
5	Perempuan	1963-	61 tahun	Ki		0000	Baru Kaler	620	BPJS PB
6	Perempuan	1957-	67 tahun	G		0000	Baru Kaler	620	BPJS/JK
7	Perempuan	1960-	65 tahun	G		0000	Lama	620	BPJS PB
8	Perempuan	1964-	61 tahun	G		0000	Baru Kaler	620	BPJS PB
9	Perempuan	1956-	68 tahun	Bi		0000	Lama	62859301	UMUM
10	Perempuan	1954-	70 tahun	Se		0000	Lama	620	BPJS PB

Gambar 4.4 Variabel yang Dihilangkan

Pada tahap prapemrosesan data, dilakukan penghapusan terhadap variabel-variabel yang tidak relevan dengan tujuan analisis klasifikasi hipertensi. Variabel yang dihapus antara lain: tanggal kunjungan, nomor rekam medis, NIK, nama pasien (disamarkan), nama orang tua, tanggal lahir, alamat, Desa, nomor BPJS, jenis kunjungan, nomor telepon, jenis pembayaran, tekanan nadi, diagnosa, kasus, nama klinik, Tindakan, Obat, nama dokter, dan nama perawat. Variabel-variabel tersebut dihilangkan karena tidak berkontribusi langsung terhadap proses klasifikasi dan hanya bersifat administratif, identitas atau informasi tambahan yang tidak mempengaruhi status hipertensi pasien.

	A	B	C	D	E	F	G
1	Nama_Inisial	Jenis_Kelamin	Umur	Tinggi_Badan	Berat_Badan	Tekanan_Darah_Sistolik	Tekanan_Darah_Diastolik
2	A1	Perempuan	70 tahun	156	56	160	77
3	A2	Laki - Laki	74 tahun	165	58.8	165	94
4	A3	Laki - Laki	64 tahun	157	41	120	70
5	A4	Laki - Laki	46 tahun	165	75	120	70
6	A5	Laki - Laki	39 tahun	175	85	147	103
7	A6	Perempuan	56 tahun	153	67.8	166	90
8	A7	Laki - Laki	33 tahun	179	87	136	88
9	A8	Laki - Laki	71 tahun	155	56	120	70
10	A9	Laki - Laki	62 tahun	160	60	140	92
11	A10	Laki - Laki	52 tahun	156	72.3	129	83
12	A11	Laki - Laki	73 tahun	154	55.9	152	82
13	A12	Laki - Laki	68 tahun	168	73.3	143	89
14	A13	Perempuan	52 tahun	156	68.2	142	96
15	A14	Laki - Laki	31 tahun	166	66.4	140	87
16	A15	Laki - Laki	47 tahun	170	79.8	132	82
17	A16	Laki - Laki	36 tahun	160	72.1	133	84
18	A17	Laki - Laki	47 tahun	160	79	40	110
19	A18	Laki - Laki	54 tahun	161	66.2	144	89
20	A19	Laki - Laki	59 tahun	165	71.7	124	73
21	A20	Laki - Laki	38 tahun	152	70.6	132	79
22	A21	Laki - Laki	58 tahun	160	91.4	185	96

Gambar 4.5 Variabel yang digunakan

Dalam pemodelan klasifikasi hanya atribut yang dianggap berpengaruh langsung terhadap status hipertensi pasien yang dipertahankan seperti usia, jenis kelamin, tekanan darah sistolik dan diastolik yang digunakan dalam pemodelan klasifikasi.

c) Menghitung Nilai IMT dari Berat dan Tinggi Badan

	A	B	C	D	E	F	G	H	I
1	Nama_Inisial	Jenis_Kelamin	Umur	Tinggi_Badan	Berat_Badan	Tekanan_Darah_Sistolik	Tekanan_Darah_Diastolik	Indeks_Massa_Tubuh	Keterangan_Indeks_Massa_Tubuh
2	A1	Perempuan	70 tahun	156	56	160	77	23	Overweight
3	A2	Laki - Laki	74 tahun	165	58.8	165	94	21.6	Normal
4	A3	Laki - Laki	64 tahun	157	41	120	70	16.6	Underweight
5	A4	Laki - Laki	46 tahun	165	75	120	70	27.5	Obesitas I
6	A5	Laki - Laki	39 tahun	175	85	147	103	27.9	Obesitas I
7	A6	Perempuan	56 tahun	153	67.8	166	90	27.2	Obesitas I
8	A7	Laki - Laki	33 tahun	179	87	136	88	23.3	Obesitas I
9	A8	Laki - Laki	71 tahun	155	56	120	70	23.4	Overweight
10	A9	Laki - Laki	62 tahun	160	60	140	92	29.7	Obesitas I
11	A10	Laki - Laki	52 tahun	156	72.3	129	83	23.6	Obesitas I
12	A11	Laki - Laki	73 tahun	154	55.9	152	82	26	Obesitas I
13	A12	Laki - Laki	68 tahun	168	73.3	143	89	28	Obesitas I
14	A13	Perempuan	52 tahun	156	68.2	142	96	24.1	Obesitas I
15	A14	Laki - Laki	31 tahun	166	66.4	140	87	27.6	Obesitas I
16	A15	Laki - Laki	47 tahun	170	79.8	132	82	28.2	Obesitas I
17	A16	Laki - Laki	36 tahun	160	72.1	133	84	30.9	Obesitas II
18	A17	Laki - Laki	47 tahun	160	79	40	110	25.5	Obesitas I
19	A18	Laki - Laki	54 tahun	161	66.2	144	89	26.3	Obesitas I
20	A19	Laki - Laki	59 tahun	165	71.7	124	73	30.6	Obesitas II
21	A20	Laki - Laki	38 tahun	152	70.6	132	79	35.7	Obesitas II
22	A21	Laki - Laki	58 tahun	160	91.4	185	96		

Gambar 4.6 Indeks Massa Tubuh

Berdasarkan data di atas, Indeks Massa Tubuh (IMT) dihitung untuk setiap pasien kemudian dikategorikan ke dalam lima kelompok yaitu: *underweight* (IMT < 18,5), normal (IMT 18,5–22,9), *overweight* (IMT 23–24,9), obesitas tingkat I (IMT 25–29,9), dan obesitas tingkat II (IMT > 30).

d) Mengubah Data Kategorikal ke Bentuk *Numerik*

	Inisial	Jenis_Kelamin	Usia	Tinggi_Badan	Berat_Badan	Tekanan_Darah_Sistolik	Tekanan_Darah_Diastolik	Indeks_Massa_Tubuh	KETERANGAN Indeks_Massa_Tubuh
1	A1	2	70	156	56,0	160	77	23,0	3
2	A2	1	74	165	58,8	165	94	21,6	2
3	A3	1	64	157	41,0	120	70	16,6	1
4	A4	1	46	165	75,0	120	70	27,5	4
5	A5	1	39	175	85,0	147	103	27,8	4
6	A6	2	56	153	67,8	166	90	29,0	4
7	A7	1	33	179	87,0	136	88	27,2	4
8	A8	1	71	155	56,0	120	70	23,3	3
9	A9	1	62	160	60,0	140	92	23,4	3
10	A10	1	52	156	72,3	129	83	29,7	4
11	A11	1	73	154	55,9	152	82	23,6	3
12	A12	1	68	168	73,3	143	89	26,0	4
13	A13	2	52	156	68,2	142	96	28,0	4
14	A14	1	31	166	66,4	140	87	24,1	3
15	A15	1	47	170	79,8	132	82	27,6	4
16	A16	1	36	160	72,1	133	84	28,2	4
17	A17	1	47	160	79,0	40	110	30,9	5
18	A18	1	54	161	66,2	144	89	25,5	4
19	A19	1	59	165	71,7	124	73	26,3	4
20	A20	1	38	152	70,6	132	79	30,6	5

Gambar 4.7 Data Numerik

Berdasarkan data di atas, dilakukan transformasi variabel kategorikal menjadi bentuk numerik agar dapat diproses oleh algoritma klasifikasi. Adapun pengkodean yang digunakan adalah sebagai berikut: (1) Jenis kelamin dikode dengan nilai 1 untuk laki-laki dan 2 untuk perempuan; (2) IMT diklasifikasikan menjadi lima kategori yaitu *Underweight* = 1, *Normal* = 2, *Overweight* = 3, *Obesitas I* = 4, dan *Obesitas II* = 5 serta (3) Status hipertensi dikode menjadi *Normal* = 1, *Prehipertensi* = 2, *Hipertensi tingkat 1* = 3, dan *Hipertensi tingkat 2* = 4.

e) Melakukan Pembagian Data Menjadi Data Latih dan Data Uji

	A	B	C	D	E	F	G
1043	A1044	Laki - Laki	40 tahun	165	76.6	140	102
1044	A1045	Perempuan	59 tahun	140	55.9	157	85
1045	A1046	Perempuan	75 tahun	152	44.2	174	87
1046	A1047	Perempuan	88 tahun	145	52	120	80
1047	A1048	Laki - Laki	73 tahun	165	68	139	73
1048	A1049	Laki - Laki	71 tahun	165	78	150	79
1049	A1050	Perempuan	53 tahun	156	53	134	71
1050	A1051	Perempuan	56 tahun	147	43.1	150	85
1051	A1052	Perempuan	73 tahun	148	57	149	77
1052	A1053	Perempuan	49 tahun	148	46.6	162	103
1053	A1054	Perempuan	70 tahun	153	60.5	180	84
1054	A1055	Laki - Laki	76 tahun	162	45.1	143	73
1055	A1056	Perempuan	53 tahun	161	67	140	79
1056	A1057	Perempuan	68 tahun	146	42.2	167	93
1057	A1058	Perempuan	71 tahun	146	65	163	81
1058	A1059	Perempuan	59 tahun	156	67.6	133	88
1059	A1060	Perempuan	46 tahun	155	78.9	179	100
1060	A1061	Perempuan	42 tahun	153	62.3	140	90
1061	A1062	Perempuan	94 tahun	149	36.5	187	77
1062	A1063	Perempuan	100 tahun	158	50	235	108
1063	A1064	Perempuan	84 tahun	145	46	162	82
1064	A1065	Perempuan	36 tahun	163	74.5	123	94

Gambar 4.8 Data Latih dan Data Uji

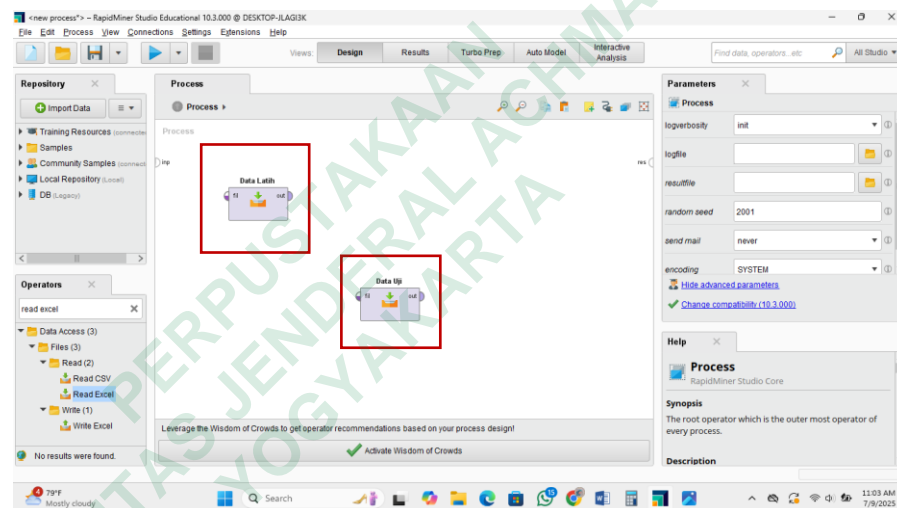
Data yang telah diproses selanjutnya dibagi menjadi dua bagian yaitu data latih (*training data*) dan data uji (*testing data*)

dengan proporsi 60:40. Sebanyak 1.252 data (60%) digunakan sebagai data latih untuk membangun model klasifikasi sedangkan 835 data sisanya (40%) digunakan sebagai data uji untuk mengevaluasi kinerja model.

d. Modeling

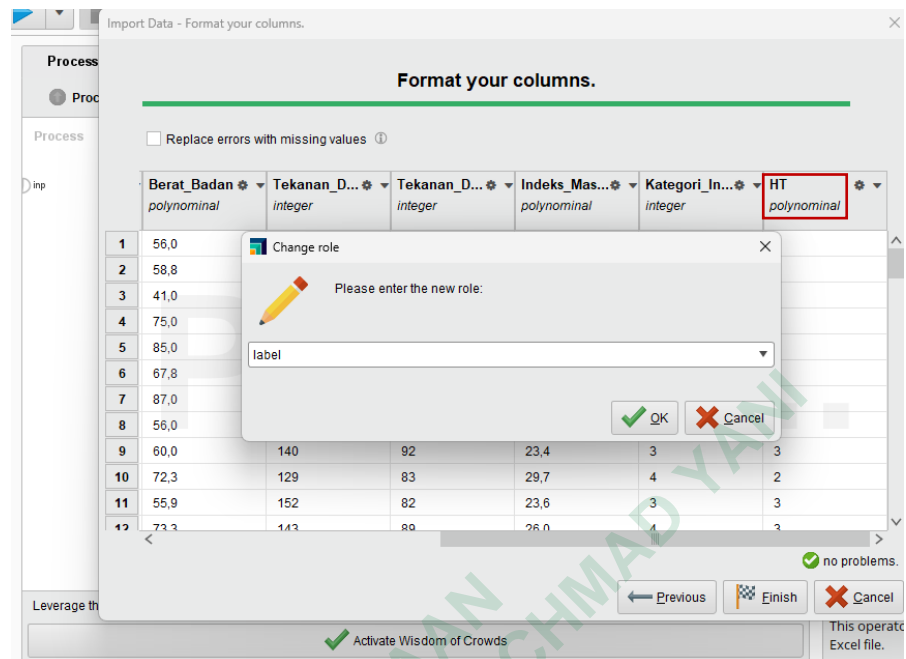
Proses analisis dalam penelitian ini melibatkan penerapan teknik data mining yang terdiri dari *clustering* sebagai metode pengelompokan data awal serta algoritma *Decision Tree* yang digunakan untuk mengklasifikasikan data berdasarkan pola yang terbentuk sebagai berikut:

1) Memasukkan Data Latih dan Data Uji pada *Rapid Miner*



Gambar 4.9 *Read Excel* Data Latih dan Data Uji

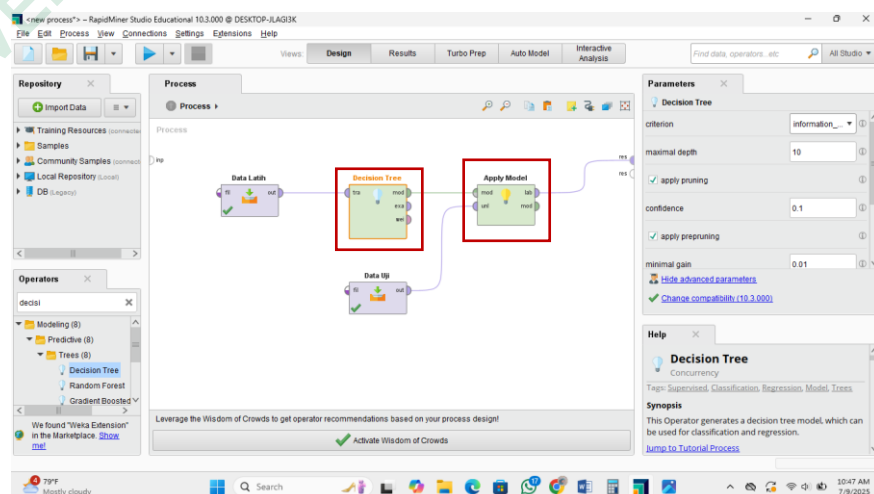
Pada tahap ini dilakukan proses pemasukan data latih dan data uji ke dalam workspace *RapidMiner* 10,3 dengan menggunakan operator *Read Excel* 2019 yang tersedia pada menu *Operators*. *Operators* ini digunakan untuk membaca file *Excel* yang berisi dataset yang telah dipersiapkan sebelumnya, baik untuk pelatihan model (*training*) maupun pengujian model (*testing*).



Gambar 4.10 Format Kolom HT

Selanjutnya format kolom pada variabel HT diubah dari tipe *integer* menjadi *polynomial* agar dapat dikenali sebagai variabel kelas (label) oleh algoritma klasifikasi. Setelah itu dilakukan pengaturan peran (*change role*) pada variabel HT sebagai label melalui menu '*Change Role*' sehingga dapat digunakan sebagai target klasifikasi dalam proses pelatihan model.

2) Memasukan *Decision Tree* dan *Apply Model*

Gambar 4.11 *Decision Tree* dan *Apply Model*

Langkah selanjutnya dalam proses klasifikasi adalah menyusun operator ‘*Decision Tree*’ dan ‘*Apply Model*’ pada *workspace*. Sebelum dijalankan, dilakukan penyesuaian terhadap parameter algoritma *Decision Tree* yaitu dengan mengganti kriteria pemilihan atribut (*criterion*) dari *Gain Ratio* menjadi *Information Gain*. Penggantian ini bertujuan agar model fokus pada atribut yang memberikan penurunan entropi terbesar dalam pembentukan pohon keputusan. Setelah konfigurasi selesai, proses dijalankan (*running*) untuk menghasilkan prediksi status hipertensi pada data uji.

3) Hasil Klasifikasi

Row No.	HT	prediction(HT)	confidence(4)	confidence(3)	confidence(1)	confidence(2)	Inisial	Jenis_Kela...	Usia
1	?	1	0	0	0.885	0.015	A1	1	80
2	?	3	0.007	0.993	0	0	A2	1	62
3	?	3	0.007	0.993	0	0	A3	1	58
4	?	3	0	0.975	0	0.125	A4	2	44
5	?	4	0.990	0.010	0	0	A5	1	80
6	?	3	0	0.941	0	0.059	A6	2	45
7	?	4	0.995	0	0.004	0	A7	1	58
8	?	3	0.007	0.993	0	0	A8	1	70
9	?	1	0	0	0.885	0.015	A9	1	70
10	?	3	0.007	0.993	0	0	A10	2	83
11	?	3	0.007	0.993	0	0	A11	1	80
12	?	2	0	0	0	1	A12	2	63
13	?	4	0.990	0.010	0	0	A13	1	72
14	?	4	0.996	0	0.004	0	A14	1	66

Gambar 4. 12 Hasil Klasifikasi

Setelah mengganti kriteria pemisahan pada algoritma *Decision Tree* menjadi *Information Gain*, model diterapkan pada data uji sebanyak 835 data pasien. Hasil prediksi ditampilkan dalam bentuk label *prediction* (HT) dan tingkat kepercayaan (*confidence*) terhadap masing-masing kelas. Mayoritas prediksi menunjukkan tingkat kepercayaan yang tinggi, seperti pada prediksi kelas hipertensi tingkat 2 dengan *confidence* mencapai 0,990. Namun, terdapat pula beberapa kasus di mana prediksi yang dihasilkan tidak selaras dengan *confidence* tertinggi.

e. Evaluation

Tujuan dari evaluasi model adalah untuk mengukur sejauh mana tingkat keakuratan model klasifikasi yang telah dibuat dalam mengelompokkan pasien hipertensi berdasarkan data rekam medis di Puskesmas Godean 1. Proses ini menggunakan metode *confusion matrix* sebagai alat penilaian dengan menghitung nilai *accuracy*, *precision*, *recall* dan *F-1 Score* pada masing-masing kategori pasien.

Model klasifikasi yang digunakan adalah algoritma *Decision Tree* yang dievaluasi menggunakan data uji sebanyak 835 pasien. Evaluasi ini dilakukan melalui aplikasi *RapidMiner versi 10,3* dengan hasil yang diperoleh sebagai berikut:

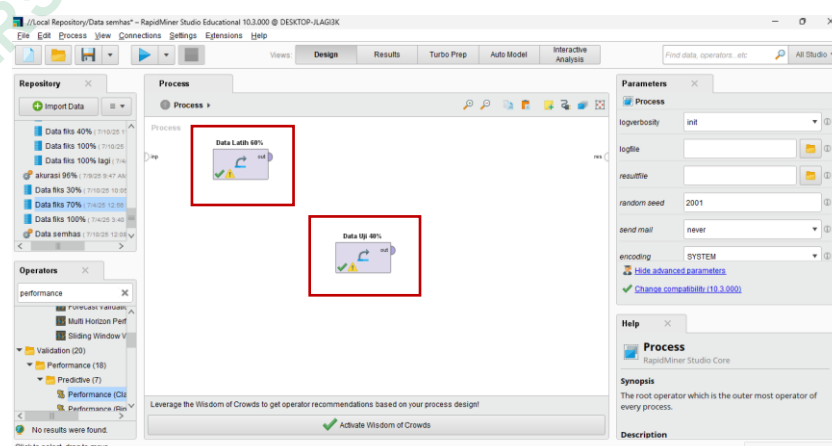
1) Accuracy

Evaluasi *accuracy* dilakukan untuk mengukur seberapa tepat model dalam mengklasifikasikan data hipertensi. Akurasi dihitung sebagai persentase jumlah prediksi yang benar dibandingkan dengan total data uji.

Langkah-langkah yang dilakukan dalam proses evaluasi akurasi antara lain:

a) Import Data

Membagi dataset menjadi data latih dan data uji



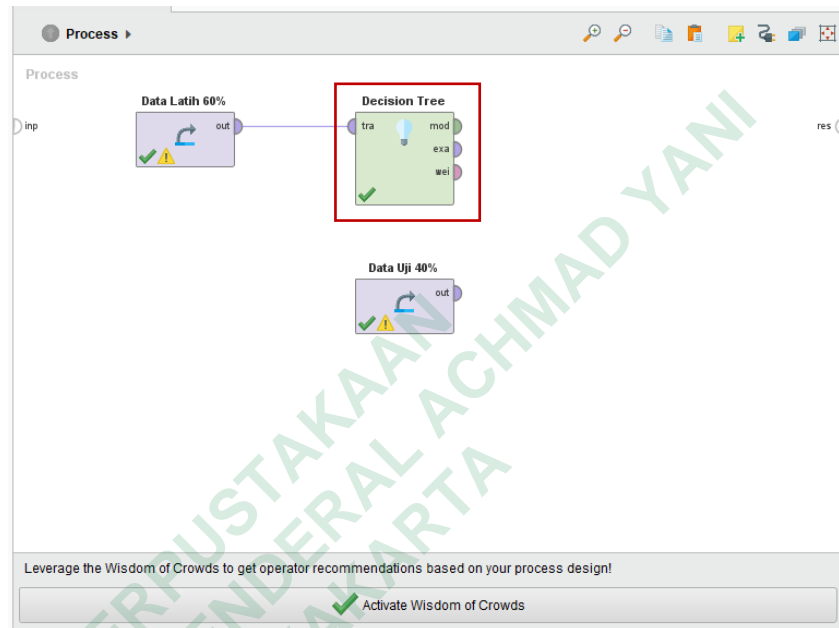
Gambar 4. 13 Dataset Data Latih dan Data Uji

Penambahan data latih sebesar 60% dan data uji sebesar 40% dilakukan untuk membangun model klasifikasi dan menguji tingkat

akurasi agar mampu mengenali pola serta memprediksi status hipertensi secara optimal pada data baru.

b) *Decision Tree*

Membangun model klasifikasi menggunakan algoritma *decision tree* berdasarkan data latih.

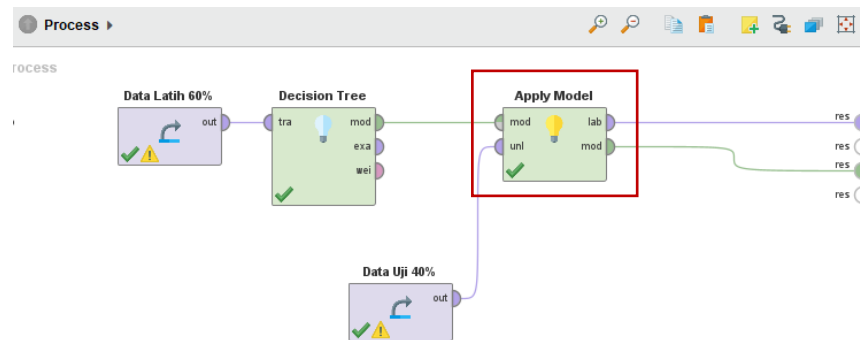


Gambar 4.14 *Decision Tree*

Penambahan algoritma *Decision Tree* dilakukan untuk membentuk model klasifikasi berdasarkan pola data latih, sehingga dapat menghasilkan aturan keputusan yang digunakan untuk memprediksi status hipertensi pasien.

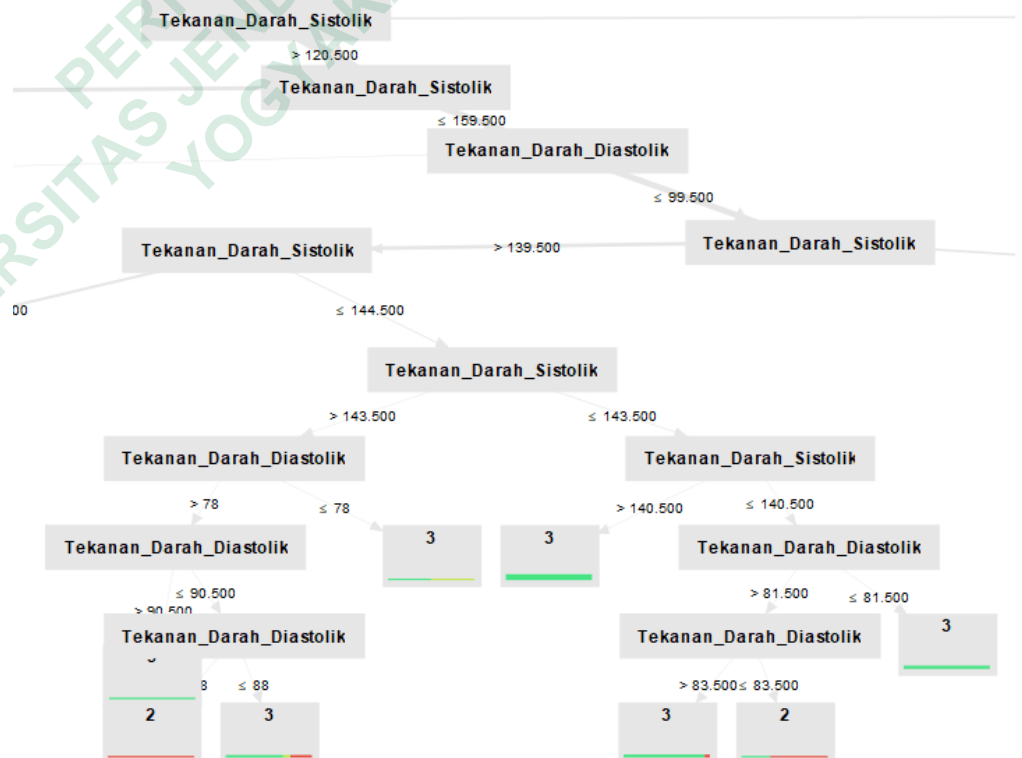
c) *Apply Model*

Menerapkan model terhadap data uji menggunakan operator *Apply Model*.



Gambar 4.15 *Apply Model*

Langkah selanjutnya yaitu menambahkan algoritma *Apply Model* kemudian *running* yang berfungsi untuk menerapkan model klasifikasi *Decision Tree* yang telah dibangun terhadap data uji, sehingga dapat menghasilkan pohon keputusan prediksi status hipertensi pada pasien berdasarkan pola yang telah dibentuk dari data latih kemudian di *running*.



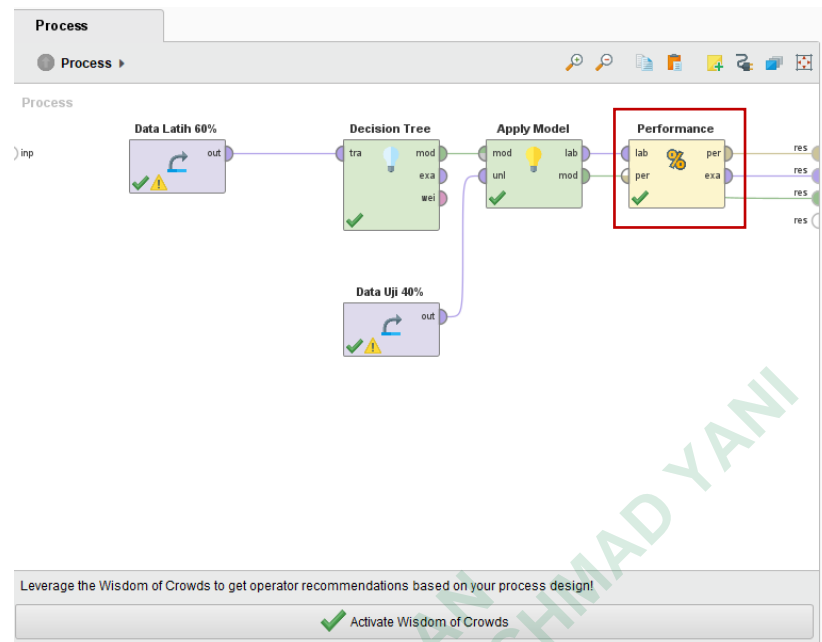
Gambar 4. 16 Pohon Keputusan

Berdasarkan hasil pemodelan klasifikasi menggunakan algoritma *Decision Tree* diperoleh pohon keputusan dengan variabel utama yaitu Tekanan Darah Systolik dan Tekanan Darah Diastolik. Pohon ini digunakan untuk menentukan klasifikasi status hipertensi pasien berdasarkan nilai tekanan darah yang tercatat dalam data rekam medis. Pada bagian akar pohon, algoritma pertama-tama memisahkan data berdasarkan nilai Tekanan Darah Systolik. Jika nilai Tekanan Darah Systolik pasien lebih dari 120,5 mmHg maka data akan diproses ke percabangan selanjutnya. Artinya pasien dengan tekanan darah systolik di bawah atau sama dengan 120,5 mmHg cenderung diklasifikasikan sebagai kategori Normal, meskipun hasil klasifikasinya tidak ditampilkan secara eksplisit dalam visualisasi.

Percabangan selanjutnya terus memanfaatkan nilai ambang batas (*threshold*) dari variabel Tekanan Darah Systolik dan Tekanan Darah Diastolik untuk menentukan kategori hipertensi. Sebagai contoh jika Tekanan Darah Systolik berada di rentang 139,5–144,5 mmHg dan Tekanan Darah Diastolik berada di atas 78 mmHg., maka model akan mengklasifikasikan pasien ke dalam kategori Hipertensi (kelas 3). Namun jika pada kondisi yang sama nilai diastolik $\leq 90,5$ mmHg, maka hasil klasifikasinya berubah menjadi Prehipertensi (kelas 2).

d) *Performance*

Mengukur performa model menggunakan operator *Performance* dengan parameter evaluasi berupa akurasi.

Gambar 4.17 *Performance*

Penambahan algoritma *Accuracy* bertujuan untuk mengevaluasi kinerja model klasifikasi yang telah dibangun dengan menghitung persentase prediksi yang benar dari keseluruhan data uji, sehingga dapat diketahui sejauh mana model mampu mengklasifikasikan status hipertensi secara tepat dan konsisten.

1) *Accuracy*

accuracy: 98.56%					
	true 1	true 3	true 4	true 2	class precision
pred. 1	123	0	0	0	100.00%
pred. 3	0	301	2	2	98.69%
pred. 4	0	1	274	0	99.64%
pred. 2	1	6	0	125	94.70%
class recall	99.19%	97.73%	99.28%	98.43%	

Gambar 4.18 *Accuracy*

Berdasarkan hasil evaluasi, model klasifikasi *Decision Tree* menghasilkan nilai akurasi sebesar 98,56%, yang menunjukkan bahwa dari seluruh data uji yang digunakan sebanyak 98,56% berhasil diklasifikasikan dengan benar.

accuracy: 98.56%

	true 1	true 3	true 4	true 2	class precision
pred. 1	123	0	0	0	100.00%
pred. 3	0	301	2	2	98.69%
pred. 4	0	1	274	0	99.64%
pred. 2	1	6	0	125	94.70%
class recall	99.19%	97.73%	99.28%	98.43%	

Gambar 4.19 Hasil Klasifikasi Model

Confusion matrix di atas menunjukkan distribusi hasil klasifikasi model terhadap data uji. Angka yang diberi tanda menunjukkan jumlah prediksi yang benar sesuai kelas sebenarnya, sedangkan angka di luar tanda adalah prediksi yang salah. Contohnya sebanyak 123 pasien yang sebenarnya Normal berhasil diklasifikasikan dengan benar dan hanya 1 pasien Normal salah diklasifikasikan sebagai Prehipertensi. Untuk kelas Hipertensi Tingkat 2 terdapat 274 prediksi yang benar dan hanya 1 kasus salah diklasifikasikan sebagai Hipertensi Tingkat 1. Hal ini menunjukkan bahwa model memiliki kemampuan klasifikasi yang sangat baik di semua kelas dengan kesalahan klasifikasi yang sangat minim.

2) Precision

accuracy: 98.56%

	true 1	true 3	true 4	true 2	class precision
pred. 1	123	0	0	0	100.00%
pred. 3	0	301	2	2	98.69%
pred. 4	0	1	274	0	99.64%
pred. 2	1	6	0	125	94.70%
class recall	99.19%	97.73%	99.28%	98.43%	

Gambar 4.20 Precision

Nilai *precision* yang diperoleh yaitu 100,00% untuk kelas Normal, 98,69% untuk Hipertensi Tingkat 1, 99,64% untuk Hipertensi Tingkat 2, dan 94,70% untuk Prehipertensi. Hal ini menunjukkan bahwa model memiliki tingkat

ketepatan yang sangat baik dalam mengklasifikasikan masing-masing kategori hipertensi.

3) Recall

accuracy: 98.56%

	true 1	true 3	true 4	true 2	class precision
pred. 1	123	0	0	0	100.00%
pred. 3	0	301	2	2	98.69%
pred. 4	0	1	274	0	99.64%
pred. 2	1	6	0	125	94.70%
class recall	99.19%	97.73%	99.28%	98.43%	

Gambar 4.21 Recall

Nilai *recall* yang diperoleh yaitu 99,19% untuk kelas Normal, 97,73% untuk Hipertensi Tingkat 1, 99,28% untuk Hipertensi Tingkat 2 dan 98,43% untuk Prehipertensi menunjukkan bahwa model memiliki kemampuan yang sangat tinggi dalam mengenali setiap kategori secara benar atau dengan kata lain hampir seluruh data aktual dari masing-masing kelas berhasil diklasifikasikan dengan tepat oleh model.

4) F1-Score

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

$$f1\ score = 2 \times \frac{recall \times presisi}{recall + presisi}$$

Berikut adalah perhitungan F-Score untuk masing-masing kelas:

a) Normal

$$F-Score = 2 \times \frac{99,19 \times 100}{99,19 + 100} = 2 \times \frac{9919}{199,9} = 2 \times 49,62 = 99,24\%$$

b) Prahipertensi

$$F-Score = 2 \times \frac{98,43 \times 94,70}{98,43 + 94,70} = 2 \times \frac{9321,32}{193,13} = 2 \times 48,26 = 96,52\%$$

c) Hipertensi Tingkat 1

$$F\text{-Score} = 2 \times \frac{97,73 \times 98,69}{97,73 + 98,69} = 2 \times \frac{9644,9}{196,42} = 2 \times 49,10 = 98,2\%$$

d) Hipertensi Tingkat 2

$$F\text{-Score} = 2 \times \frac{99,28 \times 99,64}{99,28 + 99,64} = 2 \times \frac{9892,2}{198,56} = 2 \times 49,82 = 99,64\%$$

Berdasarkan hasil evaluasi, nilai *F1-Score* yang diperoleh menunjukkan performa model yang sangat baik dalam mengklasifikasikan setiap kategori hipertensi. *F1-Score* tertinggi terdapat pada kelas Hipertensi Tingkat 2 sebesar 99,64%, disusul oleh kelas Normal sebesar 99,24%, Hipertensi Tingkat 1 sebesar 98,2% dan Prehipertensi sebesar 96,52%. Seluruh nilai *F1-Score* berada di atas 96%, dihasilkan bahwa model tidak hanya akurat dalam melakukan prediksi, tetapi juga konsisten dalam mengenali data yang sebenarnya, sehingga layak digunakan untuk mendukung sistem pengambilan keputusan klinis di Puskesmas.

B. Pembahasan

1. Pengumpulan dan Pengolahan Data Rekam Medis Pasien Hipertensi

Data yang digunakan berasal dari sistem SIMPUS Puskesmas Godean 1 dengan jumlah awal 8.058 data pasien rawat jalan tahun 2024. Setelah dilakukan pembersihan data seperti penghapusan duplikat, data kosong dan variabel tidak relevan diperoleh 2.087 data valid yang siap digunakan untuk pemodelan. Pengolahan data dilakukan dengan mengikuti pendekatan CRISP-DM yang setiap tahapan memiliki peran penting dalam memastikan bahwa model yang dihasilkan tidak hanya akurat tetapi juga relevan dengan kebutuhan lapangan dalam hal ini pelayanan kesehatan di Puskesmas Godean 1 yang terdiri dari lima tahap utama sebagai berikut:

a. *Business Understanding*

Penelitian ini diawali dengan tahap *business understanding* untuk memahami permasalahan serta menentukan pendekatan solusi yang tepat. Selama ini data rekam medis di Puskesmas Godean 1 hanya dimanfaatkan untuk pelaporan administratif, padahal memiliki potensi besar dalam mendukung deteksi dini hipertensi. Oleh karena itu penelitian ini bertujuan mengoptimalkan pemanfaatan data tersebut melalui pembangunan model klasifikasi menggunakan algoritma *Decision Tree* guna mendukung pengambilan keputusan klinis. Penelitian ini didukung oleh temuan Sari et al., (2023) yang menyatakan bahwa pemanfaatan sistem rekam medis elektronik di fasilitas kesehatan masih terbatas akibat kurangnya pemahaman dari petugas serta belum optimalnya analisis data untuk keperluan klinis. Hal ini sejalan dengan Wilandini., (2022) yang menekankan pentingnya tahap *business understanding* dalam pengolahan data untuk merumuskan solusi, mengidentifikasi kebutuhan data dan menyusun rencana kerja yang tepat.

Selaras dengan itu Yanti et al., (2024) menjelaskan bahwa *business understanding* berfokus pada pengenalan tujuan dari sisi operasional, identifikasi permasalahan utama serta perumusan strategi awal dalam pengolahan data, termasuk dalam konteks layanan atau sistem kesehatan. Supriono., (2024) juga menambahkan bahwa tahap ini penting untuk menganalisis tren penyakit berdasarkan data HIS yang dapat digunakan sebagai dasar dalam pengambilan kebijakan kesehatan dan keputusan manajerial di rumah sakit. Dengan demikian model klasifikasi yang dibangun dalam penelitian ini diharapkan mampu membantu Puskesmas dalam menangani peningkatan kasus hipertensi secara lebih efektif dan berbasis data.

b. *Data Understanding*

Tahap ini dilakukan untuk memahami struktur dan karakteristik data yang digunakan dalam proses klasifikasi. Data dalam penelitian ini berasal dari file *Excel* yang diekspor melalui Sistem Informasi Puskesmas

(SIMPUS) yang mencatat data rekam medis pasien rawat jalan dengan diagnosis hipertensi. Jumlah total data yang diperoleh sebanyak 8.058 kasus, dikumpulkan selama periode Januari hingga Desember 2024. Proses pemahaman data ini penting dilakukan sebelum tahap data *mining* sebagaimana disampaikan Zulkarnaen et al., (2023) bahwa penggunaan *RapidMiner* pada data kesehatan memerlukan peninjauan terhadap struktur dan kualitas data terlebih dahulu karena kualitas data sangat memengaruhi hasil klasifikasi yang diperoleh. Hal ini sejalan dengan penelitian Sholeh et al., (2023) yang menyatakan bahwa proses ini melibatkan identifikasi atribut dalam lembar data serta penentuan atribut yang relevan untuk membangun model klasifikasi.

Penelitian oleh Cholifatul., (2024) menunjukkan bahwa pemahaman awal terhadap data seperti membersihkan data kosong dan memilih variabel yang tepat bisa meningkatkan akurasi model sampai 99,9%. Hal ini membuktikan bahwa proses awal sebelum model dijalankan sangat berpengaruh terhadap hasil akhirnya. Sementara itu juga menunjukkan bahwa pemahaman awal terhadap data seperti membersihkan data kosong dan memilih variabel yang tepat dapat meningkatkan akurasi model hingga 99,9%. Sementara itu menurut Ariani., (2025) membahas tantangan dalam dalam menggabungkan data dari berbagai sistem rekam medis, di mana ketidaksesuaian format dapat mengganggu proses analisis. Oleh karena itu pemahaman menyeluruh terhadap struktur data sangat dibutuhkan. Selain itu berdasarkan penelitian Desiska et al., (2024) juga menekankan pentingnya pemilihan variabel berdasarkan pemahaman data sebelum dilakukan analisis. Dalam penelitiannya menggunakan metode *K-Means*, proses pemahaman dan persiapan data yang baik mampu meningkatkan kualitas hasil pengelompokan pasien.

c. *Data Preparation*

Sebelum membangun model klasifikasi, dilakukan tahap *data preparation* untuk memastikan kualitas data yang optimal. Tahap ini mencakup proses penggabungan data rekam medis per bulan menjadi satu

file tahunan, penghapusan data duplikat, penanganan *missing value* serta penghapusan variabel yang tidak relevan seperti nama pasien, alamat, nomor BPJS, dan informasi non-klinis lainnya. Hasil dari proses ini menghasilkan sebanyak 2.087 data pasien hipertensi yang siap digunakan dalam proses pemodelan. Proses *data preparation* ini mengacu pada berbagai penelitian sebelumnya yang menekankan bahwa kualitas data sangat menentukan akurasi model klasifikasi. Rahmadayanti et al., (2023) menyatakan bahwa *data preparation* mencakup proses pembersihan data untuk menghilangkan informasi yang tidak relevan agar tidak mengganggu proses pengolahan data. Putra et al., (2024) juga menegaskan bahwa tahap ini sangat krusial karena kualitas dan akurasi data sangat memengaruhi hasil klasifikasi. Pemilihan atribut yang tepat serta pemetaan label klasifikasi yang akurat juga terbukti meningkatkan performa model yang dihasilkan.

Pada penelitian oleh Sajjadnia et al., (2020) dalam penelitiannya menunjukkan bahwa tahapan *preprocessing* seperti penghapusan kesalahan (*error removal*) dan imputasi nilai hilang mampu meningkatkan performa model secara signifikan. Studi tersebut menunjukkan bahwa data yang telah melalui proses *data preparation* menghasilkan nilai akurasi, sensitivitas dan presisi yang lebih tinggi dibandingkan dengan data mentah. Hal serupa juga dikemukakan oleh Li et al., (2025) yang menyatakan bahwa teknik *data preparation* seperti imputasi nilai kosong dan konversi data numerik ke kategorikal berdampak positif terhadap akurasi dan interpretasi model klasifikasi hiperurisemia. Dengan ini menunjukkan bahwa kualitas *data preparation* sangat menentukan apakah model dapat memberikan hasil yang lebih akurat maupun lebih mudah diinterpretasikan secara klinis.

d. *Modeling*

Tahap *modeling* dilakukan dengan membagi data menjadi dua bagian yaitu 60% sebagai data latih (1.252 data) dan 40% sebagai data uji (835 data). Pembagian ini mengacu pada penelitian Prasatya et al., (2020)

yang menyatakan bahwa pada dataset berukuran kecil (< 5.000 data) rasio 60:40 mampu menjaga keseimbangan antara akurasi dan efisiensi pelatihan model. Data kemudian diolah menggunakan *software RapidMiner* versi 10,3 dengan membangun model klasifikasi menggunakan algoritma *Decision Tree* dan kriteria pemisahan *Information Gain*. Berdasarkan penelitian Yudiana et al., (2023) tahap *modeling* bertujuan untuk menerapkan teknik pemodelan yang tepat serta menampilkan pola-pola tersembunyi dalam data. Dengan algoritma *Decision Tree* hubungan antar atribut seperti tekanan darah dan usia dapat divisualisasikan dalam bentuk pohon keputusan yang mudah dipahami serta diinterpretasikan oleh tenaga medis.

Model ini didukung oleh berbagai studi sebelumnya, berdasarkan penelitian oleh Setyaningsih et al., (2022) menggunakan algoritma C4.5 dalam *RapidMiner* dengan pendekatan *Information Gain* untuk membangun pohon keputusan yang efektif. Pada penelitian Muraina., (2022) menyatakan bahwa rasio pembagian data 60:40 menciptakan representasi terbaik dari dataset, yang berkontribusi pada hasil evaluasi model yang lebih andal. Selaras dengan penelitian Bozorgnezhad., (2024) menggunakan *Decision Tree* pada data klinis dan demografis lebih dari 10.000 pasien dan memperoleh akurasi 79,5% serta AUC 0,80 dengan usia dan IMT sebagai variabel penting. Shariful et al., (2022) juga mencatat bahwa meskipun algoritma *ensemble* seperti *Random Forest* mencapai akurasi lebih tinggi ($\sim 90\%$), *Decision Tree* tetap dipilih karena kesederhanaannya dan kemampuannya memberikan interpretasi langsung yang sangat relevan dalam layanan kesehatan dasar.

e. *Evaluation*

Pada tahap ini dilakukan evaluasi performa model berdasarkan hasil klasifikasi pada data uji dengan menggunakan *confusion matrix* serta penghitungan metrik evaluasi seperti akurasi, presisi, *recall* dan *F1-score*. Model yang dihasilkan menunjukkan hasil yang sangat baik dengan akurasi sebesar 98,56% dan nilai *F1-score* di atas 96% pada seluruh kelas. Evaluasi

ini sangat penting untuk mengukur kemampuan model secara menyeluruh dan memastikan bahwa hasil klasifikasi sesuai dengan tujuan klinis penelitian. Metode evaluasi serupa juga digunakan dalam penelitian oleh Sathyanarayanan., (2024) yang memanfaatkan *confusion matrix* dan metrik evaluasi untuk menilai kualitas klasifikasi secara komprehensif. Penelitian Fitrianti et al., (2023) juga menegaskan bahwa tahap evaluasi sangat krusial guna memastikan kesesuaian hasil model dengan tujuan serta kebutuhan aplikasi penelitian. Selain itu Aprianur et al., (2024) menyatakan bahwa model dengan akurasi $\geq 75\%$ sudah dapat dikategorikan sebagai model yang baik dan dapat diandalkan dalam aplikasi prediktif klinis serta pengambilan keputusan.

Hasil evaluasi ini juga sejalan dengan temuan penelitian terdahulu di bidang klasifikasi hipertensi. Şen et al., (2022) mengungkapkan bahwa algoritma *Decision Tree* mampu mencapai akurasi sebesar 76,9% dan *F1-score* 76,5% yang meskipun bukan yang tertinggi tetap dianggap memadai karena struktur pohon keputusan yang dihasilkan mudah dipahami oleh tenaga medis. Penelitian Iaccarino et al., (2022) menunjukkan bahwa performa *Decision Tree* cenderung lebih rendah dibandingkan dengan model *ensemble* seperti *Random Forest* dan *AdaBoost*, khususnya pada metrik AUC. Namun kemudahan interpretasi model menjadi alasan utama algoritma ini tetap banyak digunakan dalam studi klinis. Dengan mempertimbangkan hasil-hasil tersebut, evaluasi model pada penelitian ini tidak hanya berfokus pada tingkat akurasi, tetapi juga pada kemudahan interpretasi hasil oleh tenaga kesehatan dan menjadi salah satu keunggulan utama penggunaan algoritma *Decision Tree* dalam konteks klinis.

2. Penerapan Algoritma *Decision Tree* untuk Mengelompokkan Pasien Berdasarkan Risiko Hipertensi

Model *Decision Tree* yang dibangun mampu mengelompokkan pasien ke dalam empat kategori berdasarkan tekanan darahnya: Normal, Prehipertensi, Hipertensi Tingkat 1 dan Hipertensi Tingkat 2. Variabel yang digunakan dalam pemodelan antara lain: tekanan sistolik, tekanan diastolik, usia, IMT dan jenis

kelamin. Menurut Aditya et al., (2021) *Decision Tree* cocok untuk klasifikasi klinis karena model yang dihasilkan bersifat visual dan mudah dipahami. Hal ini didukung oleh penelitian Setyaningsih et al., (2022) menambahkan bahwa algoritma ini akan memilih atribut dengan *information gain* tertinggi untuk dijadikan sebagai akar pohon keputusan. Klasifikasi yang digunakan dalam penelitian ini memanfaatkan sejumlah variabel klinis yang memiliki pengaruh besar terhadap status hipertensi pasien. Variabel-variabel tersebut diambil dari data rekam medis dan dipilih berdasarkan relevansinya secara medis. Berikut analisis peran klinis tiap variabel terhadap hasil klasifikasi:

a. Tekanan Darah *Sistolik*

Hasil penelitian di Puskesmas Godean 1 menunjukkan bahwa tekanan darah sistolik merupakan variabel paling dominan dalam klasifikasi hipertensi. Pasien dengan nilai sistolik ≥ 160 mmHg umumnya dikategorikan sebagai Hipertensi Tingkat 2, meskipun tekanan diastoliknya masih dalam batas normal. Hal ini sejalan dengan penelitian Fengler et al., (2020) yang menjelaskan bahwa secara patofisiologis, peningkatan tekanan sistolik disebabkan oleh arteriosklerosis dan penurunan elastisitas pembuluh darah yang menyebabkan kekakuan arteri meningkat. Kekakuan ini mengakibatkan gelombang denyut balik (*pulse wave*) datang lebih cepat dan memperkuat tekanan sistolik. Kondisi ini umum terjadi pada usia lanjut dan dikenal sebagai *isolated systolic hypertension* (hipertensi sistolik terisolasi). Panggabean., (2023) juga menyatakan bahwa tekanan sistolik yang melebihi ambang batas dapat mengganggu mekanisme autoregulasi tubuh dalam menjaga aliran darah tetap stabil. Bila mekanisme ini terganggu, perfusi darah ke organ vital seperti otak dapat menurun dan meningkatkan risiko komplikasi neurologis.

Hal ini selaras dengan penelitian Siransy et al. (2020) menyatakan bahwa hipertensi sistolik lebih sering terjadi pada kelompok lansia akibat perubahan fisiologis, seperti meningkatnya resistensi perifer dan menurunnya kemampuan homeostasis sistem kardiovaskular. Kaiyin et al. (2024) juga menekankan pentingnya lokasi pengukuran tekanan darah di

mana *central systolic blood pressure* (central SBP) dianggap sebagai prediktor kejadian kardiovaskular yang lebih kuat dibandingkan tekanan sistolik *perifer* (misalnya pada lengan). Hal ini menunjukkan bahwa tidak hanya nilai SBP yang penting, tetapi juga metode dan lokasi pengukurannya. Bassil (2025) mendukung hal ini dengan temuan bahwa pengukuran tekanan darah menggunakan metode *auskultatori* (manset) dapat meremehkan nilai SBP, terutama pada individu dengan kekakuan arteri yang tinggi. Akibatnya hingga sepertiga kasus hipertensi sistolik dapat tidak terdeteksi secara klinis. Temuan-temuan ini menggaris bawahi pentingnya perhatian terhadap tekanan sistolik dalam proses klasifikasi dan pengambilan keputusan klinis berbasis data.

b. Tekanan Darah *Diastolik*

Berdasarkan hasil penelitian di Puskesmas Godean 1, sebagian besar pasien hipertensi memiliki tekanan darah diastolik ≥ 90 mmHg. Hal ini menunjukkan bahwa tekanan diastolik merupakan salah satu faktor penting dalam menentukan klasifikasi hipertensi, terutama pada kategori Hipertensi Tingkat 1 dan 2 berdasarkan pedoman *JNC 7*. Hal ini sejalan dengan penelitian Husnu et al., (2020) Secara patofisiologis tekanan diastolik yang tinggi disebabkan oleh peningkatan resistensi pembuluh darah *perifer* akibat *vasokonstriksi* kronis. Kondisi ini menyebabkan tekanan tetap tinggi saat jantung berada dalam fase relaksasi sehingga meningkatkan beban kerja jantung. Peningkatan DBP dapat menyebabkan *hipertrofi* ventrikel kiri dan meningkatkan risiko gagal jantung, stroke, dan penyakit jantung koroner. Sebagaimana selajalan dengan penelitian Riyada et al., (2024) Tekanan diastolik tinggi juga berkaitan dengan aktivasi sistem *renin-angiotensin-aldosteron* (RAAS) yang memperparah retensi cairan dan *vasokonstriksi*. Selain itu Lukitaningtyas., (2023) menyatakan bahwa tekanan ini dapat mencerminkan kekakuan pembuluh darah akibat perubahan struktural yang memperburuk aliran darah ke organ vital. Dengan demikian tekanan diastolik tidak hanya menjadi variabel klinis tetapi juga mencerminkan proses patofisiologi penting

dalam hipertensi yang perlu diperhatikan dalam upaya deteksi dini dan klasifikasi risiko pasien.

c. IMT

Hasil klasifikasi di Puskesmas Godean 1 menunjukkan bahwa pasien dengan kategori *overweight* dan obesitas lebih banyak mengalami hipertensi dibandingkan pasien dengan IMT normal. Pada pohon keputusan, IMT muncul sebagai variabel pendukung dalam klasifikasi menuju prehipertensi hingga hipertensi tingkat 2. Artinya semakin tinggi IMT semakin besar kemungkinan pasien masuk ke dalam kategori hipertensi lanjut. Hal ini sejalan dengan penelitian Prastowo et al., (2020) Secara patofisiologis obesitas memicu peningkatan aktivitas sistem saraf simpatik dan RAAS (*renin-angiotensin-aldosteron*) yang menyebabkan peningkatan tekanan darah. Selain itu menurut Dharmawan et al., (2022) berat badan berlebih menyebabkan tekanan lebih besar pada dinding arteri dan membuat jantung bekerja lebih keras, sehingga meningkatkan risiko hipertensi. Oleh karena itu IMT menjadi salah satu variabel penting dalam klasifikasi hipertensi karena berhubungan langsung dengan kondisi fisik pasien dan tekanan darah.

d. Usia

Hasil penelitian di Puskesmas Godean 1 menunjukkan bahwa mayoritas pasien hipertensi berada pada kelompok usia di atas 40 tahun. Ini menunjukkan bahwa risiko hipertensi meningkat seiring bertambahnya usia. Pada hasil pohon keputusan variabel usia muncul sebagai penentu lanjutan setelah tekanan darah dan IMT terutama dalam mengklasifikasikan pasien ke kategori prehipertensi dan hipertensi tingkat 1. Artinya pada pasien dengan tekanan darah mendekati ambang batas, usia menjadi faktor penentu apakah pasien tergolong ringan atau sudah masuk tahap lanjut. Hal ini sejalan dengan penelitian (Adila., 2023) yang menyatakan bahwa peningkatan usia berkaitan dengan penurunan elastisitas pembuluh darah dan fungsi endotel yang memicu peningkatan tekanan darah. Usia juga merupakan faktor risiko *non-modifikabel* yang

perlu dipantau. Selain itu menurut Siransy et al., (2020) menambahkan bahwa pada usia lanjut terjadi stensilisasi arteri (kekakuan pembuluh), penurunan fungsi ginjal dalam membuang natrium serta gangguan keseimbangan tekanan vaskular yang mempercepat peralihan dari hipertensi ringan ke berat. Dengan demikian usia berperan penting dalam klasifikasi status hipertensi dan harus diperhitungkan sebagai faktor risiko fisiologis yang berdampak langsung terhadap progresi penyakit.

e. Jenis Kelamin

Hasil penelitian di Puskesmas Godean 1 menunjukkan bahwa variabel jenis kelamin tidak berpengaruh signifikan dalam hasil klasifikasi hipertensi berdasarkan pohon keputusan. Meskipun jumlah pasien perempuan lebih banyak, algoritma tidak menjadikan jenis kelamin sebagai faktor utama dalam klasifikasi. Namun secara klinis jenis kelamin memiliki peran dalam risiko hipertensi. Menurut penelitian Ali et al., (2020) menyebutkan bahwa laki-laki cenderung lebih cepat mengalami hipertensi karena pengaruh hormon testosteron yang bersifat vasokonstriktor sedangkan estrogen pada perempuan bersifat vasodilator dan melindungi pembuluh darah. Pada penelitian Mustika., (2023) juga menjelaskan bahwa risiko hipertensi pada perempuan meningkat setelah *menopause* akibat penurunan kadar estrogen.

3. Evaluasi Hasil Klasifikasi Model *Decision Tree*

Evaluasi dilakukan untuk mengetahui seberapa baik model *Decision Tree* dalam mengelompokkan pasien berdasarkan risiko hipertensi. Dalam penelitian ini evaluasi menggunakan *confusion matrix* yang terdiri dari empat metrik utama, yaitu akurasi, *precision*, *recall* dan *F1-score*. Keempat metrik ini digunakan agar hasil klasifikasi tidak hanya akurat tapi juga seimbang dan bisa diterapkan dalam dunia nyata, khususnya di fasilitas pelayanan kesehatan seperti puskesmas. Berdasarkan hasil evaluasi, model menunjukkan kinerja yang sangat baik. Akurasi model mencapai 98,56%, sementara nilai *precision*, *recall*, dan *F1-score* untuk setiap kategori berada di atas 96%. Hal Ini

menunjukkan bahwa model tidak hanya tepat dalam memprediksi, tapi juga konsisten dan mampu bekerja dengan baik meskipun digunakan pada data baru.

Hal ini sejalan dengan penelitian Clara et al., (2021) menyebutkan bahwa *confusion matrix* merupakan metode evaluasi paling efektif dalam klasifikasi klinis karena menunjukkan proporsi prediksi yang benar dan salah. Utami., (2021) menambahkan bahwa *F1-score* digunakan untuk menilai keseimbangan antara *precision* dan *recall* yang penting ketika klasifikasi terdiri dari beberapa kelas seperti dalam kasus hipertensi. Hal ini juga didukung oleh penelitian Octariadi., (2025) mengemukakan bahwa nilai evaluasi seperti ini penting untuk menilai apakah model bisa diandalkan dalam proses pengambilan keputusan di lapangan. Selain dilihat secara keseluruhan performa model juga dianalisis berdasarkan masing-masing kategori hipertensi. Jika dibandingkan dengan penelitian lain hasil ini menunjukkan keunggulan yaitu pada penelitian Elisa et al., (2021) mencatatkan akurasi 80% pada klasifikasi hipertensi lansia dan pada penelitian Putra et al., (2024) mencatatkan akurasi 91,67% dalam klasifikasi pasien diabetes menggunakan metode yang sama. Dalam hal ini model klasifikasi yang dibangun dalam penelitian tidak hanya akurat tetapi juga stabil dan dapat diterapkan dalam sistem informasi puskesmas untuk mendukung deteksi dini hipertensi.

C. Keterbatasan Penelitian

Penelitian ini memiliki beberapa keterbatasan yang perlu diperhatikan. Pertama penelitian hanya dilakukan di satu lokasi yaitu Puskesmas Godean 1, sehingga hasil dan model klasifikasi yang dikembangkan belum dapat digeneralisasi ke populasi yang lebih luas. Kedua variabel yang digunakan dalam model masih terbatas pada data rekam medis yang tersedia tanpa mempertimbangkan faktor eksternal lain seperti gaya hidup, status ekonomi, dan tingkat stres yang juga berpengaruh terhadap kejadian hipertensi. Selain itu kajian dalam penelitian ini masih berfokus pada aspek data *mining* dan belum memasukkan pendekatan *patofisiologis* secara menyeluruh dalam pembentukan model klasifikasi.