

**PENERAPAN ALGORITMA *DECISION TREE* DAN SMOTE UNTUK
KLASIFIKASI IMBALANCED DATASET PADA PASIEN HIPERTENSI
UNTUK PREDIKSI *PLAN PERAWATAN PASIEN*.**

TUGAS AKHIR

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Strata-1
Program Studi Teknik Industri - Fakultas Teknologi Industri
Universitas Islam Indonesia**



Nama : Rizha Syukrilah
No. Mahasiswa : 19 522 122

**PROGRAM STUDI TEKNIK INDUSTRI
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA
YOGYAKARTA
2023**

PERNYATAAN KEASLIAN

ii

PERNYATAAN KEASLIAN

Saya mengakui bahwa tugas akhir ini adalah hasil karya saya sendiri kecuali kutipan dan ringkasan yang seluruhnya sudah saya jelaskan sumbernya. Jika dikemudian hari ternyata terbukti pengakuan saya ini tidak benar dan melanggar peraturan yang sah maka saya bersedia ijazah yang telah saya terima ditarik kembali oleh Universitas Islam Indonesia.



SURAT BUKTI PENELITIAN



RUMAH SAKIT AS-SYIFA
Jl. Gerak Alam RT. 13 Kelurahan Kota Medan
Kecamatan Kota Manna Kabupaten Bengkulu Selatan
Telp. (0739) 21888 Email: rs.assyifa.manna@gmail.com



SURAT KETERANGAN

Nomor : 001/RSAS/B/SPAL/VII/2023

Kepada Yth.

Universitas Islam Indonesia
Kepala Jurusan Teknik Industri

Di Tempat

Dengan Hormat,

Dengan ini kami sampaikan bahwa mahasiswa tersebut di bawah ini :

Nama : Rizha Syukrilah
No. Mhs : 19522122
Jurusan : Teknik Industri

Yang bersangkutan telah selesai melakukan penelitian di **Rumah Sakit As-Syifa** Manna Bengkulu Selatan, Jl. Gerak Alam, Kel. Kota Medan, Kec. Kota Manna, Kab. Bengkulu Selatan selama 1 (Satu) bulan, terhitung mulai tanggal 16 Januari 2023 Sampai dengan 17 Februari 2023. Untuk memperoleh data dalam rangka penyusunan skripsi penelitian yang berjudul **“Penerapan Algoritma Decision Tree Dan Smote Untuk Klasifikasi Imbalanced Dataset Pada Pasien Hipertensi Untuk Prediksi Plan Perawatan Pasien”**.

Demikian surat keterangan ini dibuat dan diberikan kepada yang bersangkutan untuk dapat digunakan sebagaimana mestinya, atas perhatian dan kerjasamanya yang baik, kami ucapkan terimakasih.

Manna, 20 Februari 2023

DIREKTUR
Rumah Sakit As-syifa

dr. Andanu Sulaksana
NIK: 1701016-1-001



LEMBAR PENGESAHAN PEMBIMBING

iv

LEMBAR PENGESAHAN PEMBIMBING

PENERAPAN ALGORITMA *DECISION TREE* DAN SMOTE UNTUK
KLASIFIKASI IMBALANCED DATASET PADA PASIEN HIPERTENSI
UNTUK PREDIKSI *PLAN* PERAWATAN PASIEN.



Disusun Oleh :

Nama : Rizha Syukrilah

No. Mahasiswa : 19522122

Yogyakarta, 07 Juli 2023

Dosen Pembimbing

(Annisa Uswatun Khasanah, ST., M.B.A., M.Sc.)

LEMBAR PENGESAHAN DOSEN PENGUJI

LEMBAR PENGESAHAN DOSEN PENGUJI

**PENERAPAN ALGORITMA DECISION TREE DAN SMOTE UNTUK
KLASIFIKASI IMBALANCED DATASET PADA PASIEN HIPERTENSI
UNTUK PREDIKSI PLAN PERAWATAN PASIEN.**

TUGAS AKHIR

Disusun Oleh :

Nama : Rizha Syukrilah

No. Mahasiswa : 19 522 122

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Strata-1 Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia

Yogyakarta, 09 - Agustus - 2023

Tim Penguji

Annisa Uswatun Khasanah, ST., M.B.A., M.Sc.

Ketua

Ir. Muhammad Ridwan Andi Purnomo, ST, M.Sc, PhD,
IPM

Anggota I

Dian Janari, ST., MT

Anggota II

Mengetahui,

Ketua Program Studi Teknik Industri
Fakultas Teknologi Industri
Universitas Islam Indonesia
* YOGYAKARTA *

Ir. Muhammad Ridwan Andi Purnomo, ST, Msc, PhD, IPM

015220101

HALAMAN PERSEMBAHAN

Tugas Akhir ini saya persembahkan untuk kedua orang tua serta kakak dan adik saya tercinta yang senantiasa mendoakan, mendukung, memberi motivasi, serta memberikan jasa yang tak terbalaskan untuk kehidupan saya serta orang-orang yang selalu ada di kehidupan saya.

MOTTO

Sesungguhnya jika kamu bersyukur, pasti Kami akan menambah (nikmat) kepadamu, namun jika kamu mengingkari (nikmat-Ku), maka sesungguhnya azab-Ku sangatlah pedih.

-Q.S. Ibrahim: 7-

KATA PENGANTAR

Alhamdulillah puji dan syukur penulis panjatkan atas kehadiran Allah SWT yang telah senantiasa memberikan rahmat dan karunia-Nya sehingga penulis dapat menyelesaikan tugas akhir yang berjudul **“Penerapan Algoritma *Decision Tree* Dan SMOTE Untuk Klasifikasi *Imbalanced Dataset* Pada Pasien Hipertensi Untuk Prediksi Plan Perawatan Pasien.”**. Penyusunan tugas akhir ini dilakukan dalam rangka memenuhi salah satu syarat kelulusan untuk mencapai gelar Sarjana Teknik Program Studi Teknik Industri Fakultas Tekonologi Industri Universitas Islam Indonesia.

Penulis menyadari bahwa dalam penulisan tugas akhir ini banyak mengalami kendala, namun berkat bantuan, bimbingan, serta kerja sama dari berbagai pihak dan ridho dari Allah SWT sehingga kendala-kendala yang penulis hadapi dapat terselesaikan. Oleh karena itu, penulis ingin mengucapkan terimakasih dan penghargaan kepada :

1. Bapak Prof., Dr., Ir., Hari Purnomo, M.T., IPU, ASEAN.Eng selaku Dekan Fakultas Teknologi Industri Universitas Islam Indonesia.
2. Bapak Ir., Muhammad Ridwan Andi Purnomo, S.T., M.Sc., Ph.D., IPM selaku Ketua Prodi Teknik Industri Fakultas Teknologi Industri Universitas Islam Indonesia.
3. Ibu Annisa Uswatun Khasanah, ST., M.B.A., M.Sc. selaku Dosen Pembimbing Tugas Akhir yang telah meluangkan waktu, memberikan bimbingan, ilmu, petunjuk, nasehat, motivasi, dan dorongan kepada penulis selama penyusunan Tugas Akhir ini.
4. Dr. Andanu Sulaksana selaku direktur Rumah Sakit As-Syifa Manna yang telah mengizinkan penulis untuk melakukan penelitian serta memperoleh informasi yang dibutuhkan dalam penelitian ini.
5. Kedua orang tua tercinta, ayahanda Budi Prayitno, S.P dan Ibunda saya Eka Yulianti, S.Pd yang telah menjadi sosok paling berpengaruh dalam hidup penulis yang selalu memberikan doa, nasehat, motivasi, dukungan, kasih sayang, perhatian dan cinta yang tak terhingga kepada penulis dalam menempuh pendidikan.
6. Saudara kandung penulis dr. Ridho Prizka Putro dan Khesyfa Hanan Kinanti. yang telah banyak memberikan bantuan serta dukungan untuk penulis dalam menyelesaikan Tugas Akhir ini.
7. Muhammad Oriwardana Karil, S.T yang telah menjadi sahabat seperjuangan penulis dari dimulainya masa perkuliahan hingga akhir penulisan Tugas Akhir penulis.
8. Rekan-rekan seperjuangan Teknik Industri 2019, atas segala bantuan dan kerjasamanya. Tekhusus untuk Oka, Dio, Imam, Qowi, Majid, Kevin, Daffa, Fairuz, Tiara, Gadis yang selalu memberikan dukungan dan banyak membantu penulis.
9. Semua pihak yang tidak dapat penulis tuliskan satu persatu yang telah membantu dalam penyelesaian penulisan Tugas Akhir.

Penulis memahami sepenuhnya bahwa Tugas Akhir ini tidak luput dari kesalahan. Oleh karena itu, kritik dan saran yang membangun sangat diharapkan demi perbaikan dimasa yang akan datang. Semoga Tugas Akhir ini dapat menjadi inspirasi bagi para pembaca untuk melakukan hal yang lebih baik lagi dan Semoga Tugas Akhir ini bermanfaat dalam dunia pendidikan di masa mendatang.

ABSTRAK

Kualitas pelayanan dalam industri kesehatan merupakan faktor yang sangat krusial dan penting untuk dipertimbangkan. Rumah sakit As-Syifa adalah salah satu rumah sakit swasta bertipe D yang berlokasi di Kota Manna, Kabupaten Bengkulu Selatan, Provinsi Bengkulu. Rumah sakit ini merupakan salah satu rumah sakit yang sering menjadi rujukan untuk daerah atau kabupaten di sekitar kabupaten Manna. Dalam proses pelayanan penerimaan pasien rawat inap, system kerja rumah sakit memberikan waktu tunggu yang lama bagi pasien. Hal ini tentunya memengaruhi performa pelayanan sekaligus *output* pasien rawat inap itu sendiri. Penelitian ini dilakukan untuk melakukan klasifikasi menggunakan metode *decision tree C4.5* yang digunakan untuk klasifikasi dan prediksi *plan perawatan* pasien *hypertensive heart disease*. Dari hasil pengujian menggunakan seluruh data rekam medis mulai tanggal 1 Januari hingga 31 Desember 2022, diperoleh sebanyak 20 aturan dengan akurasi sebesar 99% untuk model tanpa metode *oversampling*. Sedangkan 33 aturan untuk model integrasi dengan metode *oversampling* dengan akurasi 92.08%. Penggunaan metode SMOTE untuk menangani *imbalance data* dalam kasus ini membuat performa model lebih baik secara menyeluruh yang dibuktikan dengan meningkatnya nilai *recall true positive* dan *precision* rawat inap.

Kata Kunci: *Decision Tree C4.5*, *Imbalance data*, Rekam Medis, Rawat Inap, Rawat Jalan, SMOTE.

DAFTAR ISI

PERNYATAAN KEASLIAN.....	ii
SURAT BUKTI PENELITIAN	iii
LEMBAR PENGESAHAN PEMBIMBING	iv
LEMBAR PENGESAHAN DOSEN PENGUJI.....	v
HALAMAN PERSEMBAHAN	vi
MOTTO	vii
KATA PENGANTAR.....	viii
ABSTRAK	ix
DAFTAR ISI	x
DAFTAR TABEL	xiii
DAFTAR GAMBAR.....	xv
BAB I	
PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	8
1.3 Tujuan Penelitian	8
1.4 Manfaat Penelitian	9
1.5 Batasan Penelitian	9
1.6 Sistematika Penulisan.....	10
BAB II TINJAUAN PUSTAKA	12
2.1 Kajian Literatur	12
2.2 Research Gap	15
2.3 Landasan Teori.....	16
2.3.1 Knowledge discovery in database	16
2.3.2 Data mining.....	17
2.3.3 Decision tree.....	19
2.3.4 Perbandingan decision tree C4.5 dan ID3.....	21
2.3.5 Imbalance dataset	26
2.3.6 Metode Resampling	26

2.3.7	Synthetic Minority Over-sampling Technique (SMOTE)	29
2.3.8	Rekam Medis	30
2.3.9	Manfaat Rekam Medis	31
2.3.10	Hypertensive Heart Disease	31
2.3.11	Indeks Masa Tubuh	33
2.3.12	Denyut Nadi	33
2.3.13	Klasifikasi Suhu Tubuh Manusia	33
2.3.14	Saturasi Oksigen (SPO ₂).....	34
2.3.15	Frekuensi Pernapasan (Respirasi)	34
2.3.16	Glasgow Coma Scale	35
2.3.17	Usia	35
BAB III	METODE PENELITIAN.....	36
3.1	Subjek Penelitian.....	36
3.2	Objek Penelitian	36
3.3	Variabel Penelitian	36
3.4	Alur Penelitian	38
BAB IV	PENGUMPULAN DAN PENGOLAHAN DATA.....	44
4.1	Pengumpulan Data	44
4.2	<i>Pre-Processing</i>	44
4.2.1	Merge data	44
4.2.2	Data setelah penanganan missing value	46
4.2.3	Data Diskritisasi	47
4.2.4	Transformasi data	49
4.2.5	Data reduksi	50
4.2.6	Data akhir	52
4.3	Analisis Deskriptif	53
4.4	Pengolahan Data.....	56
4.4.1	Decision tree C4.5 sebelum SMOTE.....	56
4.4.2	Decision Tree C4.5 Setelah SMOTE	60
BAB V	PEMBAHASAN.....	80
5.1	Analisis <i>Rules</i> Analisis Klasifikasi Algoritma <i>Decision Tree</i> C4.5.....	80
5.2	Evaluasi Performa Klasifikasi.....	81
5.3	Perbandingan Performa Model SMOTE dan Non-SMOTE	82
5.4	Analisis Hasil Penggunaan Tipe Data Berbeda	84
5.5	Analisis Perbaikan Masalah	85
BAB VI	PENUTUP.....	87

6.1	Kesimpulan	87
6.2	Saran.....	88
DAFTAR PUSTAKA.....		89
LAMPIRAN		98

DAFTAR TABEL

Tabel 2. 1 Kajian Literatur.....	12
Tabel 2. 2 <i>Research Gap</i>	15
Tabel 2. 3 <i>Rain Level</i>	23
Tabel 2. 4 Klasifikasi indeks masa tubuh	33
Tabel 2. 5 Klasifikasi denyut nadi	33
Tabel 2. 6 Klasifikasi tubuh Manusia	34
Tabel 2. 7 Tingkat Saturasi Oksigen.....	34
Tabel 2. 8 Kategori nilai GCS	35
Tabel 2. 9 Kategori usia.....	35
Tabel 2. 10 Kategori nilai GCS	48
Tabel 2. 11 Diskritisasi data usia	48
Tabel 2. 12 Indeks masa tubuh	48
Tabel 4. 1 Variabel penelitian.....	36
Tabel 4. 2 Data ID pasien	44
Tabel 4. 3 Data hasil pemeriksaan	45
Tabel 4. 4 Hasil <i>merge</i> data	45
Tabel 4. 5 Data setelah proses penanganan <i>missing value</i>	46
Tabel 4. 6 Transformasi data	49
Tabel 4. 7 Hasil statistik uji variabel independen.....	51
Tabel 4. 8 Hasil keputusan uji independensi terhadap variabel independen	51
Tabel 4. 9 Kesimpulan uji independensi.....	51
Tabel 4. 10 Data akhir.....	52
Tabel 4. 11 Tabel kontingensi jenis kelamin dan jenis rawat	53
Tabel 4. 12 Tabel kontingensi tingkat hipertensi dan jenis rawat	54
Tabel 4. 13 Tabel kontingensi variabel level nadi dengan jenis rawat	54
Tabel 4. 14 Tabel kontingensi level SpO2 dan Jenis rawat	55
Tabel 4. 15 Tabel kontingensi usia dan jenis rawat	55

Tabel 4. 16 tabel kontingensi kesadaran dan jenis rawat.....	56
Tabel 4. 17 CM Sebelum SMOTE.....	58
Tabel 4. 18 <i>Confusion matrix</i> setelah SMOTE.....	62
Tabel 4. 19 Jumlah data keseluruhan.....	64
Tabel 4. 20 <i>record instance</i> level nadi.....	64
Tabel 4. 21 <i>record instance</i> level SpO2	66
Tabel 4. 22 <i>record instance</i> Usia.....	67
Tabel 4. 23 <i>Record instance</i> tingkat hipertensi.....	69
Tabel 4. 24 <i>Record instance</i> kesadaran	71
Tabel 4. 25 <i>Record instance</i> tingkat hipertensi.....	72
Tabel 4. 26 Record instance Jenis kelamin.....	74
Tabel 4. 27 Rekapitulasi perhitungan manual	75
Tabel 5. 1 Performa klasifikasi	81
Tabel 5. 2 Perbandingan performa SMOTE dan Non-SMOTE.....	82

DAFTAR GAMBAR

Gambar 1. 1 Jumlah kunjungan pasien hipertensi	3
Gambar 1. 2 Jumlah kunjungan pasien 2022	5
Gambar 2. 1 Struktur diagram <i>decision tree</i>	20
Gambar 2. 2 <i>Simple split</i>	22
Gambar 2. 3 <i>Cross validation</i>	23
Gambar 2. 4 ilustrasi cara kerja SMOTE.....	30
Gambar 3. 1 Alur Penelitian	39
Gambar 4. 1 Distribusi jenis rawat pasien	53
Gambar 4. 2 <i>Decision tree</i> sebelum SMOTE	57
Gambar 4. 3 <i>Decision tree</i> setelah SMOTE	60

BAB I

PENDAHULUAN

1.1 Latar Belakang

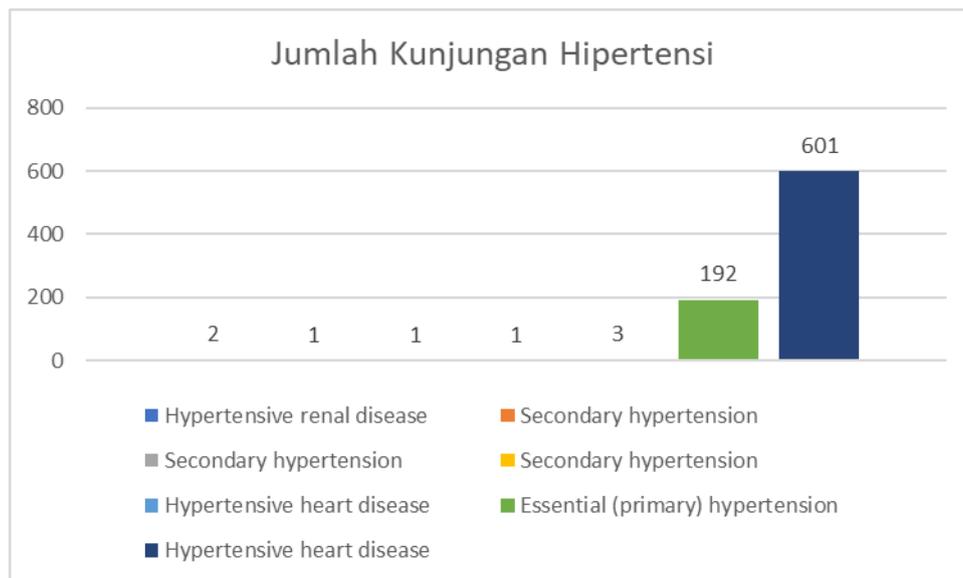
Perkembangan di industri kesehatan dunia saat ini tengah mengalami skala perubahan yang sangat besar. Transformasi ini dilakukan untuk melakukan penghematan dari sisi biaya operasional namun tetap berupaya untuk meningkatkan kualitas pelayanan dan efisiensi layanan kesehatan (Wu et al., 2016). Salah satu tantangan terbesar bagi industri kesehatan di Indonesia adalah penyakit tidak menular. Sekitar 70% kematian di seluruh dunia diakibatkan oleh penyakit tidak menular. Penyakit tidak menular (PTM) di prediksi akan terus mengalami kenaikan (WHO, 2018). Hal ini disebabkan oleh kebiasaan buruk pada pola makan masyarakat yang sering memakan makanan tinggi kalori, rendah serat, tinggi natrium, mengandung gula yang tinggi dan mengandung lemak yang banyak ditambah dengan gaya hidup *sedentary* yang lebih memilih makanan siap saji yang diikuti dengan pola hidup yang tidak baik diketahui dengan pasti akan berdampak buruk bagi fungsi fisiologis dan kesehatan tubuh (Kaur et al., 2022). Kombinasi kebiasaan buruk ini memicu timbulnya penyakit seperti hipertensi, diabetes melitus, obesitas, kanker, jantung dan hiperkolestrol (Reskiaddin et al., 2020). Data tersebut menunjukkan bahwa kebutuhan masyarakat untuk mendapatkan perawatan yang tepat dan optimal akan meningkat seiring berjalannya waktu.

Sesuai dengan peraturan perundang-undangan Indonesia No. 44 tahun 2009, rumah sakit merupakan suatu institusi pelayanan berbasis kesehatan yang mana rumah sakit merupakan tempat penyelenggaraan pelayanan kesehatan yang menyediakan berbagai macam pelayanan seperti pelayanan rawat inap, rawat jalan dan gawat darurat. Rumah sakit As-Syifa adalah salah satu rumah sakit swasta bertipe D yang berlokasi di Kota Manna, Kabupaten Bengkulu Selatan, Provinsi Bengkulu. Rumah sakit ini merupakan salah satu rumah sakit yang sering menjadi rujukan untuk daerah atau kabupaten di sekitar kabupaten Manna. Berdasarkan Peraturan Menteri Kesehatan Republik Indonesia No. 56 alinea 4 mengenai klasifikasi rumah sakit umum kelas D, rumah sakit tipe D merupakan rumah sakit yang wajib memiliki dan menjamin pelayanan kelas 3 yaitu pelayanan secara medis, farmasi, kebidanan, penunjang klinik maupun non-klinik dan rawat inap. Sebagai penyelenggara layanan kesehatan rumah sakit tipe D, rumah sakit As-Syifa sering mendapat pasien rujukan dari puskesmas di daerah.

Berdasarkan hasil wawancara dengan pihak internal rumah sakit, diketahui bahwa penyakit dengan pasien terbanyak adalah pasien penderita penyakit dalam yaitu hipertensi. Di sepanjang tahun 2022 tercatat total pasien hipertensi yang mengunjungi rumah sakit As-Syifa menyentuh angka 800 pasien. Hipertensi merupakan penyakit yang dinilai bertanggung jawab atas 8,5 juta kematian di seluruh dunia akibat *stroke*, ginjal dan penyakit pembuluh darah lainnya (Zhou, Perel, et al., 2021).

Dikutip dari laman resmi organisasi kesehatan dunia (WHO), hipertensi merupakan gangguan kesehatan yang ditandai dengan meningkatnya tekanan darah yang disebabkan oleh pembuluh darah yang mengalami peningkatan tekanan secara terus menerus. Seseorang dikategorikan hipertensi ketika memiliki tekanan darah sistolik sebesar 140 mmHg atau lebih dan tekanan darah *diastolic* kurang dari 90 mmHg (Zhou, Carrillo-Larco, et al., 2021). Tercatat di tahun 2021, sebanyak 1,4 juta orang di seluruh dunia mengidap penyakit hipertensi yang mana rata-rata terdapat pada negara berpenghasilan rendah hingga menengah sekaligus menegaskan bahwa hipertensi merupakan salah satu gangguan kesehatan dengan kasus tertinggi di dunia saat ini (Insani et al., 2022). Penelitian yang sama juga mengatakan hal yang sama, angka prevalensi kasus hipertensi di dunia di tahun 2019 menyentuh angka 22% dari total keseluruhan populasi manusia di dunia yang mana 2/3 pengidap hipertensi berada di negara yang memiliki ekonomi menengah ke bawah (Kemenkes, 2019). Berdasarkan data hasil riset kesehatan dasar yang dilakukan oleh Kementerian Kesehatan Republik Indonesia, diketahui prevalensi hipertensi di Provinsi Bengkulu menyentuh angka 8,4% yang didasarkan kepada diagnostik dokter untuk usia penduduk ≥ 18 tahun.

Pada penelitian yang akan dilakukan, peneliti berfokus kepada salah satu penyakit yang disebabkan oleh komplikasi penyakit hipertensi, yaitu *Hypertensive Heart Disease* (HHD). Jumlah penderita HHD di rumah sakit As-Syifa sendiri merupakan penyakit dalam dengan jumlah kunjungan tertinggi selama periode tahun 2022. Berikut ini merupakan jumlah kunjungan pasien penderita hipertensi yang tercatat di *database* rumah sakit.

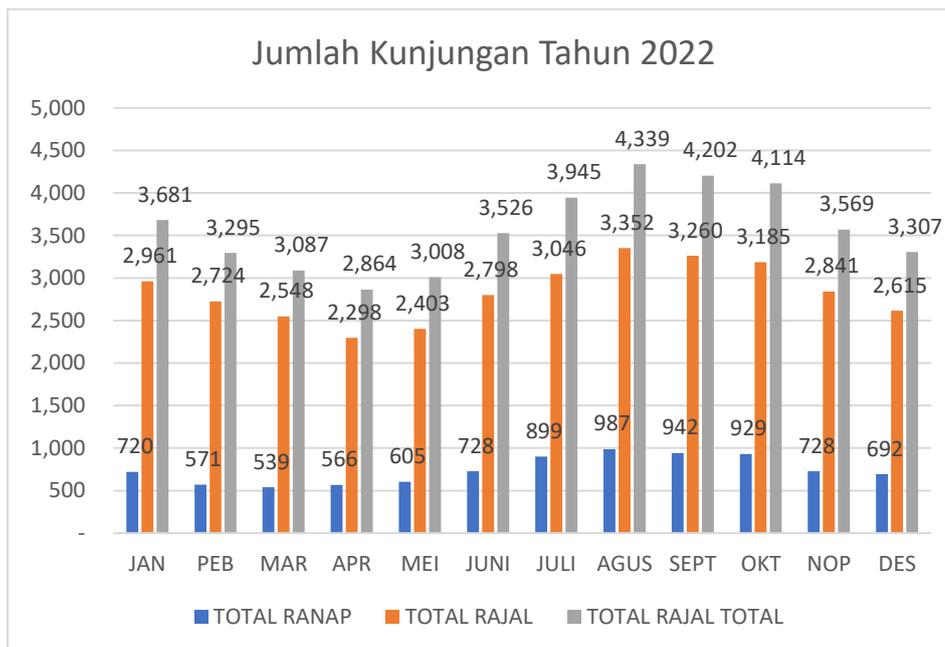


Gambar 1. 1 Jumlah kunjungan pasien hipertensi

Selain itu, *Hypertensive Heart Disease* (HHD) merupakan penyakit hipertensi yang memiliki angka kematian tertinggi yaitu sebanyak 5 orang selama periode tahun 2022. *Hypertensive Heart Disease* (HHD) terjadi ketika terdapat sejumlah perubahan pada ventrikel bagian kiri atau *left ventricular hypertrophy* (LVH), atrium kiri, dan pembuluh darah coroner karena mengalami peningkatan tekanan darah secara terus menerus dalam jangka waktu yang lama. Hal ini menyebabkan kerja jantung yang semakin berat dengan akibat, otot jantung semakin lama semakin melemah. Selain itu, tekanan darah tinggi juga membuat pembuluh darah akan selalu menegang yang membuat otot jantung mengalami pembesaran, kaku dan tidak lagi elastis. *Hypertensive Heart Disease* (HHD) merupakan salah satu penyakit yang disebabkan oleh hipertensi yang menjadi faktor utama terjadinya gagal jantung sehingga harus dilakukan tindakan yang tepat dan terukur.

Manajemen informasi di rumah sakit As-Syifa saat ini telah berbasis sistem informasi manajemen (SIM). Dalam proses pelayanannya baik administrasi maupun *record data screening* pasien, rumah sakit telah menggunakan Sistem Informasi Manajemen Rumah Sakit (SIMRS). Namun pihak rumah sakit mencatat seluruh data pasien hanya untuk arsip rumah sakit. Berdasarkan Peraturan Menteri Kesehatan No. 269 Tahun 2008 mengenai rekam medis pasien, rumah sakit wajib melakukan pemusnahan terhadap data rekam medis selambat-lambatnya 5 tahun dari terakhir pasien berobat. Sedangkan, rumah sakit hingga saat ini tidak ada proses pengolahan data lanjutan dan kondisi data yang ada saat ini mengalami peningkatan

dari sisi kecepatan data. Sehingga apabila tidak dimanfaatkan tentunya arsip data rekam medis hanya menjadi tumpukan arsip yang tidak memiliki nilai tambah. Hal ini sangat disayangkan apabila rumah sakit tidak mengolah data rekam medis yang memiliki banyak *insight* berarti bagi rumah sakit. Permasalahan lain yang dihadapi rumah sakit adalah saat ini belum ada model atau bahan evaluasi dan alat bantu bagi tenaga medis dalam menentukan seorang pasien dikategorikan sebagai pasien rawat inap atau rawat jalan. Ketika pasien HHD datang ke rumah sakit, pasien akan diterima oleh perawat dan dilakukan pemeriksaan secara cepat untuk mengetahui kondisi pasien secara umum. Setiap pasien yang mendaftar di rumah sakit akan diidentifikasi sebagai pasien rawat jalan terlebih dahulu. Kemudian data yang telah didapat oleh perawat akan diberikan kepada dokter spesialis jantung dan paru untuk mendapatkan pengesahan *plan* perawatan yang akan dijalani oleh pasien apakah rawat jalan atau rawat inap. Namun system seperti ini dinilai kurang efisien. Hal ini dikarenakan, ketika pasien telah mendapatkan vonis rawat inap yang diberikan oleh dokter, pasien harus kembali ke bagian administrasi guna mendaftar kembali sebagai pasien rawat inap. Hal ini terjadi karena, setiap pasien yang baru masuk akan diidentifikasi sebagai pasien rawat jalan. Sehingga, perawat bagian administrasi masih belum mengetahui bahwa kondisi pasien terkini telah berubah karena mendapatkan vonis rawat inap dari dokter. Oleh karena itu, dibutuhkan pendaftaran ulang untuk memperbaharui seluruh informasi pasien. Setelah melakukan pendaftaran dan menyelesaikan administrasi, perawat akan berkoordinasi dengan divisi keuangan terkait tarif yang akan dibayarkan oleh pihak pasien. Setelah semua urusan administrasi selesai, perawat akan mempersiapkan kamar yang akan ditempati oleh pasien. Sistem seperti ini dinilai menyulitkan bagi perawat ketika jumlah pasien sedang tinggi ditambah dengan jumlah perawat yang terbatas. Berikut ini merupakan jumlah pasien rumah sakit As-Syifa di sepanjang tahun 2022.



Gambar 1. 2 Jumlah kunjungan pasien 2022

Selain itu, proses ini memberikan waktu tunggu yang cukup lama bagi pasien rawat inap untuk mendapatkan kamar yang akan ditempati. Waktu tunggu bagi pasien rawat inap sendiri bisa berdampak negative, karena penundaan dapat meningkatkan morbiditas dan mortalitas sehingga dapat berdampak pada kondisi pasien kedepannya (Dewi, 2015). Rumah sakit As-Syifa saat ini merupakan rumah sakit tipe D yang artinya memiliki jumlah dokter yang terbatas. Tercatat, rumah sakit saat ini hanya memiliki 1 dokter spesialis jantung dan paru. Sedangkan, dalam mengambil keputusan apakah seseorang dikategorikan sebagai pasien rawat inap atau rawat jalan, perawat harus mendapat persetujuan dari dokter. Pada penelitian yang dilakukan oleh *National Association of healthcare Acces Management* (NAHAM) juga menyatakan bahwa waktu tunggu untuk penerimaan pasien rawat inap merupakan salah satu masalah yang umum ditemui di berbagai rumah sakit. Untuk menangani permasalahan tersebut, rumah sakit dapat memanfaatkan data rekam medis pasien yang dimiliki untuk diolah agar dapat mendapatkan *insight* serta manfaat dan dapat dijadikan alat bantu di beberapa *station* di rumah sakit.

Cara yang dikenal *powerful* untuk melakukan ekstraksi pada data baru dan memiliki informasi yang berguna dalam suatu *dataset* yang besar adalah *data mining* (Luo et al., 2022). Salah satu metode yang terdapat dalam *data mining* adalah Analisis Klasifikasi. Klasifikasi bertujuan untuk melakukan ekstraksi *rules* yang penting pada data yang digunakan untuk memprediksi *output* baru atau prediksi yang belum pernah terlihat sebelumnya (Jin et al., 2022).

Dalam dunia *machine learning*, klasifikasi sendiri termasuk ke dalam *supervised learning* (Sanjeev et al., 2023). Selama 2 dekade atau 20 tahun terakhir, sangat banyak dilakukan pengembangan pada metode mining khususnya analisis klasifikasi seperti *rule based system*, *rough set*, *algoritma evolusi*, *artificial neural Network*, dan *decision tree* (Jin et al., 2022).

Algoritma yang sering digunakan dalam menyelesaikan kasus klasifikasi adalah algoritma *decision tree* atau pohon keputusan. Salah satu kelebihan yang dimiliki oleh algoritma *decision tree* adalah algoritma ini dapat dibentuk menjadi tampilan grafis yang mudah untuk dipahami, lebih transparan dan lebih mudah untuk diinterpretasikan atau dijelaskan kepada *stakeholder* (Delen et al., 2013). Berdasarkan penelitian terdahulu yang dilakukan oleh Nour & Polat (2020), dengan judul penelitian “*Automatic Classification of Hypertension Types Based on Personal Features by Machine Learning Algorithms*” dari 4 algoritma yang dilakukan percobaan yaitu *decision tree C4.5*, *Random Forest Classifier*, *Linear Discriminant Analysis (LDA) Classifier* dan *Linear SVM Classifier* diketahui bahwa untuk memprediksi penyakit hipertensi, yang menjadi algoritma terbaik adalah C4.5 dengan akurasi sebesar 99,5%. Penelitian lain yang dilakukan oleh Chang et al (2019) menunjukkan performa algoritma *decision tree C4.5* dengan hasil yang baik dengan akurasi sebesar 86,30%. Kelebihan utama dari algoritma *decision tree C4.5* adalah algoritma ini menghasilkan pohon keputusan yang sangat mudah untuk dipahami dan diinterpretasikan, akurasi yang dihasilkan tergolong baik, efisien dalam mengolah atribut yang memiliki tipe diskrit sekaligus, algoritma *decision tree* dapat menangani dua tipe data yang berbeda, yaitu diskrit dan numerik (Kamagi & Hansun, 2016). Pemilihan algoritma *decision tree* juga dinilai tepat dalam kasus ini karena, *decision tree* memiliki output yang unik dibandingkan dengan algoritma lainnya. Dengan menggunakan *decision tree*, informasi yang terdapat dalam big data akan diekstrak dan direpresentasikan ke dalam bentuk *if-THEN rules* dalam bentuk grafik pohon sehingga lebih mudah untuk dipahami (Fiarni et al., 2019). Selain itu, *decision tree* merupakan algoritma klasifikasi yang paling efektif dikarenakan metode ini dinilai sederhana dan hasil yang diperoleh *user friendly* (Mitrofanov & Semenkin, 2021).

Penelitian yang akan dilakukan juga menggunakan metode *oversampling* yaitu SMOTE untuk mengatasi ketimpangan pada label data yang terjadi. Penggunaan SMOTE dalam penelitian ini dikarenakan kondisi data pasien penderita *Hypertensive Heart Disease* mengalami ketimpangan antara jumlah pasien penderita *Hypertensive Heart Disease* rawat jalan dan rawat inap. Kondisi data yang mengalami ketimpangan dalam dunia machine learning

disebut sebagai *imbalanced data*. *Imbalanced data* merupakan kondisi dimana kelas data mayoritas mengalami jumlah selisih (*gap*) yang besar dibandingkan dengan kelas minoritas sehingga mengalami ketimpangan pada label kelas data (Mutmainah, 2021). Berdasarkan data yang telah diberikan oleh pihak rumah sakit, diketahui bahwa terjadi ketimpangan data rawat jalan dan rawat inap pada data rekam medis pasien penderita *Hypertensive Heart Disease*. Untuk data yang terekam di sistem *database* rumah sakit, tercatat penderita HHD adalah sejumlah 601 pasien. Berikut ini merupakan hasil rekapitulasi jumlah kunjungan pasien hipertensi di sepanjang tahun 2022. Namun untuk jumlah data secara menyeluruh dengan mempertimbangkan kunjungan kembali yang dilakukan oleh pasien, tercatat sebanyak 827 data pasien rawat jalan dan 10 data pasien rawat inap. Sehingga berdasarkan kondisi tersebut, dibutuhkan solusi untuk menangani ketimpangan data yang terjadi di lapangan. Pada penelitian yang dilakukan oleh (Ishaq et al., 2021) dengan penelitian yang berjudul “*Improving the Prediction of Heart Failure Patients’ Survival Using SMOTE and Effective Data Mining Techniques*” memberikan hasil bahwa teknik *oversampling* SMOTE memberikan peningkatan yang signifikan terhadap performa algoritma klasifikasi yang digunakan untuk melakukan prediksi pada khususnya pada algoritma klasifikasi *tree-based*. Setelah mengatasi ketimpangan data yang terjadi, perbandingan label data yaitu data rawat inap dan data rawat jalan menjadi 50:50 dengan jumlah keseluruhan data menjadi 1654 data.

Dengan kondisi yang telah dijelaskan sebelumnya, peneliti akan melakukan klasifikasi terhadap *plan perawatan* pasien yang mengidap penyakit hipertensi khususnya *Hypertensive heart disease* berdasarkan data hasil pemeriksaan awal pasien yang meliputi *plan* perawatan (rawat inap, rawat jalan), sistolik, distolik, *heart rate*/menit, respirasi/menit, tinggi badan (cm), berat badan (kg) dan *glasgow coma scale* (GCS) yang akan dilakukan pada rumah sakit As-Syifa Kota Manna, Kabupaten Bengkulu Selatan, Provinsi Bengkulu. Data tekanan dara sistolik dan diastolik akan dilakukan pengolahan agar mendapatkan tingkat hipertensi pasien. Selain itu, data tinggi dan berat badan juga diolah untuk mendapatkan indeks masa tubuh pasien. Penelitian dilakukan dengan menggunakan algoritma klasifikasi *decision tree* C4.5 yang bertujuan untuk menemukan *rules* atau kriteria pasien *hypertensive heart disease* yang dikatakan layak untuk mendapatkan *plan* perawatan rawat inap dan rawat jalan. Sehingga, dengan menggunakan hasil penelitian yang dilakukan, diharapkan perawat telah memiliki prediksi serta gambaran *plan* perawatan yang akan diberikan oleh dokter kepada pasien sejak pasien berada di *nurse station*. Dengan system seperti ini, perawat dapat segera mempersiapkan

seluruh kebutuhan yang dibutuhkan oleh pasien rawat inap tanpa perlu mengulangi proses yang ada. Dengan adanya bantuan model prediksi yang dimiliki perawat, diharapkan dapat memangkas waktu proses dari awal kedatangan pasien hingga pasien mendapatkan kamar inap. Dengan terpangkasnya waktu pelayanan pasien, antrian pasien akan lebih cepat terurai dan pasien bisa segera masuk ke kamar inap setelah mendapatkan vonis dari dokter tanpa adanya waktu tunggu. Sehingga dapat menurunkan angka morbiditas dan mortalitas. Selain itu, penelitian yang dilakukan diharapkan dapat meningkatkan kinerja dari perawat dalam memberikan pelayanan, memperbaiki system yang saat ini dijalankan oleh rumah sakit serta mengembangkan dan memvalidasi alat baru untuk menilai siapa yang layak untuk menjadi pasien rawat inap atau rawat jalan pada pasien.

1.2 Rumusan Masalah

Berikut ini merupakan masalah yang dapat disusun berdasarkan latar belakang penelitian.

1. Bagaimana kriteria pasien yang direkomendasikan untuk mendapat tindakan rawat inap dan rawat jalan?
2. Bagaimana hasil model klasifikasi untuk prediksi *plan* perawatan pasien rumah sakit As-Syifa dengan menggunakan algoritma *decision tree C45*?
3. Bagaimana performa klasifikasi sebelum dan sesudah mengatasi *imbalanced dataset* yang terjadi menggunakan SMOTE berdasarkan algoritma *decision tree C4.5*?
4. Apa usulan perbaikan yang dapat diberikan kepada pihak rumah sakit untuk menyelesaikan permasalahan yang terjadi?

1.3 Tujuan Penelitian

Berikut ini merupakan tujuan dilakukannya penelitian.

1. Mengetahui kriteria pasien yang direkomendasikan untuk mendapatkan rawat inap dan rawat jalan.
2. Mengetahui hasil model klasifikasi untuk prediksi *plan* perawatan pasien rumah sakit As-Syifa dengan menggunakan algoritma *decision tree C45*.

3. Mengetahui pengaruh model klasifikasi sebelum dan sesudah dilakukannya penanganan *imbalanced dataset* dengan menggunakan SMOTE berdasarkan algoritma *decision tree* C4.5.
4. Memberikan usulan perbaikan yang dapat diberikan kepada pihak rumah sakit untuk menyelesaikan permasalahan yang terjadi

1.4 Manfaat Penelitian

Berikut ini merupakan manfaat penelitian yang dilakukan untuk peneliti dan pihak rumah sakit.

1. Bagi Peneliti

Penelitian yang dilakukan diharapkan mampu memberikan gambaran secara nyata terkait implementasi dari rumpun ilmu jurusan Teknik Industri khususnya konsentrasi studi operasional riset.

2. Bagi Rumah sakit

Penelitian diharapkan dapat menjadi bahan evaluasi dalam menentukan kriteria pasien yang layak untuk mendapatkan tindakan rawat inap dan rawat jalan berdasarkan aturan yang terbentuk dan menyelesaikan permasalahan waktu tunggu pasien rawat inap. Selain itu, algoritma klasifikasi yang digunakan dapat disematkan pada aplikasi dimasa mendatang untuk memudahkan pihak tenaga Kesehatan dalam melakukan *controlling* dan *monitoring* pasien dengan *hypertensive heart disease* (HHD) sehingga probabilitas terjadinya komplikasi pasien HHD dapat dikurangi.

1.5 Batasan Penelitian

Berikut ini merupakan batasan yang ditetapkan oleh peneliti dalam penelitian yang dilakukan

1. Data rekam medis yang digunakan antara lain plan perawatan pasien, Suhu(C), Tensi, Nadi(/mnt), Respirasi(/mnt), Tinggi (m), Berat (Kg) dan *Glasgow coma scale* (GCS).
2. Software yang digunakan untuk menganalisis data adalah *RapidMiner* dengan menggunakan algoritma *decision tree* C4.5, Jasper Reports, Microsoft Excel, Python dan SPSS.
3. Rekam medis yang digunakan yaitu dalam jangka waktu 1 Januari 2022 hingga 30 Desember 2022.

1.6 Sistematika Penulisan

Sistematika penulisan merupakan bagian yang dibuat agar penulisan laporan dilakukan secara jelas dan terstruktur. Berikut ini merupakan sistematika penulisan laporan tugas akhir secara garis besar.

BAB I PENDAHULUAN

Bab 1 merupakan bab pertama yang berisi mengenai penjelasan secara singkat terkait latar belakang penelitian ini dilakukan, masalah yang terjadi pada objek yang diteliti, Batasan-batasan masalah agar penelitian memperoleh tujuan yang jelas dan terarah, tujuan dilakukannya penelitian, manfaat penelitian yang dilakukan, bagi peneliti maupun objek penelitian dan universitas, serta sistematika penelitian.

BAB II KAJIAN PUSTAKA

Bab 2 dalam penelitian ini yaitu kajian Pustaka. Kajian Pustaka sendiri terdiri dari landasan teori dan juga kajian literatur yang digunakan sebagai landasan dalam proses penyelesaian masalah yang terjadi. Pada bagian kajian literatur berisi mengenai penelitian terdahulu yang memiliki *similarity* terhadap penelitian yang dilakukan. Sedangkan landasan teori, berisi teori maupun konsep yang akan digunakan dalam penelitian yang dilakukan.

BAB III METODE PENELITIAN

Bab 3 berisi penjelasan secara singkat mengenai metode penelitian yang akan digunakan dalam penelitian. Penjelasan secara singkat meliputi objek penelitian, metode yang digunakan, subjek penelitian serta alur penelitian yang digambarkan dengan menggunakan diagram alir (*flowchart*).

BAB IV PENGUMPULAN DAN PENGOLAHAN DATA

Pada bagian bab 4 berisi terkait data yang akan digunakan dalam penelitian, mulai dari tahapan *pre-processing* hingga menghasilkan suatu *insight* yang berasal dari pengolahan data yang dilakukan menggunakan metode analisis klasifikasi. Data yang digunakan dalam penelitian ini adalah data rekam medis pasien pada rumah sakit As-Syifa Manna, Bengkulu Selatan.

BAB V PEMBAHASAN

Bab 5 berisi pembahasan atau analisis terhadap hasil pengolahan data yang telah dilakukan. Bab ini kan memberikan informasi berharga dari suatu kumpulan data besar serta memberikan penjelasan terkait performa metode yang digunakan.

BAB VI PENUTUP

Bab 6 merupakan bab terakhir yang berisi penutup. Bagian ini berisi kesimpulan serta saran dari penelitian yang dilakukan oleh penulis. Kesimpulan yang diberikan akan menjawab rumusan masalah penelitian serta berdasarkan kesimpulan, penulis dapat memberikan saran kepada pihak rumah sakit maupun penelitian yang akan dilakukan di masa mendatang.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Literatur

Berikut ini merupakan kajian induktif guna mendukung penelitian yang dilakukan.

Tabel 2. 1 Kajian Literatur

No	Penulis dan Tahun	Metode	Hasil
1.	(Tayefi et al., 2017)	Klasifikasi, <i>Decision Tree</i>	Hasil yang penelitian yang diperoleh dalam penelitian ini dengan menggunakan subjek sebanyak 9078 subjek adalah terdapat sebanyak 14 aturan yang terbentuk dengan nilai sensitivity 68%, specificity 78%, accuracy 73% dan AUC 0.72
2.	(Santos et al., 2022)	<i>Decision Tree</i>	Data yang digunakan dalam penelitian ini terdiri atas 43,400 <i>instances</i> . Dalam percobaan eksperimen yang pertama dilakukan, pendekatan yang dilakukan memperoleh hasil sensitivitas dan spesifisitas yaitu sebesar 70% dan 80%. Sedangkan untuk eksperimen kedua dilakukan untuk melakukan evaluasi
3.	(Islam & Shamsuddin, 2021)	<i>CNN, Decision Tree</i>	Penelitian yang dilakukan menghasilkan sebanyak 9 aturan yang terbentuk dari algoritma decision tree. Secara performa, dapat dikatakan bahwa klasifikasi dengan algoritma CNN memiliki performa yang lebih konsisten dari segi akurasi, sensitifitas, spesifisitas dan <i>ROC Curve</i>
4.	(Ozcan & Peker, 2023)	Klasifikasi, <i>Decision Tree</i>	Berdasarkan penelitian yang telah dilakukan dapat disimpulkan bahwa model yang terbentuk telah berhasil tervalidasi dan memilih performa yang baik dengan akurasi sebesar 87% diikuti dengan parameter pengukuran performa sensitivitas, spesifisitas dan presisi yang masing-masing sebesar 85%, 90% dan 88%
5.	(Mohammed, 2020)	Klasifikasi	Berdasarkan penelitian yang dilakukan, dapat diketahui bahwa efek penggunaan metode

No	Penulis dan Tahun	Metode	Hasil
			resampling SMOTE pada algoritma klasifikasi dikatakan memberikan dampak positif. Akurasi model yang terbentuk mengalami peningkatan yang signifikan setelah diterapkan metode resampling SMOTE yang dikombinasikan dengan <i>Z-Score Normalization</i> untuk menangani <i>imbalanced data</i> yang terjadi.
6.	(D. Kugelman et al., 2022)	Klasifikasi	Model yang dihasilkan menunjukkan bahwa sebanyak 1091 pasien diberikan keputusan untuk mendapatkan tindakan rawat jalan dan 318 pasien mendapat tindakan rawat inap. Perbedaan signifikan antara pasien rawat jalan dan rawat inap dapat dilihat melalui BMI yang lebih tinggi, peningkatan usia, jenis kelamin laki-laki dan beberapa komorbid. Model menghasilkan nilai akurasi sebesar 78.7% dengan RUC sebesar 81.5%
7.	(Dyas et al., 2022)	<i>Multiple Logistic Regression.</i>	Dari seluruh pasien yang berjumlah 2,822,379 pasien, sebanyak 3,6% direkomendasikan mendapat perubahan perawatan yang awalnya rawat jalan menjadi rawat inap. Model yang terbentuk menghasilkan akurasi sebesar 90.6% berdasarkan desil ke 10 dan 71.3% berdasarkan indeks youden.
8.	(D. N. Kugelman et al., 2021)	Klasifikasi	Model yang dihasilkan berhasil melakukan prediksi dengan akurasi 63,3%. Total pasien yang memperoleh prediksi rawat jalan yaitu sebanyak 625 pasien sedangkan sebanyak 274 dinilai layak untuk rawat inap.
9.	(Nogay & Nogay, 2022)	<i>Decision Tree</i>	Berdasarkan penelitian yang telah dilakukan dengan menggunakan berbagai macam jenis Decision Tree, dapat diketahui performa dari algoritma berdasarkan hasil evaluasi dengan menggunakan parameter, hasil performa fine decision tree, medium decision tree, optimizable decision tree masing-masing 81.1%, 86.2%, 87.8%, 88.1%
10.	(Sun et al., 2018)	<i>Decision Tree, SMOTE</i>	Berdasarkan hasil penelitian yang dilakukan, model yang telah dilakukan pengujian diketahui bahwa metode oversampling SMOTE lebih berfokus kepada kelas minoritas atau kelas dengan label positif dalam suatu dataset. Selain itu model yang telah dilakukan evaluasi dapat diketahui bahwa SMOTE-DT dan Bagging-DT dapat dilakukan integrasi dengan decision tree yang mana kedua metode oversampling ini memberikan hasil performa yang baik bahkan lebih meningkatkan <i>G - measure</i> , <i>F -measure</i> dan akurasi secara menyeluruh.
11.	(Santoni et al., 2020)	<i>Decision Tree,</i>	Dalam penelitian yang dilakukan, model yang dievaluasi diketahui bahwa algoritma analisis

No	Penulis dan Tahun	Metode	Hasil
		<i>Naïve Bayes dan Artificial Neural Network</i>	klasifikasi artificial neural network memiliki nilai akurasi, recall dan presisi masing-masing yaitu 94.7%, 91.5%, 97.7%. decision tree memiliki nilai akurasi, recall dan presisi masing-masing yaitu 83%, 76.6%, 87.8% diikuti dengan algoritma naive bayes memiliki nilai akurasi, recall dan presisi masing-masing yaitu 74.5%, 68.1%, 78%.
12.	(Pambudi et al., 2022)	<i>Decision Tree</i>	Berdasarkan hasil penelitian yang dilakukan, diperoleh sebanyak 23 aturan berdasarkan algoritma decision tree C4.5 dengan 14 aturan menyatakan tidak stroke dan 9 aturan lainnya menyatakan stroke. Selain itu setelah dilakukan evaluasi pada model diketahui bahwa akurasi yang didapatkan adalah sebesar 99.07%.
13.	(Azwanti & Elisa, 2019)	<i>Decision Tree</i>	Berdasarkan hasil penelitian yang dilakukan dapat diketahui bahwa algoritma Decision Tree C4.5 dapat digunakan untuk mengklasifikasikan pola penyakit hipertensi sebagai salah satu penyebab gagal jantung, gagal ginjal, diabetes, stroke dan hipoglikemia. Penelitian yang dilakukan juga memberikan informasi bahwa penyakit hipertensi merupakan salah satu penyebab utama dalam gagal jantung.
14.	(Alham, 2021)	<i>Decision Tree</i>	Berdasarkan penelitian yang dilakukan dengan menggunakan pembagian data rasio 70:30 yang berdasarkan atribut penelitian yaitu usia, jenis kelamin, faktor keturunan, tekanan darah sistolik dan diastolik, level kolestrol, HDL, LDL dan tingkat <i>triglyceride</i> didapatkan akurasi sebesar 94,4%.
15.	(Nurlia & Enri, 2021)	<i>Decision Tree</i>	Berdasarkan penelitian yang dilakukan, pada percobaan pertama dengan menggunakan decision tree C4.5 dengan basis normal diperoleh akurasi sebesar 77.89%. sedangkan untuk performa decision tree dengan menggunakan basis <i>forward selection</i> didapatkan akurasi sebesar 84.29% dengan nilai AUC sebesar 0,785.

Berdasarkan kajian literatur yang telah dilakukan, menunjukkan bahwa, analisis klasifikasi khususnya algoritma *decision tree* memang dapat digunakan untuk melakukan prediksi, salah satunya melakukan prediksi terhadap keputusan *plan* perawatan pasien berupa rawat jalan atau rawat inap dengan menghasilkan performa yang baik. Selain itu mengangani ketimpangan data (*imbalanced data*) menggunakan metode SMOTE, terbukti membantu performa model klasifikasi yang dihasilkan menjadi lebih baik.

2.2 Research Gap

Berikut ini merupakan perbedaan penelitian yang saat ini dilakukan dengan penelitian terdahulu.

Tabel 2. 2 *Research Gap*

Penulis	Klasifikasi	Decision Tree	SMOTE	Imbalanced Data	Hipertensi	Rekam Medis	Prediksi Plan perawatan
(D. Kugelman et al., 2022)	√					√	√
(Dyas et al., 2022)						√	√
(D. N. Kugelman et al., 2021)	√					√	√
(Fiandra et al., 2017)	√	√				√	
(Islam & Shamsuddin, 2021)	√	√			√	√	
(Chu et al., 2021)	√	√			√	√	
(Sun et al., 2018)	√	√	√	√			
(Lestari, 2020)	√	√	√	√		√	
(Ozcan & Peker, 2023)	√	√				√	
(Santos et al., 2022)	√	√			√	√	
(Santoni et al., 2020)	√	√	√	√	√	√	
(Tayefi et al., 2017)	√	√			√	√	
(Mohammed, 2020)	√	√	√	√		√	
(Quesado et al., 2022)	√	√				√	
(Jayasri & Aruna, 2022)	√	√				√	
(Martinez-Ríos et al., 2021)	√	√			√	√	
(Nurmasani & Pristyanto, 2021)	√			√	√	√	

Penulis	Klasifikasi	Decision Tree	SMOTE	Imbalanced Data	Hipertensi	Rekam Medis	Prediksi Plan perawatan
(Gumelar et al., 2021)	√		√	√			
(Nisa' et al., 2020)	√			√			
(Mienye & Sun, 2021)	√	√	√	√		√	
Usulan	√	√	√	√	√	√	√

Berdasarkan tabel 2.2 yang menggambarkan *research gap* dalam penelitian ini, *novelty* ataupun keterbaharuan yang terkandung dalam penelitian ini yaitu terdapat pada tujuan penelitian, kondisi data di lapangan, data penyakit yang di gunakan, lokasi penelitian dan waktu penelitian yang dilakukan. Penelitian yang dilakukan bertujuan untuk melakukan prediksi *plan* perawatan terhadap pasien pengidap penyakit *hypertensive heart disease*, dengan menggunakan data rekam medis berupa data tanda-tanda vital tubuh manusia. Data rekam medis dalam penelitian ini mengalami ketimpangan data (*imbalanced data*) antara pasien rawat inap dan rawat jalan sehingga dibutuhkan metode tambahan untuk mengatasi ketimpangan data yang terjadi. Pada penelitian ini juga akan dilakukan perbandingan terhadap performa model sebelum dan setelah ketimpangan yang terjadi diatasi. Prediksi *plan* perawatan pasien akan dilakukan dengan memanfaatkan algoritma analisis klasifikasi yaitu *decision tree* C4.5

2.3 Landasan Teori

2.3.1 Knowledge discovery in database

Knowledge Discovery in Database merupakan salah satu cara yang digunakan untuk menggali pengetahuan secara lebih mendalam pada *database*. Dalam suatu basis data, terdapat sejumlah tabel yang saling berhubungan atau berelasi satu sama lain. *Knowledge Discovery in Database* digunakan untuk untuk menjelaskan proses untuk menemukan atau menggali informasi yang tersembunyi dalam data yang memiliki volume yang besar (Yuli Mardi, 2019). Dalam prosesnya, KDD dapat digambarkan secara garis besar dengan beberapa proses, yaitu sebagai berikut.

1. Data Selection

Pada tahapan ini, akan dilakukan proses pemilihan data (seleksi) yang akan digunakan dalam penerapan dari *Knowledge Discovery in Database*. Data yang telah berhasil

diseleksi, akan disimpan dalam berkas yang terpisah dari *dataset* operasional aslinya. Sehingga data yang digunakan adalah data yang murni akan digunakan untuk proses *Knowledge Discovery in Database*.

2. *Pre-Processing (Data Cleaning)*

Setelah dilakukannya proses pemilihan data yang akan digunakan, perlu dilakukan proses untuk membersihkan data yang menjadi fokus dalam proses yang akan dilakukan. *Cleaning* yang dimaksud dalam tahapan ini adalah antara lain membuang duplikasi data, memeriksa *missing value*, memeriksa data yang mengalami inkonsisten hingga memperbaiki kesalahan dalam penulisan data.

3. *Transformation*

Pada tahapan ini, data yang telah dipilih dan dibersihkan akan mengalami transformasi. Transformasi dilakukan dengan tujuan agar data yang tersedia sesuai dengan proses *data mining* yang akan dilakukan. Proses transformasi dalam KDD dapat dilakukan dengan *coding* yang mana hal ini merupakan proses kreatif dan sangat tergantung pada metode *data mining* yang akan dilakukan dan informasi seperti apa yang ingin dicari dalam suatu basis data.

4. *Data Mining*

Data mining merupakan proses pencarian atau penggalian pola informasi yang menarik dalam data yang telah diproses pada proses KDD sebelumnya dengan menggunakan metode dan algoritma tertentu. Pemilihan metode dan algoritma dalam data mining disesuaikan dengan tujuan yang dilakukan tergantung informasi seperti apa yang ingin didapatkan.

5. *Interpretation/Evaluation*

Proses terakhir dalam KDD adalah menyajikan data yang telah diolah sebelumnya dengan teknik *data mining* menjadi lebih mudah untuk dipahami oleh para *stakeholder*. Selain itu, pada tahapan ini juga dilakukan evaluasi dari model dan hasil yang telah terbentuk dengan melakukan pemeriksaan apakah informasi yang diperoleh melalui proses *data mining* telah sesuai dengan fakta yang ada dilapangan atau tidak.

2.3.2 *Data mining*

Menurut penelitian yang dilakukan oleh Yang et al (2020), *data mining* merupakan metode yang digunakan untuk menggali informasi dan pengetahuan yang tersembunyi dari sekumpulan

data yang bervolume besar. informasi yang ingin diraih harus memenuhi 3 karakteristik, yaitu belum pernah diketahui sebelumnya, efektif dan praktis. *Data mining* sendiri terbagi menjadi 2 jenis, yaitu *predictive* dan *descriptive*. Menurut Hamonangan Nasution & Iqbal Sofyan (2020), kata “*mining*” merupakan perumpamaan dari bahasa Inggris yang artinya menambang atau menggali sumber daya di dalam tanah, sehingga data mining merupakan proses penambangan informasi yang tersembunyi di dalam suatu kumpulan data yang bervolume sangat besar. *Data mining* berkaitan erat dengan bidang ilmu lainnya seperti *Artificial Intelligent*, *machine learning*, statistik dan *database*. Data mining sendiri dibagi menjadi beberapa teknik sesuai dengan fungsinya masing-masing, yaitu sebagai berikut.

1. Analisis Klasifikasi

Teknik yang dilakukan dengan melihat label kelas dan atribut dari suatu kelompok. Teknik ini digunakan untuk memberikan klasifikasi pada data dan menggunakan hasil dari klasifikasi menjadi sejumlah aturan (*rules*). Dalam penerapannya, analisis klasifikasi menggunakan beberapa algoritma yaitu antara lain *Decision Tree*, *Naive Bayes*, dan *Support Vector machine* (Erwansyah et al., 2021). Secara definisi, klasifikasi merupakan suatu proses yang digunakan untuk mendapatkan model yang dapat mendeskripsikan dan membedakan setiap data ke dalam kelas-kelas data. Hal ini dilakukan untuk menemukan pola data yang memiliki informasi dari sekumpulan data yang besar. Dilihat dari datanya, klasifikasi hanya memiliki satu atribut yang akan diprediksi yang disebut sebagai label kelas atau atribut target. Selain dari atribut target atau label kelas disebut sebagai atribut prediktor (Dwiasnati & Devianto, 2019).

2. *Clustering*

Analisis cluster merupakan suatu teknik dalam ilmu statistic yang bertujuan untuk mengelompokkan objek-objek yang ada ke dalam suatu kelompok sehingga objek yang telah dikelompokkan menjadi satu memiliki tingkat kesamaan yang tinggi apabila dibandingkan dengan objek yang berada di kelompok lain maupun diluar kelompoknya (Ningsih et al., 2019).

3. Asosiasi

Asosiasi merupakan salah satu tahapan yang digunakan untuk mencari hubungan atau relasi dari suatu item dengan item lainnya pada sekumpulan data yang berjumlah besar. Tujuan dari asosiasi sendiri adalah untuk menggali informasi yang tersembunyi dalam

item yang saling berhubungan dan melaporkan hasil pengolahan ke dalam bentuk *rules* (aturan) yang ditandai dengan penggunaan “*if*” dan “*THEN*”.

2.3.3 *Decision tree*

Merupakan salah satu algoritma supervised learning dalam data mining, algoritma ini berisi berbagai faktor yang dapat digunakan sebagai pemecah masalah. Penggunaan algoritma *decision tree* dinilai sangat baik dalam pemecahan masalah klasifikasi. Hal ini dikarenakan hasil prediksi dapat diketahui hanya dengan melihat pohon keputusan yang telah terbentuk (Ayudhitama et al., 2020). Algoritma *decision tree* C4.5 adalah algoritma hasil pengembangan algoritma *decision tree* yang telah ada sebelumnya yaitu algoritma ID3. Pada *decision tree* terdapat 3 tipe jenis *node* yang menyusun berdirinya diagram pohon keputusan, yaitu.

1. *Root node*

Root node dalam diagram pohon keputusan terletak dibagian tertinggi yang mana, pada *node* ini tidak memiliki *input* dan tidak memiliki *output* dan juga dapat memiliki *output* lebih dari satu.

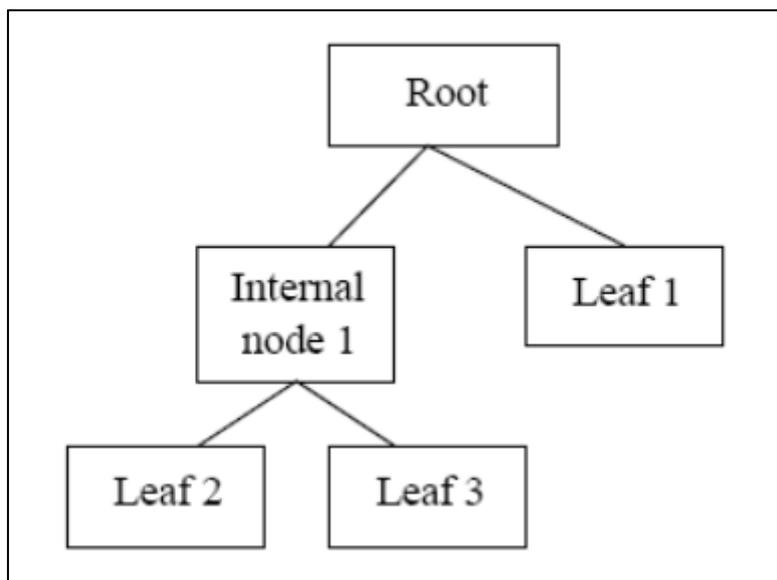
2. *Internal Node*

Node ini merupakan percabangan yang mana pada *node* ini hanya memiliki satu input dan mempunyai output paling sedikit sebanyak 2 *output*.

3. *Leaf node*

Leaf node merupakan node paling akhir pada diagram pohon keputusan. Node ini hanya memiliki satu input dan tidak memiliki output.

Berikut ini merupakan contoh dari diagram *Decision tree*.



Gambar 2. 1 Struktur diagram *decision tree*
 Sumber: (Defiyanti, 2019)

Kelebihan yang paling utama dari algoritma *decision tree* menurut penelitian yang dilakukan oleh Ratniasih (2016), antara lain:

1. Pengambilan keputusan dengan menggunakan pohon keputusan dapat menjadi lebih sederhana dan lebih jelas. Dibandingkan tanpa menggunakan pohon keputusan yang bersifat sangat rumit dan sangat umum.
2. Dalam *decision tree*, *sample* yang diuji hanya berdasarkan kriteria dan kelas atribut yang terpilih, sehingga bisa lebih efektif dan efisien.
3. Pemilihan fitur *internal node* yang berbeda dalam algoritma *decision tree* lebih fleksibel. Fitur yang telah terpilih akan membedakan antara setiap atribut atau kriteria yang berbeda pada *node* yang sama. Fleksibilitas inilah salah satu yang membuat metode *decision tree* mengalami peningkatan performa dalam kualitas keputusan dibandingkan dengan metode perhitungan bertahap yang lebih tergolong kepada metode konvensional
4. *Decision tree* merupakan salah satu teknik analisis multivariat yang memiliki atribut dan kelas data yang sangat besar. *Decision tree* dapat meminimalisir terjadinya kesalahan dalam mengestimasi parameter tertentu dengan menggunakan kriteria yang telah di seleksi pada setiap *internal node* tanpa harus mengorbankan kualitas keputusan.

Fleksibilitas dari *decision tree* membuat metode ini sangat dapat diandalkan, hal ini dikarenakan *decision tree* memiliki hasil akhir berupa visualisasi diagram pohon keputusan. Berdasarkan hal tersebut, prediksi dalam *decision tree* dapat diamati dari awal hingga akhir.

2.3.4 Perbandingan *decision tree* C4.5 dan ID3

Terdapat beberapa algoritma dalam metode klasifikasi *decision tree* yang dapat digunakan untuk melakukan prediksi kelas data, contohnya adalah C4.5 dan ID3. Salah satu jenis algoritma dari metode *decision tree* yaitu ID3 yang merupakan singkatan dari *Iterative Dichotomiser* atau *Induction of Decision 3*, yang pada awal penemuannya dicetuskan dan dioptimalkan oleh Quinlan. Untuk menggunakan algoritma ID3, menurut penelitian yang dilakukan oleh Munthe & Sihombing (2018). Secara umum, terdapat beberapa langkah yang harus dilalui, yaitu:

1. Langkah pertama pada induksi *decision tree* akan selalu dimulai dengan *root node*
2. Perhitungan nilai *entropy* dengan menggunakan seluruh data latih (*training data*) pada *node*
3. Pemilihan *root node* dilihat berdasarkan nilai *information gain* tertinggi
4. *Root node* yang telah terpilih digunakan sebagai *node* pemecahan menjadi cabang
5. Langkah poin b hingga poin d dilakukan secara terus menerus dan berulang (rekursif) hingga seluruh data dalam setiap *node* menghasilkan *output* satu label kelas (*leaf node*), *leaf node* merupakan *node* yang tidak dapat dilakukan pemecahan lanjutan yang berisi keputusan akhir (label kelas)
6. Ketika atribut telah terpilih menjadi *node*, atribut tersebut tidak lagi dimasukkan ke dalam perhitungan nilai *gain*.
7. Proses yang berulang secara terus menerus (rekursif) dapat berhenti dengan kondisi sebagai berikut.
 - a) Semua data dari anak cabang sudah termasuk ke dalam kelas yang sama.
 - b) Seluruh atribut dalam suatu *dataset* telah digunakan, namun masih ada data yang belum digunakan pada kelas yang berbeda.

Sedangkan pada algoritma *decision tree*, menurut penelitian yang dilakukan oleh (Setio et al., 2020) algoritma *decision tree* C4.5 melewati beberapa tahapan, antara lain:

1. Menentukan *data training* dan *data testing*

Algoritma *Decision Tree* C4.5 memiliki 2 input yang harus dipenuhi, yaitu *data training* dan *data testing*. *Data training* dalam klasifikasi berfungsi sebagai data yang digunakan

untuk membangun model yang akan menghasilkan diagram pohon keputusan. Sedangkan testing data (data uji) merupakan data yang berfungsi untuk menguji dan mengevaluasi model klasifikasi yang telah terbentuk dengan menggunakan beberapa parameter (Setio et al., 2020). Terdapat beberapa metode dalam pemberian label pada data, yaitu (Saputri, 2018):

a. *Simple Split*

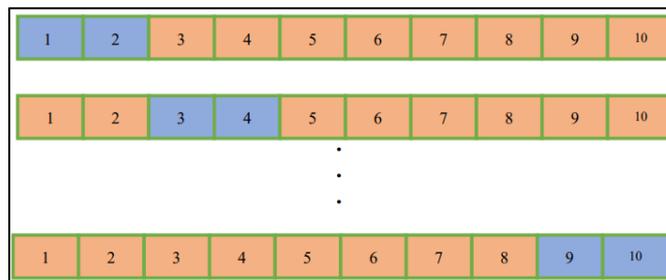
Simple split atau dengan nama lain *Hold Out Method* merupakan metode untuk membagi *dataset* yang ada menjadi data *training* dan testing. Secara umum, data testing mengambil bagian sebanyak 1/3 dari keseluruhan data dan sisanya dijadikan data *training*. Berikut ini merupakan ilustrasi dari pembagian data *simple split*.

Training 90%	Test 10%
Training 80%	Test 20%
Training 70%	Test 30%
Training 60%	Test 40%
Training 50%	Test 50%
Training 40%	Test 60%
Training 30%	Test 70%
Training 20%	Test 80%
Training 10%	Test 90%

Gambar 2. 2 *Simple split*
Sumber: (Dwi Untari et al., 2018)

b. *Cross Validation Method*

Metode *cross validation* adalah menjalankan sebuah model menggunakan k-1 *folds* yang dijadikan data *training*. Lalu hasil dari model tersebut divalidasi pada bagian data yang tersisa yaitu data testing. Dengan adanya validasi dapat dimanfaatkan untuk mengetahui seberapa akurat model klasifikasi yang telah dihasilkan.

Gambar 2. 3 *Cross validation*

Sumber: (Nurhopipah & Hasanah, 2020)

2. Menghitung nilai *Entropy*

Entropy adalah fungsi atau formula yang digunakan untuk mengetahui tingkat heterogenitas dari suatu kumpulan sampel data. Semakin kecil keberagaman (heterogenitas), semakin baik juga model dalam melakukan ekstraksi suatu kelas. Berikut ini merupakan formula yang digunakan untuk melakukan perhitungan *entropy*.

$$Entropy = \sum_{i=1}^n -p_i \times \log_2 p_i \quad (2.1)$$

Dengan,

S : Himpunan dataset (kasus)

n : banyaknya partisi S

i : Jumlah partisi S

p_i : probabilitas S_i terhadap S = $\frac{S_i}{S} \left(\frac{\text{himpunan}(ya)}{\text{himpunan total kasus}} \right)$

Dengan menggunakan rumus (1) yang telah dipaparkan sebelumnya, dapat dilakukan perhitungan pada kasus berikut. Dimisalkan ada kasus dengan variabel independen yang memiliki total sebanyak 13 kasus dengan label kelas *yes* sebanyak 8 kasus dan *no* sebanyak 5 kasus. Salah satu variabel independen yang menyusun kasus tersebut adalah variabel *rain* dengan dua tingkatan yaitu *strong* dan *medium*.

Tabel 2. 3 *Rain Level*

	Total	No	Yes
<i>High</i>	7	2	5
<i>Low</i>	6	3	3

Maka berdasarkan formula *entropy*, didapatkan hasil perhitungan *entropy* pada partisi *high* dan *low* adalah:

$$\begin{aligned} Entropy (S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{2}{7} \times \log_2 \left(\frac{2}{7} \right) \right) + \left(-\frac{5}{7} \times \log_2 \left(\frac{5}{7} \right) \right) = 0,86 \end{aligned}$$

$$\begin{aligned} Entropy (S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{3}{6} \times \log_2 \left(\frac{3}{6} \right) \right) + \left(-\frac{3}{6} \times \log_2 \left(\frac{3}{6} \right) \right) = 1 \end{aligned}$$

3. Menghitung nilai *Gain*

Nilai gain dalam analisis klasifikasi digunakan untuk mengetahui tingkat pengaruh dari suatu atribut terhadap keputusan yang dihasilkan. Sehingga dapat disimpulkan bahwa nilai *gain* merupakan ukuran yang menyatakan tingkat efektivitas suatu atribut dalam melakukan klasifikasi data. Nilai *gain* dapat dihitung dengan menggunakan rumus berikut.

$$Gain (S, A) = Entropy (S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(A) \quad (2.2)$$

Keterangan:

S = himpunan *dataset* (kasus)

A = Atribut

n = jumlah partisi atribut A

$|S_i|$ = Jumlah kasus pada partisi ke- i

$|S|$ = jumlah kasus S atau himpunan kasus

Jika dalam kasus ini diketahui *entropy* untuk himpunan dataset atau himpunan kasus (S) adalah sebesar 0.96, maka hasil perhitungan nilai *gain* yaitu sebagai berikut.

$$\begin{aligned} Gain (S, A) &= Entropy (S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(A) \\ &= 0.96 - \left(\left(\frac{7}{13} \times 0.86 \right) + \left(\frac{6}{13} \times 1 \right) \right) = 0,03 \end{aligned}$$

4. *Split information*

Setelah mendapat nilai *gain*, proses perhitungan dilanjutkan untuk menghitung nilai *split information*. Berdasarkan penelitian yang dilakukan oleh (Sinam & Lawan, 2019), *split information* pertama kali diperkenalkan oleh Quinlan Ross untuk memberikan perbaikan dari keterbatasan kemampuan yang terjadi pada algoritma ID3, yaitu latensi dan *over-fitting* yang memberatkan dalam sisi komputasi. Berikut ini merupakan persamaan yang digunakan untuk menghitung nilai *split information*.

$$\begin{aligned} \text{split info} &= \sum_{i=1}^c -\frac{S_i}{S} \times \log_2 \frac{S_i}{S} & (2.3) \\ &= \left(-\frac{7}{14} \times \log_2 \left(\frac{7}{14} \right) \right) - \left(\frac{6}{14} \times \log_2 \left(\frac{6}{14} \right) \right) = 0.061 \end{aligned}$$

5. *Gain ratio*

Menurut penelitian yang dilakukan oleh (Gupta, 2017), *gain ratio* digunakan untuk mengurangi efek bias akibat penggunaan *information gain*. Berikut ini merupakan persamaan yang digunakan untuk menghitung *gain ratio*.

$$\begin{aligned} \text{Gain ratio} &= \frac{\text{Gain}(S, A)}{\text{split info}} & (2.4) \\ \text{Gain ratio} &= \frac{\text{Gain}(S, A)}{\text{split info}} = \frac{0.03}{0.06} = 0.5 \end{aligned}$$

6. Melakukan *testing* terhadap model yang telah dibangun.

Pengujian model akan menggunakan 2 bagian data yang berbeda, yaitu data training dan data *testing*. Dalam analisis klasifikasi, terdapat 2 metode yang dapat digunakan untuk menentukan besaran dari jumlah data *training* dan data *testing*, yaitu *simple split* dan *cross validation*. Menurut penelitian yang dilakukan oleh (Pamuji & Ramadhan, 2021), *10-fold cross validation* merupakan metode validasi yang membagi suatu kumpulan data menjadi 10 bagian yang berukuran sama besar dengan cara pengacakan data. Proses dari metode ini sendiri adalah dengan cara melakukan perulangan sebanyak k kali, sehingga setiap bagian dari *dataset* berperan menjadi *data testing* sebanyak satu kali (Nurhopipah & Hasanah, 2020). Pemisahan masing-masing bagian berisi sebanyak $(k-1)$ *fold data training* dan *1-fold* untuk *data testing*. *10-fold cross validation* merupakan *k-fold cross validation* yang direkomendasikan untuk pembagian data

karena relatif memberikan akurasi yang menghasilkan bias paling kecil dibandingkan dengan *fold cross validation* lain (Mubaroroh et al., 2022).

7. Evaluasi hasil klasifikasi.

Evaluasi dalam analisis klasifikasi dapat dilakukan dengan alat bantu bernama *confusion matrix*. Tabel *confusion matrix* merupakan tabel yang dapat menjelaskan seberapa baik model klasifikasi yang dihasilkan dalam melakukan prediksi. Dalam *confusion matrix* akan dijabarkan jumlah baris data uji yang diprediksi benar dan tidak benar. Melalui *confusion matrix*, model klasifikasi yang dihasilkan dapat diketahui performanya dalam melakukan klasifikasi data.

2.3.5 *Imbalance dataset*

Imbalance data merupakan suatu kondisi dimana terjadinya ketimpangan data antara data dengan kelas mayoritas dengan data kelas minoritas. Kondisi ketimpangan yang terjadi dapat membuat *machine learning* menjadi keliru dalam melakukan klasifikasi sehingga performa yang dihasilkan tidak maksimal (Mutmainah, 2021). *Imbalance data* dapat diatasi dengan menyeimbangkan distribusi kelas data mayoritas dan kelas data minoritas. Untuk menyeimbangkan kelas data mayoritas dan kelas data minoritas, dapat dilakukan dengan menerapkan teknik *random oversampling*.

Salah satu teknik *random oversampling* yang dikenal sangat populer dan merupakan salah satu metode *pre-processing* dan *sampling data* yang paling berpengaruh dalam dunia *machine learning* maupun *data mining* adalah SMOTE (Fernández et al., 2018). SMOTE merupakan metode *oversampling* yang digunakan untuk meningkatkan sampel kelas data rawat inap (minoritas) dengan data rawat jalan (mayoritas) dengan cara menduplikasi data Minoritas agar memiliki jumlah yang seimbang dengan data mayoritas sehingga hasil evaluasi model akan lebih baik dibandingkan sebelum digunakannya SMOTE (Saputro & Rosiyadi, 2022).

2.3.6 Metode *Resampling*

Terdapat beberapa Teknik yang dapat digunakan untuk mengatasi ketimpangan data yang terjadi. Berikut ini merupakan Teknik metode *oversampling* dan *undersampling* yang dapat digunakan menurut penelitian yang dilakukan oleh (Sir & Soepranoto, 2022).

Tabel 2. 4 Perbandingan teknik oversampling

Metode	Jenis Sampling	Cara Kerja
Random Oversampling	<i>Oversampling</i>	<p>Metode ini bekerja dengan cara memilih <i>instance</i> kelas data minoritas dalam dataset secara acak, yang kemudian dilakukan duplikasi secara berkali-kali hingga jumlah kelas data minoritas menyamai jumlah kelas data mayoritas. Cara kerja dari ROS membuat kemungkinan terjadinya <i>overfitting</i> sangat tinggi.</p> <p>Teknik <i>oversampling</i> yang digunakan cenderung lebih baik dibandingkan dengan ROS. SMOTE lebih cerdas dibandingkan dengan ROS yang ditandai dengan cara kerja yang mengambil secara acak <i>nearest neighbour</i> sebanyak k dari kelas minoritas dan membuat data sintesis baru. Dengan pendekatan SMOTE, model dapat lebih kebal terhadap terjadinya <i>overfitting</i>.</p>
SMOTE	<i>Oversampling</i>	<p>Cara kerja metode ini yaitu dengan menggunakan bobot distribusi pada kelas data</p>
ADASYN	<i>Oversampling</i>	

Random Undersampling

Undersampling

Tomek Links

Undersampling

minoritas berdasarkan pada tingkat kesulitan model dalam melakukan pembelajaran. Data sintetis dalam metode ini dihasilkan melalui kelas data minoritas yang sulit untuk di pelajari dibandingkan dengan data kelas minoritas yang mudah untuk dipelajari. Metode ini bekerja dengan cara memilih secara acak instance yang terdapat dalam *dataset* pada kelas mayoritas kemudian mengeliminasi kelas data tersebut. Proses yang sama akan dilakukan secara berulang hingga target *sampling* dapat terpenuhi. Metode ini memiliki kelemahan yaitu dikarenakan banyaknya instance yang dihapus, kemungkinan untuk kehilangan informasi penting menjadi semakin besar, akibatnya akan memengaruhi kinerja dari model klasifikasi itu sendiri. Metode ini bekerja dengan cara mengurangi data minoritas dan data mayoritas

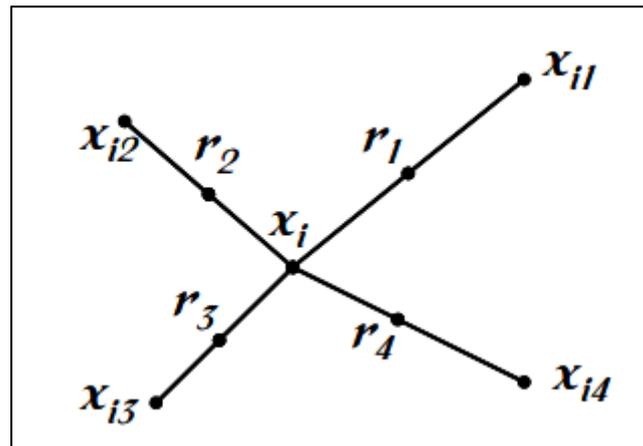
*ENN**Undersampling*

yang saling berdekatan. Sehingga bisa dikatakan metode ini merupakan metode yang akan mengeliminasi data *outlier*. Penggunaan metode ini tidak dapat benar-benar menyeimbangkan kelas data minoritas dan mayoritas karena pengurangan jumlah kelas data mayoritas bergantung pada jumlah kelas data minoritas dan mayoritas yang saling berdekatan. Metode ini menghilangkan sampel dari kelas data mayoritas yang memiliki perbedaan label jika dibandingkan dengan data-data yang saling berdekatan. Output dari metode ini bergantung pada jumlah k yang dipilih.

2.3.7 Synthetic Minority Over-sampling Technique (SMOTE)

Synthetic Minority Over-sampling Technique atau yang lebih dikenal dengan nama singkatan SMOTE, merupakan salah satu teknik *random oversampling* yang digunakan untuk menyeimbangkan kelas data minoritas dengan kelas data mayoritas pada suatu kumpulan data. Cara kerja dari teknik ini adalah dengan cara mengambil perbedaan yang terdapat antara vektor *sample* lalu mengalikan perbedaan yang ada dengan bilangan acak antara 0 dan 1 dan menambahkan hasilnya ke dalam vektor fitur yang telah ada sebelumnya. Hal ini akan memilih

secara acak titik di sepanjang segmen garis diantara dua fitur tertentu. Sehingga melalui pendekatan ini, dapat membuat kelas minoritas untuk menjadi lebih umum secara efektif (Chawla et al., 2002). Menurut penelitian yang dilakukan oleh Fernández et al (2018), SMOTE merupakan teknik *random oversampling* yang sangat terkenal dan berpengaruh di dunia *machine learning*. Berikut ini merupakan ilustrasi bagaimana SMOTE dapat menghasilkan titik data sintesis.



Gambar 2. 4 ilustrasi cara kerja SMOTE
Sumber: (Fernández et al., 2018)

Ilustrasi yang ditunjukkan oleh gambar 2.4 merupakan ilustrasi yang diwujudkan melalui rumus untuk membangkitkan data buatan dengan menggunakan metode SMOTE. Berikut ini merupakan rumus pembangkitan data buatan dengan SMOTE menurut (Wijayanti et al., 2021).

$$X_{new} = X_i + (\widehat{X}_k - X_i) \times \delta \quad (2.3)$$

X_{new} = data sintesis baru

X_i = data ke-i dari kelas minoritas

\widehat{X}_k = data k tetangga terdekat dari data X_i

δ = bilangan acak 0 dan 1

2.3.8 Rekam Medis

Berdasarkan peraturan menteri kesehatan No 269 tahun 2008, rekam medis diartikan sebagai berkas yang memuat di dalamnya catatan dan dokumen mengenai pasien yang berisi identitas pasien, hasil pemeriksaan, pengobatan, tindakan medis lainnya pada sarana pelayanan

kesehatan dengan tindakan rawat jalan maupun rawat inap baik dalam pengelolaan pemerintah maupun pihak swasta.

Menurut Departemen Kesehatan Republik Indonesia pada tahun 2002, rekam medis merupakan keterangan yang dibuat secara tertulis maupun terekam mengenai tindakan medis. Rekam medis memuat berbagai informasi, yaitu identitas pasien, anamnese pasien, hasil cek fisik laboratorium, diagnosa penyakit, pelayanan dan tindakan medis yang disarankan dan diterima, tindakan rawat inap, rawat jalan ataupun gawat darurat.

2.3.9 Manfaat Rekam Medis

Berdasarkan Peraturan Menteri Kesehatan No 269/MenKes/Per/III/ 2008 tentang rekam medis, terdapat beberapa manfaat dari rekam medis yang dijabarkan dalam permenkes tersebut, yaitu sebagai berikut.

1. Rekam medis dapat dijadikan sebagai acuan dasar dan petunjuk guna melakukan analisis dan perencanaan terhadap penyakit serta menentukan perawatan maupun tindakan medis yang akan diberikan kepada pasien
2. Pembuatan rekam medis untuk praktik kedokteran dengan jelas dan rinci akan meningkatkan kualitas pelayanan dan juga dapat melindungi tenaga kesehatan sehingga dapat terwujud kesehatan masyarakat yang optimal.
3. Rekam medis menjadi sumber informasi perkembangan kronologis penyakit, pelayanan medis, pengobatan dan tindakan medis. Selain itu juga bermanfaat untuk bahan studi untuk perkembangan pembelajaran, pengajaran dan penelitian pada bidang kedokteran dan kedokteran gigi.
4. Rekam medis dapat dijadikan sebagai petunjuk dan bahan pertimbangan untuk menetapkan biaya dalam pelayanan kesehatan pasien.
5. Rekam medis berperan dalam statistik kesehatan khususnya, untuk menentukan jumlah penderita hingga mempelajari perkembangan kesehatan dari masyarakat.
6. Rekam medis dapat dijadikan pembuktian masalah hukum. Disiplin dan etik rekam medis adalah alat bukti utama sehingga dapat dijadikan bukti apabila terdapat permasalahan hukum, disiplin dan etik.

2.3.10 Hypertensive Heart Disease

Hypertensive heart disease merupakan penyakit jantung yang pada awalnya disebabkan oleh komplikasi dari penyakit hipertensi. HHD terjadi ketika terdapat sejumlah perubahan pada

ventrikel bagian kiri atau *left ventricular hypertrophy (LVH)*, atrium kiri, dan pembuluh darah coroner karena mengalami peningkatan tekanan darah secara terus menerus dalam jangka waktu yang lama (Zulfa et al., 2022). Menurut penelitian yang dilakukan oleh Tackling & Borhade (2021), seseorang dikatakan memiliki penyakit hipertensi apabila memiliki tekanan darah sistolik lebih besar dari 120 mmHg atau memiliki tekanan *diastolic* lebih dari 80 mmHg. Risiko kematian akibat penyakit *cardiovascular* akan meningkat dua kali lipat setiap peningkatan tekanan darah sistolik sebesar 20 mmHg dan diastolik sebesar 10 mmHg ketika tekanan darah dasar adalah 115/75.

Kondisi pasien hipertensi sangat berpengaruh kepada kondisi genetik dari pasien dan lingkungan maupun gaya hidup dari pasien itu sendiri. Beberapa faktor risiko yang memengaruhi kondisi pasien hipertensi adalah usia, obesitas, diet sodium (lebih dari 3g/hari), tidak banyak melakukan gerak hingga konsumsi alkohol memiliki korelasi independen terhadap perkembangan kondisi hipertensi pasien. Peningkatan tekanan darah dan risiko dari penyakit *cardiovascular*, mengalami peningkatan secara substansial seiring bertambahnya usia.

1. Pasien dengan rentang usia 45-54 tahun – 36,1% pria dan 33,2% wanita
2. Pasien dengan rentang usia 55-64 tahun – 57,6% pria dan 55,5% wanita
3. Pasien dengan rentang usia 65-74 tahun – 63,6% pria dan 65,8% wanita
4. Pasien dengan usia 75 tahun atau lebih 73,4% pria dan 81,2% wanita.

Penelitian yang dilakukan Tackling & Borhade (2021) juga mengatakan bahwa, terdapat 4 tingkatan atau kategori klasifikasi dari tekanan darah.

1. Tekanan darah normal didefinisikan dengan tekanan darah sistolik yang berada dibawah 120 mmHg dan tekanan darah distolik kurang dari 80 mmHg
2. Tekanan darah tinggi didefinisikan dengan tekanan darah sistolik yang berada pada rentang 120-129 mmHg dengan tekanan darah distolik kurang dari 80mmHg
3. *Stage 1* didefinisikan dengan tekanan darah sistolik berada pada rentang 130-139 mmHg atau tekanan darah *distolic* berada pada rentang 80-89mmHg
4. *Stage 2* didefinisikan dengan tekanan darah *sistolik* lebih dari 140 mmHg atau tekanan darah distolik 90 mmHg atau lebih.

Pasien penderita penyakit hipertensi dapat dilakukan perawatan dengan menggunakan pengobatan *antihypertensive*

2.3.11 Indeks Masa Tubuh

Body mass index atau yang lebih dikenal sebagai indeks masa tubuh, merupakan parameter yang menunjukkan kategori berat badan dari seseorang. Indeks masa tubuh dapat menunjukkan klasifikasi berat badan seseorang apakah sudah proporsional atau belum. Melalui indeks masa tubuh, seseorang dapat mengetahui apakah berat badannya saat ini tergolong normal, overweight, underweight atau obesitas. Nilai indeks masa tubuh seseorang dapat dihitung menggunakan pembagian berat badan dalam kilogram dengan kuadrat dari tinggi badan dalam satuan meter (Mahfud Imam, Gumantan Aditya, 2020). Berikut ini merupakan kateogrisasi nilai indeks masa tubuh seseorang menurut Truesdale & Stevens (2008).

Tabel 2. 5 Klasifikasi indeks masa tubuh

Indeks masa tubuh	Kategori
< 18.5	Underweight
18.5 – 25	Healthy weight
25 – 30	Overweight
>30	Obesity

Sumber: (Truesdale & Stevens, 2008)

2.3.12 Denyut Nadi

Denyut nadi merupakan irama detak jantung yang dapat dirasakan dan diukur pada permukaan kulit di lokasi tertentu pada bagian tubuh. Pada jantung manusia dengan kategori normal, denyut dapat berasal dari nodus SA. Mengetahui denyut nadi sendiri dinilai penting karena, selain merupakan salah satu *vital sign*, melalui denyut nadi juga dapat diketahui status sirkulasi darah pada pasien. Menurut Richa Rachmawati et al (2022) terdapat 3 tingkatan dalam melakukan pengukuran denyut nadi, yaitu sebagai berikut.

Tabel 2. 6 Klasifikasi denyut nadi

Umur	Bradikardia	Normal	Takikardia
<1 tahun	< 100	100-160	> 160
1-10	70	70-120	> 120
11-17	60	60-100	>100
>17	60	60-100	>100

Sumber: (Richa Rachmawati et al., 2022)

2.3.13 Klasifikasi Suhu Tubuh Manusia

Berdasarkan penelitian yang dilakukan oleh Yuniarti et al (2020) suhu tubuh manusia terbagi menjadi 2 bagian, yaitu suhu tubuh untuk manusia yang dikategorikan anak-anak dan dewasa. Berikut ini merupakan tingkatan dari suhu tubuh manusia.

Tabel 2. 7 Klasifikasi tubuh Manusia

Kategori	Suhu	
	Anak-anak	Dewasa
Hipotermia	< 36	< 36
Normal	36.1 – 37.5	36.1 – 37.7
Demam	37.6 – 40	37.8 – 40
Hiperpireksia	> 40	> 40

Sumber: (Yuniarti et al., 2020)

2.3.14 Saturasi Oksigen (*SPO2*)

Saturasi oksigen adalah ukuran yang digunakan untuk mengetahui presentase kemampuan hemoglobin dalam membawa oksigen (Fadlilah et al., 2020). Nilai dari saturasi oksigen mengindikasikan presentase semua situs pengikatan hemoglobin yang ditempati oleh hemoglobin. Saturasi oksigen dapat diukur dengan bantuan alat bernama *pulse oximetry* yang merupakan alat non invasif yang dapat mengukur saturasi oksigen darah arteri melalui pemasangan alat pada ujung jari, ibu jari, hidung, daun telinga atau dahi. Menurut penelitian terdahulu yang dilakukan oleh Dini et al (2022), kadar saturasi oksigen (*SpO2*) diklasifikasi ke dalam 4 kelas, yaitu sebagai berikut.

Tabel 2. 8 Tingkat Saturasi Oksigen

<i>SpO2</i>	Tingkat Saturasi Oksigen
95-100	Normal
91-94	cukup
86-90	sedang
<85	rendah

Sumber: (Fadlilah et al., 2020).

2.3.15 Frekuensi Pernapasan (Respirasi)

Menurut Sari Ikhsan (2019), frekuensi pernapasan merupakan jumlah pernapasan yang dihitung mulai dari seseorang menghirup hingga mengeluarkan napas yang dihitung dengan satuan napas per menit (*breath per minute*). Frekuensi pernapasan sendiri tergolong ke dalam tanda-tanda vital tubuh (*vital sign*) karena merupakan salah satu parameter untuk mengetahui keadaan keseluruhan kerja tubuh manusia. Menurut penelitian yang dilakukan oleh Nejadgholi et al (2019), manusia dewasa normal memiliki frekuensi bernapas yaitu 12-20 kali bernapas dalam satu menit yang sama dengan 0.2-0.3 Hz. Berbeda dengan manusia dewasa, anak-anak memiliki frekuensi pernapasan yang lebih tinggi yaitu sebesar 30-60 kali per menit. Namun seiring

berjalannya waktu dan penambahan usia anak, frekuensi napas akan mulai berkurang menjadi 18-30 ketika anak menginjak usia 10 tahun.

2.3.16 *Glasgow Coma Scale*

Berdasarkan penelitian yang dilakukan oleh Hartanto et al (2019), *Glasgow coma scale* merupakan skala yang digunakan secara luas untuk mengetahui tingkat kesadaran seseorang yang sudah terbukti dan tervalidasi. GCS dinilai dengan melibatkan tiga aspek yang berbeda untuk dilakukan pengukuran antara lain, pembukaan mata, respon verbal dan respon motorik. Respon yang telah disebutkan tersebut akan dikonversi ke dalam nilai angka, semakin buruk kondisi respon, maka semakin buruk pula nilai yang akan diberikan. Berikut ini merupakan klasifikasi pemberian skor pada respon motorik seseorang.

Tabel 2. 9 Kategori nilai GCS

Nilai GCS	Kategori
15	Compos Mentis
12-14	Somnolen
8-11	Sopor
3-7	Koma

Sumber: (Hartanto et al., 2019)

2.3.17 *Usia*

Berdasarkan penelitian yang dilakukan oleh Tackling & Borhade (2021), diketahui bahwa usia merupakan faktor risiko yang memperburuk dan memperbesar peluang seorang pasien dapat terjangkit penyakit hipertensi. Berikut ini merupakan diskritisasi data usia menurut (Lesmana, 2012).

Tabel 2. 10 Kategori usia

Usia	Kategori
< 40 Tahun	Muda
40 – 59 Tahun	Sedang
> 60 tahun	Tua

Sumber: (Lesmana, 2012)

BAB III

METODE PENELITIAN

3.1 Subjek Penelitian

Data yang digunakan dalam penelitian ini adalah data sekunder yang bersumber dari *database* rumah sakit. Data yang digunakan dalam penelitian ini berfokus kepada pasien penderita penyakit *Hypertensive Heart Disease* yang dilakukan rekapitulasi mulai tanggal 1 Januari 2022 hingga 30 Desember 2022.

3.2 Objek Penelitian

Penelitian ini dilaksanakan di Rumah Sakit As-Syifa Jl. Gerak Alam No.RT.13, Kota Medan, Kec. Kota Manna, Kabupaten Bengkulu Selatan, Bengkulu 38513. Metode yang digunakan dalam penelitian ini adalah kombinasi metode SMOTE dan *Decision tree* C4.5 yang digunakan untuk memprediksi tindakan medis yang akan diberikan kepada pasien.

3.3 Variabel Penelitian

Berikut ini merupakan penjelasan secara rinci terkait variabel penelitian yang digunakan untuk melakukan prediksi tindakan medis yang akan diterima oleh pasien.

Tabel 4. 1 Variabel penelitian

	Variabel	Satuan	Keterangan	Referensi
Dependen	Y = <i>plan</i> perawatan	Rawat jalan/inap	Tindakan medis yang direkomendasikan setelah hasil pemeriksaan awal dilakukan.	
Independen	X_1 = Umur	Muda, Sedang, Tua	Umur ketika pasien melakukan perawatan	(Wang et al., 2012)
	X_2 = Jenis Kelamin	Laki-laki dan Perempuan	Jenis kelamin pasien	
	X_3 = suhu	Hipotermia, normal, demam hipereksia	Suhu badan pasien saat dilakukan pemeriksaan	(Auble et al., 2005).
	X_4 = Nadi	Bradikardia, normal, takikardia	Tingkatan nadi pasien dalam satuan menit	(Wang et al., 2012)

Variabel	Satuan	Keterangan	Referensi
$X_5 = \text{Respirasi}$	Normal, tidak normal	Frekuensi bernafas dalam satu menit	
$X_6 = \text{Body Mass Index}$	<i>Underweight, healthy weight, overweight, obesity</i>	BMI merupakan satuan yang digunakan untuk memperkirakan lemak tubuh yang didasarkan kepada tinggi dan berat badan. BMI diperoleh melalui massa tubuh dibagi dengan tinggi dikuadratkan	
$X_7 = \text{Tingkat Hipertensi}$	Normal, Pra-hipertensi, stage 1 dan stage 2	Kondisi atau tingkatan hipertensi yang diderita pasien ketika mendapatkan perawatan yang diperoleh dari klasifikasi tekanan darah sistolic dan dyastolic	
$X_8 = \text{Glasgow coma scale (GCS)}$	Compos Mentis, Sopor, Somnolence, Coma	Variabel ini merupakan variabel yang menjelaskan kondisi kesadaran pasien ketika melakukan kunjungan ke rumah sakit	(Yamamoto et al., 2015)
$X_8 = \text{Level SpO2}$		Variabel ini menjelaskan terkait tingkat atau kadar oksigen dari tubuh seseorang.	(Fadlilah et al., 2020)

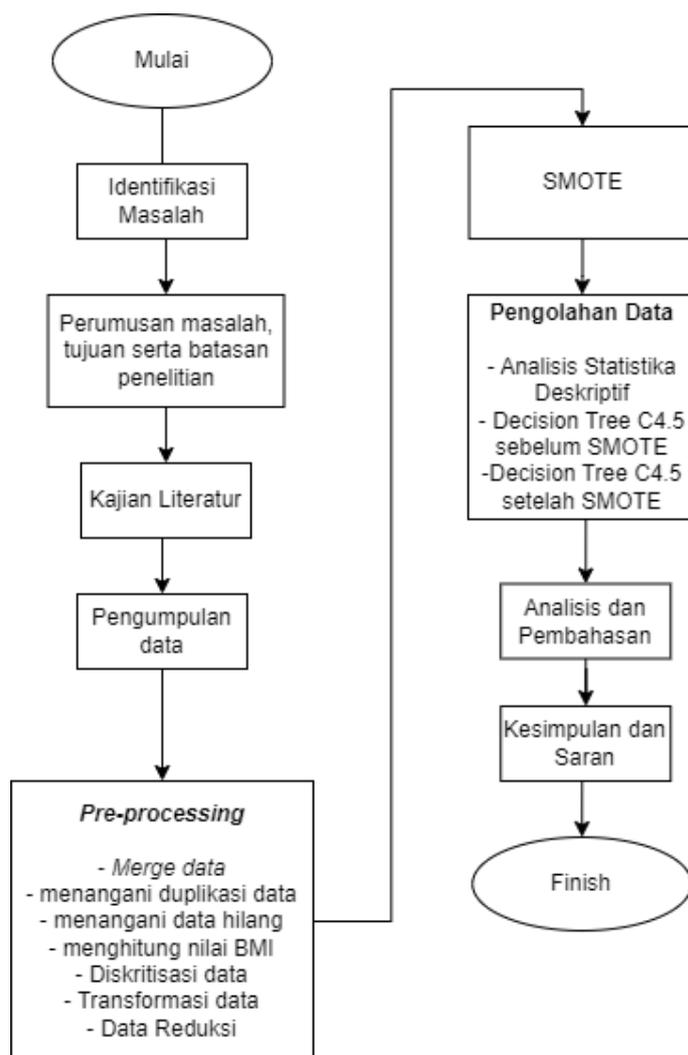
Pada penelitian yang dilakukan oleh Wang et al (2012), bertujuan untuk melakukan prediksi untuk menentukan apakah pasien layak mendapatkan tindakan medis berupa rawat inap atau dikategorikan pasien yang tidak dapat tertolong lagi (kematian). Dalam penelitian yang dilakukan, variabel yang digunakan untuk menilai kondisi pasien yang mendapat tindakan rawat inap yaitu berupa data demografis sosial pasien dan tanda-tanda vital tubuh manusia antara lain, usia, jenis kelamin, tingkat nadi, tingkat respirasi, tekanan darah sistolik dan diastolic yang digunakan untuk mengetahui tingkat hipertensi seseorang dan indeks masa tubuh. Hal yang sama juga dikemukakan oleh penelitian yang dilakukan Auble et al (2005).

Penelitian yang dilakukan bertujuan untuk menilai kondisi pasien dengan penyakit jantung yang diberi perawatan rawat inap untuk mengetahui tindakan medis yang tepat dalam menangani kondisi pasien dengan penyakit jantung pada tingkat risiko rendah dan komplikasi serius dengan menggunakan variabel prediktor usia dan jenis kelamin (data demografis) dan juga uji fisik berupa sistolik dan diastolik, nadi, respirasi dan suhu tubuh. Pada penelitian yang dilakukan oleh Yamamoto et al (2015), diketahui bahwa *glasgow coma scale* merupakan salah satu indikator penting untuk mengetahui kondisi *vital sign* dari seorang pasien. *Glasgow coma scale* (GCS) merupakan satuan yang mengukur tingkat kesadaran pasien. Selain GCS, *vital sign* yang menjadi prediktor yaitu, tekanan darah, level nadi, suhu tubuh, dan level respirasi pasien.

Penggunaan *vital sign* atau tanda-tanda vital tubuh dinilai tepat untuk digunakan sebagai variabel prediktor karena tanda-tanda vital tubuh manusia merupakan gambaran secara umum dan menyeluruh kondisi manusia saat dilakukannya assesmen atau pemeriksaan. Bagi dokter, pemeriksaan tanda-tanda vital merupakan gerbang yang paling pertama untuk mengenali indikasi pasien yang layak untuk dilakukan rawat inap ataupun rawat jalan. Ketika seseorang mengalami gangguan pada tanda-tanda vital, dibutuhkan pemeriksaan lebih lanjut untuk mengetahui kondisi pasien secara lebih rinci dan mendalam.

3.4 Alur Penelitian

Berikut ini merupakan alur penelitian yang dilakukan dengan dilampirkan ke dalam bentuk *flowchart*.



Gambar 3. 1 Alur Penelitian

Berikut ini merupakan penjelasan mengenai alur penelitian yang dilakukan.

1. Mulai

Tahapan ini merupakan tahapan paling awal dan menginisiasi seluruh tahapan yang akan dilakukan pada penelitian ini,

2. Identifikasi masalah

Pada tahapan identifikasi masalah, peneliti mencoba untuk mengetahui permasalahan yang tengah dialami oleh rumah sakit As-Syifa. Terdapat beberapa cara untuk mengidentifikasi masalah yaitu dapat dengan melakukan wawancara dan diskusi secara langsung bersama para karyawan yang bekerja di rumah sakit As-Syifa. Berdasarkan wawancara yang dilakukan, diketahui bahwa rumah sakit As-Syifa masih belum mengolah data yang bersumber dari pasien secara lebih lanjut, sehingga data yang telah diperoleh, hanya menjadi tumpukan sampah. Selain itu, perawat juga membutuhkan sistem yang dapat

merekomendasikan *plan* perawatan yang akan diberikan kepada pasien yang dapat digunakan untuk membantu kinerja dokter.

3. Perumusan masalah, tujuan serta batasan penelitian

Tahapan ini dilakukan dengan tujuan agar penelitian memiliki masalah yang ingin diselesaikan dan batasan penelitian yang jelas dan berfokus sehingga dapat menyelesaikan tujuan penelitian yang ingin dicapai

4. Kajian literatur

Kajian literatur terdiri atas dua sub-bab yaitu kajian literatur dan landasan teori. Pada kajian literatur, peneliti mengkaji penelitian-penelitian terdahulu yang memiliki kaitan dengan penelitian yang akan dilakukan. Sedangkan landasan teori berisi teori-teori yang akan dijadikan acuan awal dalam melakukan penelitian.

5. Pengumpulan data

Tahapan ini, peneliti melakukan pengumpulan data yang dibutuhkan guna mencapai tujuan penelitian dan mengatasi permasalahan yang terjadi. Penelitian ini menggunakan data hasil *record* pada *database* rumah sakit terkait pemeriksaan awal pasien. Data hasil *database* rumah sakit akan diubah (*convert*) terlebih dahulu yang awalnya berformat *Jrprints* menjadi *csv file* menggunakan *software JasperReports*.

6. *Pre-processing*

Pre-processing merupakan suatu tahapan untuk mempersiapkan data agar terhindar dari data yang terduplikat dan data yang hilang (*missing*). Selain itu, pada tahapan ini dilakukan *feature selection* yang mana proses ini dilakukan dengan menggunakan metode *chi square* uji independensi. Proses ini dilakukan untuk mengeliminasi variabel independen yang secara ilmu statistik tidak memiliki pengaruh yang signifikan terhadap variabel dependen atau variabel yang akan diprediksi. Dalam penelitian ini, proses *pre-processing* yang dilakukan yaitu sebagai berikut.

1) Menggabungkan data rawat jalan dan rawat inap

Data yang diberikan oleh pihak rumah sakit terbagi menjadi 2 bagian, yaitu data rawat jalan dan rawat inap. Data rawat jalan dan rawat inap masing-masing memiliki data yang terpisah yaitu data identitas pasien dan hasil pemeriksaan. Data identitas pasien meliputi ID, jenis kelamin, usia dan kode penyakit. Lalu untuk data hasil pemeriksaan meliputi data rekam medis tanda-tanda vital tubuh pasien (*vital sign*). Kemudian data rawat jalan dan rawat inap akan digabung dengan mengacu kepada ID para pasien sehingga diperoleh data yang akurat hanya untuk data pasien pengidap penyakit *hypertensive heart disease*.

2) Menangani duplikasi data

Langkah selanjutnya yaitu menangani duplikasi data yang terjadi dalam *dataset*. Menghapus duplikasi data dilakukan dengan cara mengacu kepada “ID” dan “Tanggal Rawat” pasien. Sehingga ketika ID dan Tanggal Rawat Pasien sama, otomatis akan dihapus. Penggunaan tanggal rawat sebagai acuan dalam menghapus duplikasi data agar proses yang dilakukan tidak menghapus data rekam medis pasien yang melakukan kunjungan ulang ke rumah sakit.

3) Menangani data yang hilang

Setelah data duplikasi dalam *dataset* telah ditangani, selanjutnya menangani data yang tidak lengkap dari masing-masing kolom yaitu dengan cara mengisi kekosongan data yang ada di dalam kolom dengan nilai rata-rata dari kolom tersebut.

4) Menghitung nilai BMI

Perhitungan nilai BMI dilakukan dengan memanfaatkan data tinggi dan berat badan pasien.

5) Diskritisasi data

Pada proses ini, dilakukan penggolongan atau pengkategorian data yang awalnya bertipe numerik menjadi ke dalam bentuk interval. Pengkategorian data yang dilakukan mengacu kepada referensi penelitian yang telah dilakukan sebelumnya. Berikut ini merupakan diskritisasi data dari masing-masing atribut.

6) Transformasi

Proses ini dilakukan untuk mengonversi data yang awalnya berbentuk numerik menjadi data dengan tipe kategori. Transformasi dilakukan dengan mengacu kepada pembagian kelas yang dilakukan pada tahapan diskritisasi data.

7) Data Reduksi

Dalam tahapan ini akan dilakukan seleksi terhadap variabel independen yang akan digunakan untuk memprediksi variabel dependen. Tahapan ini memanfaatkan uji *chi square* independensi untuk melihat pengaruh masing-masing variabel independen terhadap variabel dependen. Variabel independen yang tidak memiliki pengaruh akan dieliminasi.

7. SMOTE

Tahapan ini merupakan tahapan yang dilakukan untuk menangani ketimpangan data yang terjadi. Dalam penelitian ini ketimpangan data yang terjadi yaitu sebesar 99% data dengan

label rawat jalan dan 1% data dengan label rawat inap. Total data yang digunakan dalam penelitian ini adalah sebanyak 837 data dengan kondisi sebelum diterapkannya SMOTE. Setelah SMOTE diaplikasikan, jumlah data menjadi 1654. Prinsip kerja dari metode SMOTE ini adalah, memperbanyak data kelas minoritas yang dalam penelitian ini adalah data rawat inap agar setara dengan jumlah kelas data mayor (rawat jalan) dengan cara membuat data sintesis atau data buatan.

8. Pengolahan data

Setelah data berhasil dipersiapkan dengan melewati *pre-processing* dan penanganan ketimpangan data, peneliti melakukan pengolahan awal dengan menerapkan statistika deskriptif untuk mengetahui gambaran secara umum terkait data yang dimiliki. Setelah dilakukan analisis deskriptif, peneliti melanjutkan pengolahan dengan menggunakan algoritma *decision tree* C4.5 untuk melakukan prediksi tindakan medis yang layak untuk diterima oleh pasien. Berikut ini merupakan tahapan pengolahan data dengan menggunakan *decision tree* C4.5.

- 1) Melakukan pembagian data dengan menggunakan metode *cross validation*.
- 2) Menghitung nilai *entropy instance*.
- 3) Menghitung nilai *entropy* atribut
- 4) Menghitung nilai *gain*.
- 5) Menghitung nilai *split information*.
- 6) Menghitung nilai *gain ratio*
- 7) Melakukan pengujian model dengan menggunakan data uji (*testing*).
- 8) Membuat pohon keputusan.
- 9) Evaluasi hasil klasifikasi dengan menggunakan *confusion matrix*.
- 10) Menginterpretasi aturan/rules dari *decision tree* yang telah terbentuk.

9. Analisis dan Pembahasan

Setelah mendapatkan hasil dari pengolahan data lanjutan yang dilakukan, peneliti akan membahas terkait gambaran data secara umum melalui analisis deskriptif dan juga mengetahui hasil model yang terbentuk melalui algoritma *decision tree* C4.5 sehingga dapat diketahui tindakan medis yang tepat untuk didapatkan oleh pasien. Selain itu dilakukan perbandingan performa algoritma *decision tree* C4.5 sebelum dan sesudah menggunakan SMOTE.

10. Kesimpulan dan Saran

Pada tahapan ini, peneliti akan menyimpulkan hasil penelitian yang telah diperoleh. Kesimpulan akan menjawab rumusan masalah dan tujuan penelitian yang telah dirumuskan di awal penelitian. Hasil penelitian ini berupa model prediksi untuk menetapkan tindakan medis untuk pasien, sekaligus mengetahui performa model yang dihasilkan ketika di kombinasikan dengan SMOTE.

11. Selesai

Setelah melewati seluruh proses yang ada, penelitian akan dianggap selesai.

BAB IV

PENGUMPULAN DAN PENGOLAHAN DATA

4.1 Pengumpulan Data

pengumpulan data dilakukan bertujuan untuk memenuhi semua kebutuhan pada penelitian yang akan dilakukan. Dalam penelitian ini, data yang dibutuhkan merupakan data historis hasil pemeriksaan pasien hipertensi dengan tindakan medis rawat jalan dan rawat inap. Data yang digunakan secara menyeluruh berjumlah sebanyak 491 baris dengan kolom sebanyak 10 baris.

4.2 *Pre-Processing*

Berikut ini merupakan tahapan *pre-processing* yang dilakukan dalam penelitian ini, agar data siap untuk dilakukan pengolahan. Proses pengerjaan tahapan *pre-processing* dalam penelitian ini, dilakukan sepenuhnya dengan menggunakan Bahasa pemrograman *python*, *Microsoft Excel* dan SPSS.

4.2.1 *Merge data*

Tahapan ini dilakukan untuk menggabungkan data antara data identitas pasien dengan data hasil pemeriksaan yang pasien. Proses penggabungan data mengacu kepada *unique value* “ID” agar data yang diambil sesuai dengan yang diinginkan yaitu data pasien hipertensi.

1. Data identitas pasien

Berikut ini merupakan data identitas pasien sebelum digabungkan dengan data hasil pemeriksaan.

Tabel 4. 2 Data ID pasien

ID	Jenis Kelamin	Kode	Umur
6503	Laki-Laki	I11	54
75769	Perempuan	I11	50
75575	Laki-Laki	I11	46
74409	Perempuan	I11	80
74099	Perempuan	I11	68
12242	Laki-Laki	I11	90
19513	Perempuan	I11	66

2. Data hasil pemeriksaan pasien

Berikut merupakan data hasil pemeriksaan pasien hipertensi yang *di-record* dalam system manajemen rumah sakit.

Tabel 4. 3 Data hasil pemeriksaan

ID	Suhu(C)	Sistolik	Diastolik	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat
6503	37.7	145	92	80	20			90	Compos Mentis	Rawat Inap
75769	36.5	170	100	82	20			80	Compos Mentis	Rawat Inap
75575	36.6	190	100	58	20			99	Compos Mentis	Rawat Inap
74409	37.1	200	94	97	20			82	Compos Mentis	Rawat Inap

3. Merge data

Selanjutnya data ID akan digabungkan data hasil pemeriksaan pasien dengan acuan no ID yang dimiliki tiap pasien. Berikut ini merupakan hasil penggabungan data yang dilakukan.

Tabel 4. 4 Hasil merge data

ID	Jenis Kelamin	Kode	Umur	Suhu (C)	Sistolik	Diastolik	Nadi (/mnt)	Respirasi (/mnt)	Tinggi (m)	Berat (Kg)	SpO2	Kesadaran	Jenis Rawat
6503	Laki-Laki	I11	54	37.7	145	92	80	20			90	Compos Mentis	Rawat Inap
75769	Perempuan	I11	50	36.5	170	100	82	20			80	Compos Mentis	Rawat Inap
75575	Laki-Laki	I11	46	36.6	190	100	58	20			99	Compos Mentis	Rawat Inap
74409	Perempuan	I11	80	37.1	200	94	97	20			82	Compos Mentis	Rawat Inap

4.2.2 Data setelah penanganan *missing value*

Berikut ini merupakan data rekam medis sebelum dilakukannya proses *cleaning* dan penanganan data hilang. Pada proses ini, dilakukan penghapusan kolom ID dan Kode. Setelah itu, proses dilanjutkan dengan mengisi kolom yang memiliki data hilang dengan cara mengisi menggunakan nilai rata-rata dari masing-masing kolom. Menurut Rafsunjani et al (2019) salah satu cara untuk mengisi nilai yang hilang pada dataset dapat dengan menggunakan nilai rata-rata. Cara ini dikenal sebagai cara yang sederhana namun dikenal *powerful*. Selain itu juga dilakukan perhitungan skor indeks masa tubuh dengan menggunakan data tinggi dan berat badan pasien. Berikut ini merupakan contoh hasil setelah dilakukannya proses penanganan data hilang.

Tabel 4. 5 Data setelah proses penanganan *missing value*

Jenis Kelamin	Umur	Suhu(C)	Sistolik	Diastolik	Nadi (/mnt)	Respirasi (/mnt)	Tinggi (m)	Berat (Kg)	SpO2	BMI	Kesadaran	Jenis Rawat
Laki-Laki	54	37.7	145	92	80	20	1.59	66.38	90	26.12	Compos Mentis	Rawat Inap
Perempuan	50	36.5	170	100	82	20	1.59	66.38	80	26.12	Compos Mentis	Rawat Inap
Laki-Laki	46	36.6	190	100	58	20	1.59	66.38	99	26.12	Compos Mentis	Rawat Inap
Perempuan	80	37.1	200	94	97	20	1.59	66.38	82	26.12	Compos Mentis	Rawat Inap
Perempuan	68	38	148	96	80	20	1.59	66.38	99	26.12	Compos Mentis	Rawat Inap
Laki-Laki	90	36.5	189	84	80	22	1.5	60	76	26.66	Compos Mentis	Rawat Inap
Perempuan	66	36.5	200	110	55	21	1.5	48	98.28	21.33	Compos Mentis	Rawat Inap
Laki-Laki	59	36	160	90	90	18	1.59	66.38	98.28	26.12	Compos Mentis	Rawat Jalan

4.2.3 Data Diskritisasi

Penanganan data yang hilang dilakukan dengan cara mengganti data yang hilang dengan rata-rata masing-masing kolom yang mengalami kehilangan data (*missing values*). Berikut ini merupakan contoh data yang telah mengalami proses

1. Diskritisasi data hipertensi

Berikut ini merupakan diskritisasi data tekanan darah

Tabel 3. 1 Diskritisasi Data Hipertensi

Sistolik	Distolik	Tingkat Hipertensi
< 120 mmHg	< 80 mmHg	Normal
120 – 129 mmHg	< 80 mmHg	Pra-Hipertensi
130 - 139 mmHg	80-89 mm Hg	Stage 1
> 140 mmHg	> 90 mmHg	Stage 2

Sumber: (Tackling & Borhade, 2021)

2. Diskritisasi data denyut nadi

Berikut ini merupakan diskritisasi data denyut nadi

Tabel 3. 2 Diskritisasi data denyut nadi

Umur	Bradikardia	Normal	Takikardia
<1 tahun	< 100	100-160	> 160
1-10	70	70-120	> 120
11-17	60	60-100	>100
>17	60	60-100	>100

Sumber: (Richa Rachmawati et al., 2022)

3. Diskritisasi data suhu tubuh

Berikut ini merupakan diskritisasi data suhu tubuh manusia

Tabel 3. 3 diskritisasi data suhu tubuh

Kategori	Suhu	
	Anak-anak	Dewasa
Hipotermia	< 36	< 36
Normal	36.1 – 37.5	36.1 - 37.7
Demam	37.6 - 40	37.8 – 40
Hiperpireksia	> 40	> 40

Sumber: (Yuniarti et al., 2020)

4. Saturasi oksigen (SpO2)

Berikut ini diskritisasi dari data SpO2

Tabel 3. 4 Diskritisasi data SpO2

SpO2	Tingkat Saturasi Oksigen
95-100	Normal
91-94	cukup

86-90 sedang
 <85 rendah
 Sumber: (Fadlilah et al., 2020).

5. Respirasi

Berikut ini merupakan diskritisasi data respirasi manusia.

Tabel 3. 5 Diskritisasi data respirasi

Respirasi	Tingkat Saturasi Oksigen
12-20	Normal

Sumber: (Nejadgholi et al., 2019)

6. Glasgow coma scale

Berikut ini merupakan diskritisasi data GCS

Tabel 2. 11 Kategori nilai GCS

Nilai GCS	Kategori
15	Compos Mentis
12-14	Somnolen
8-11	Sopor
3-7	Koma

Sumber: (Hartanto et al., 2019)

7. Usia

Berikut ini merupakan diskritisasi data usia pasien

Tabel 2. 12 Diskritisasi data usia

Usia	Kategori
< 40 Tahun	Muda
40 – 59 Tahun	Sedang
> 60 tahun	Tua

Sumber: (Lesmana, 2012)

8. Indeks masa tubuh

Berikut ini merupakan diskritisasi data indeks masa tubuh pasien.

Tabel 2. 13 Indeks masa tubuh

Indeks masa tubuh	Kategori
< 18.5	Underweight
18.5 – 25	Healthy weight
25 – 30	Overweight
>30	Obesity

Sumber: (Truesdale & Stevens, 2008)

4.2.4 Transformasi data

Berikut ini merupakan hasil dari transformasi data yang dilakukan. Transformasi data bertujuan untuk mengubah tipe data yang awalnya bertipe numerik menjadi tipe kategorikal.

Tabel 4. 6 Transformasi data

Jenis Kelamin	Umur	Level BMI	Tingkat Hipertensi	Level Nadi	Level Suhu	Level SpO2	Level respirasi	Kesadaran	Jenis Rawat
Laki-Laki	54	Overweight	Stage 2	Normal	Normal	Sedang	Normal	Compos Mentis	Rawat Inap
Perempuan	50	Overweight	Stage 2	Normal	Normal	Rendah	Normal	Compos Mentis	Rawat Inap
Laki-Laki	46	Overweight	Stage 2	Bradikardia	Normal	Normal	Normal	Compos Mentis	Rawat Inap
Perempuan	80	Overweight	Stage 2	Normal	Normal	Rendah	Normal	Compos Mentis	Rawat Inap
Perempuan	68	Overweight	Stage 2	Normal	Demam	Normal	Normal	Compos Mentis	Rawat Inap

4.2.5 Data reduksi

Data reduksi dilakukan untuk melakukan seleksi terhadap variabel independen yang digunakan untuk melakukan prediksi. Data reduksi yang dilakukan memanfaatkan uji statistik yaitu *chi square* uji independensi yang digunakan untuk mengetahui pengaruh masing-masing variabel independen terhadap variabel dependen secara dengan pendalaman ilmu statistik. Variabel independen yang tidak memiliki pengaruh pada variabel dependen akan dieliminasi. Di bawah ini merupakan uji *chi square independensi* yang dilakukan untuk mengetahui pengaruh antara masing-masing variabel independen dengan variabel dependen. Proses data reduksi dalam penelitian ini menggunakan bantuan *software* SPSS.

a. Uji Hipotesis

- H_0 : Tidak terdapat pengaruh antara jenis kelamin terhadap *plan* perawatan.
 H_1 : Terdapat pengaruh antara jenis kelamin terhadap *plan* perawatan.
- H_0 : Tidak terdapat pengaruh antara umur terhadap *plan* perawatan.
 H_1 : Terdapat pengaruh antara umur terhadap *plan* perawatan.
- H_0 : Tidak terdapat pengaruh antara level BMI terhadap *plan* perawatan.
 H_1 : Terdapat pengaruh antara level BMI terhadap *plan* perawatan.
- H_0 : Tidak terdapat pengaruh antara tingkat hipertensi terhadap *plan* perawatan.
 H_1 : Terdapat pengaruh antara tingkat hipertensi terhadap *plan* perawatan.
- H_0 : Tidak terdapat pengaruh antara level nadi terhadap *plan* perawatan.
 H_1 : Terdapat pengaruh antara level nadi terhadap *plan* perawatan.
- H_0 : Tidak terdapat pengaruh antara level suhu terhadap *plan* perawatan.
 H_1 : Terdapat pengaruh antara level suhu terhadap *plan* perawatan.
- H_0 : Tidak terdapat pengaruh antara level SpO2 terhadap *plan* perawatan.
 H_1 : Terdapat pengaruh antara level SpO2 terhadap *plan* perawatan.
- H_0 : Tidak terdapat pengaruh antara level respirasi terhadap *plan* perawatan.
 H_1 : Terdapat pengaruh antara level respirasi terhadap *plan* perawatan.
- H_0 : Tidak terdapat pengaruh antara kesadaran terhadap *plan* perawatan.
 H_1 : Terdapat pengaruh antara kesadaran terhadap *plan* perawatan.

b. Tingkat signifikansi

Tingkat Keyakinan 95%, maka nilai $\alpha = 0,05$

c. Kriteria pengujian

Dengan membandingkan angka probabilitas:

- Nilai signifikansi $> 0,05$, maka H_0 diterima
- Nilai signifikansi $\leq 0,05$, maka H_0 ditolak

d. Harga statistik penguji

Berikut merupakan hasil statistik uji untuk variabel independen yang digunakan dalam penelitian ini.

Tabel 4. 7 Hasil statistik uji variabel independen

Variabel Independen	Nilai signifikansi
Jenis kelamin	0.178
Umur	0.000
Level BMI	0.128
Tingkat Hipertensi	0.000
Level Nadi	0.000
Level Suhu	0.052
Level SpO2	0.000
Level Respirasi	0.572
Kesadaran	0.000

e. Pengambilan keputusan

Berikut ini merupakan keputusan dari dilakukannya uji *chi-square* independensi.

Tabel 4. 8 Hasil keputusan uji independensi terhadap variabel independen

Variabel Independen	Nilai signifikansi		Alpha (α)	Keputusan
Jenis kelamin	0.178	>		H_0 diterima
Umur	0.003	<		H_0 ditolak
Level BMI	0.128	>		H_0 diterima
Tingkat Hipertensi	0.000	<		H_0 ditolak
Level Nadi	0.000	<	0.05	H_0 ditolak
Level Suhu	0.052	>		H_0 diterima
Level SpO2	0.000	<		H_0 ditolak
Level Respirasi	0.572	>		H_0 diterima
Kesadaran	0.000	<		H_0 ditolak

f. Kesimpulan

Berikut ini merupakan kesimpulan dari hasil uji independensi yang telah diperoleh

Tabel 4. 9 Kesimpulan uji independensi

Variabel Independen	Kesimpulan
---------------------	------------

Jenis kelamin	Tidak terdapat pengaruh antara jenis kelamin terhadap <i>plan</i> perawatan.
Umur	Terdapat pengaruh antara umur terhadap <i>plan</i> perawatan.
Level BMI	Tidak terdapat pengaruh antara level BMI terhadap <i>plan</i> perawatan.
Tingkat Hipertensi	Terdapat pengaruh antara tingkat hipertensi terhadap <i>plan</i> perawatan.
Level Nadi	Terdapat pengaruh antara level nadi terhadap <i>plan</i> perawatan.
Level Suhu	Tidak terdapat pengaruh antara level suhu terhadap <i>plan</i> perawatan.
Level SpO2	Terdapat pengaruh antara level SpO2 terhadap <i>plan</i> perawatan.
Level Respirasi	Tidak terdapat pengaruh antara level respirasi terhadap <i>plan</i> perawatan.
Kesadaran	Terdapat pengaruh antara kesadaran terhadap <i>plan</i> perawatan.

Berdasarkan uji independensi yang dilakukan, dapat diketahui bahwa terdapat 4 variabel independen yang akan dihapus, yaitu jenis kelamin, level BMI, level suhu dan level respirasi. Namun ditinjau dari keterangan dokter, terdapat beberapa hasil uji statistic yang tidak sesuai. Menurut dokter, tanda-tanda vital tubuh manusia yang mempengaruhi tingkat keparahan hipertensi yaitu jenis kelamin, umur, level BMI, tingkat hipertensi dan kesadaran. Untuk level nadi, suhu, level SpO2 dan level respirasi adalah tanda-tanda vital yang dipengaruhi oleh penyakit hipertensi itu sendiri. Peneliti mengkombinasikan variable yang akan digunakan berdasarkan uji statistic yang dilakukan dan keterangan dokter. Sehingga, variable yang digunakan dalam penelitian ini adalah jenis kelamin, usia, level BMI, tingkat hipertensi, level nadi, level SpO2 dan kesadaran.

4.2.6 Data akhir

Berikut ini merupakan pratinjau 5 data teratas yang telah melewati proses *pre-processing* dan siap untuk dilakukan pemodelan dengan algoritma *decision tree* C4.5. Data yang digunakan berjumlah sebanyak 837 baris dengan menggunakan 7 variabel.

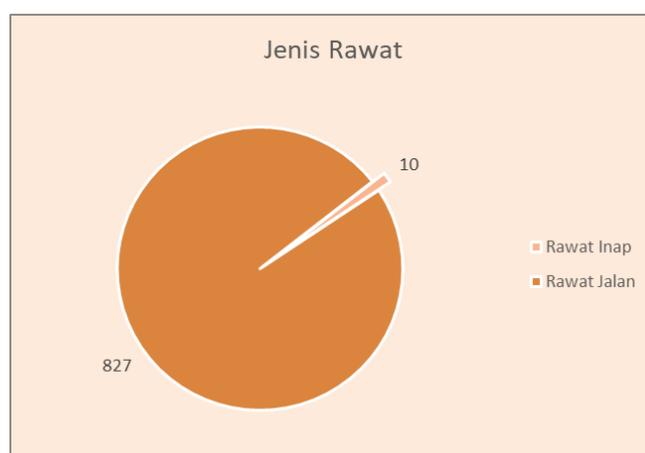
Tabel 4. 10 Data akhir

Jenis Kelamin	Usia	Level BMI	Tingkat Hipertensi	Level Nadi	Level SpO2	Kesadaran	Jenis Rawat
Laki-Laki	Medium	Overweight	Stage 2	Normal	Sedang	Compos Mentis	Rawat Inap
Perempuan	Medium	Overweight	Stage 2	Normal	Rendah	Compos Mentis	Rawat Inap
Laki-Laki	Medium	Overweight	Stage 2	Bradikardia	Normal	Compos Mentis	Rawat Inap

Jenis Kelamin	Usia	Level BMI	Tingkat Hipertensi	Level Nadi	Level SpO2	Kesadaran	Jenis Rawat
Perempuan	Tua	Overweight	Stage 2	Normal	Rendah	Compos Mentis	Rawat Inap
Perempuan	Muda	Healthy Weight	Stage 2	Normal	Normal	Compos Mentis	Rawat Inap

4.3 Analisis Deskriptif

Dalam penelitian yang dilakukan, terdapat 5 variabel independen dan 1 variabel dependen. Untuk variabel independen sendiri terdiri dari umur, tingkat hipertensi, level nadi, level SpO2 dan kesadaran. Sedangkan variabel dependen terdiri dari jenis rawat. Berikut merupakan analisis deskriptif untuk seluruh variabel yang terlibat dalam penelitian ini.



Gambar 4. 1 Distribusi jenis rawat pasien

Berdasarkan gambar 4.1, diketahui bahwa total seluruh data yang digunakan dalam penelitian ini adalah 837 pasien dengan sebanyak 827 pasien menjalani rawat jalan dan 10 diantaranya menjalani rawat inap.

Tabel 4. 11 Tabel kontingensi jenis kelamin dan jenis rawat

		Kontingensi		
		Jenis_Rawat		Total
Kelamin	Laki-Laki	Rawat Inap	Rawat Jalan	
				6

Perempuan	4	504	508
Total	10	827	837

Pada tabel 4.11 menunjukkan tabel kontingensi antara variabel jenis kelamin dan jenis rawat. Tercatat, bahwa pasien dengan jenis kelamin perempuan merupakan pasien *hypertensive heart disease* dengan jumlah tertinggi yaitu sebanyak 508 orang dengan 4 diantaranya harus menjalani rawat inap. Untuk pasien dengan jenis kelamin laki-laki, tercatat sebanyak 329 pasien dengan 6 orang diantaranya harus menjalani rawat inap.

Tabel 4. 12 Tabel kontingensi tingkat hipertensi dan jenis rawat

		Crosstab		
		Jenis_Rawat		Total
Tingkat_hipertensi		Rawat Inap	Rawat Jalan	
		Normal	0	69
	Pra-Hipertensi	2	184	186
	Stage 1	1	453	454
	Stage 2	7	121	128
	Total	10	827	837

Berdasarkan tabel 4.12, terlihat bahwa pasien yang berkunjung paling banyak ke rumah sakit adalah pasien dengan tingkat hipertensi *stage 1* yaitu sebanyak 454 pasien dengan 1 orang diantaranya harus menjalani rawat inap. Dikuti dengan tingkat hipertensi pra-hipertensi sebanyak 186 orang dengan 2 orang diantaranya menjalani rawat inap. Untuk tingkat hipertensi *stage 2* tercatat sebanyak 128 pasien melakukan kunjungan dengan 7 diantaranya menjalani rawat inap. Untuk pasien dengan tingkat hipertensi normal tercatat sebanyak 69 pasien dan seluruhnya menjalani perawatan rawat jalan.

Tabel 4. 13 Tabel kontingensi variabel level nadi dengan jenis rawat

		Kontingensi		
		Jenis_Rawat		Total
Level_nadi		Rawat Inap	Rawat Jalan	
		Bradikardia	2	1
	Normal	8	810	818

Takikardia	0	16	16
Total	10	827	837

Dengan melihat tabel 4.13, diketahui bahwa pasien yang berkunjung paling banyak adalah dengan level nadi normal sebanyak 818 pasien dengan 8 diantaranya harus menjalani rawat inap. Sebanyak 3 pasien memiliki level nadi bradikardia dengan 2 pasien rawat inap dan 1 pasien rawat jalan. Kemudian 16 pasien tercatat memiliki level nadi takikardia dengan jumlah 16 pasien dan seluruhnya menjalani rawat jalan.

Tabel 4. 14 Tabel kontingensi level SpO2 dan Jenis rawat

		Kontingensi		
		Jenis_Rawat		Total
		Rawat Inap	Rawat Jalan	
Level_SpO2	Cukup	0	1	1
	Normal	6	780	786
	Rendah	3	18	21
	Sedang	1	28	29
	Total	10	827	837

Dengan melihat tabel 4.14, pasien HHD yang berkunjung ke rumah sakit paling banyak memiliki level SpO2 yang normal dengan 786 pasien, yang 6 diantaranya harus menjalani rawat inap. Diikuti dengan level SpO2 sedang sebanyak 29 pasien dengan 1 pasien rawat inap, level SpO2 rendah sebanyak 21 pasien dengan rawat inap sebanyak 3 pasien dan level terakhir yaitu cukup dengan 1 pasien rawat jalan.

Tabel 4. 15 Tabel kontingensi usia dan jenis rawat

		Kontingensi		
		Jenis_Rawat		Total
		Rawat Inap	Rawat Jalan	
Usia	Medium	4	363	367
	Muda	3	43	46
	Tua	3	421	424
	Total	10	10	837

Berdasarkan tabel kontingensi 4.15 yaitu mengenai kontingensi variabel usia dan jenis rawat, diketahui bahwa pasien terbanyak dikategorikan memiliki usia tua dengan jumlah 424 pasien dengan 3 diantaranya rawat inap. Diikuti oleh pasien usia medium dengan jumlah 367 dengan 4 diantaranya rawat inap dan yang terakhir pasien usia muda berjumlah 46 pasien dengan 3 diantaranya menjalani rawat inap

Tabel 4. 16 tabel kontingensi kesadaran dan jenis rawat

		Kontingensi		
		Jenis_Rawat		Total
		Rawat Inap	Rawat Jalan	
Kesadaran	Compos Mantis	10	827	837
Total		10	827	837

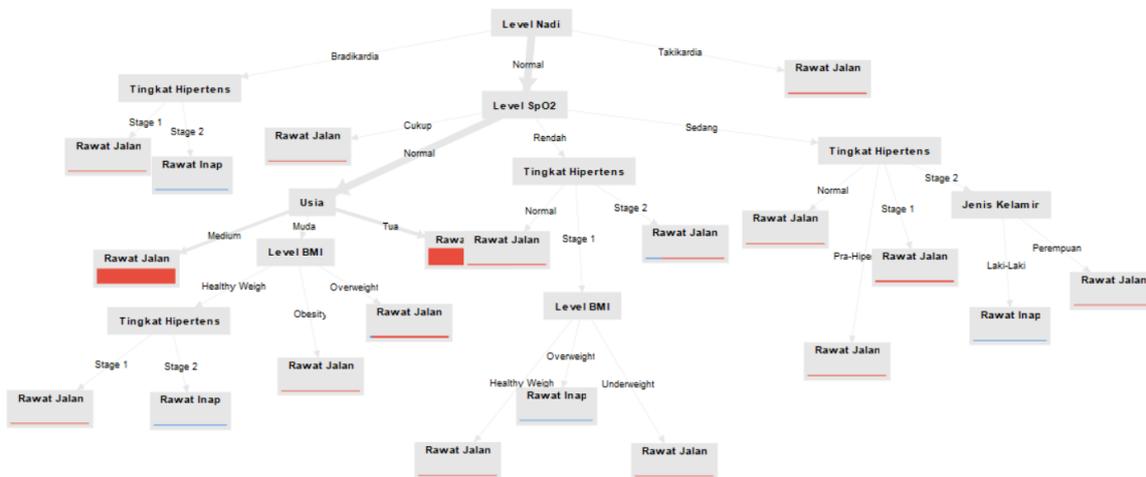
Berdasarkan tabel 4.16, diketahui bahwa seluruh pasien yang berkunjung yaitu sebanyak 837 pasien memiliki tingkat kesadaran *compos mantis* yang 10 diantaranya harus menjalani rawat inap.

4.4 Pengolahan Data

Pada penelitian ini, pengolahan data yang dilakukan sepenuhnya menggunakan bantuan *software* Rapidminer. Berikut ini merupakan pengolahan data yang dilakukan dalam penelitian ini.

4.4.1 *Decision tree* C4.5 sebelum SMOTE

Berikut ini merupakan model yang terbentuk dengan menggunakan algoritma *decision tree* C4.5 sebelum integrasi dengan metode SMOTE.



Gambar 4. 2 *Decision tree* sebelum SMOTE

Pembentukan *decision tree* diawali dengan melakukan perhitungan nilai entropi untuk mengetahui tingkat keseragaman data yang kemudian dilanjutkan dengan perhitungan nilai *gain ratio*. Variabel yang memiliki nilai *gain ratio* tertinggi akan menjadi *root node* dalam *decision tree*. Dalam penelitian ini, *decision tree* sebelum diterapkannya SMOTE, yang bertindak sebagai *root node* adalah variabel Level nadi. Kemudian variabel lainnya akan mengikuti sebagai *internal node* lainnya. Jumlah seluruh aturan atau *rules* yang terbentuk adalah sebanyak 20 *rules*. Berikut ini merupakan *rules* yang terbentuk.

1. *IF* Level nadi Bradikardia *AND* Tingkat Hipertensi Stage 1 *THEN* Rawat Jalan dengan perbandingan (1:0).
2. *IF* Level nadi Bradikardia *AND* Tingkat Hipertensi Stage 2 *THEN* Rawat Inap dengan perbandingan (2:0).
3. *IF* Level nadi Normal *AND* Level SpO2 Cukup *THEN* Rawat Jalan dengan perbandingan (1:0).
4. *IF* Level nadi Normal *AND* Level SpO2 Normal *AND* Usia Medium *THEN* Rawat Jalan dengan perbandingan (342:1).
5. *IF* Level nadi Normal *AND* Level SpO2 Normal *AND* Usia Muda *AND* Level BMI Obesity *THEN* Rawat Jalan dengan perbandingan (2:0).
6. *IF* Level nadi Normal *AND* Level SpO2 Normal *AND* Usia Muda *AND* Level BMI Healthy weight *AND* Tingkat Hipertensi Stage 1 *THEN* Rawat Jalan dengan perbandingan (3:0).

7. *IF* Level nadi Normal *AND* Level SpO2 Normal *AND* Usia Muda *AND* Level BMI *Healthy weight AND* Tingkat Hipertensi *Stage 2 THEN* Rawat Inap dengan perbandingan (2:0).
8. *IF* Level nadi Normal *AND* Level SpO2 Normal *AND* Usia Muda *AND* Level BMI *Healthy weight THEN* Rawat Jalan dengan perbandingan (31:1).
9. *IF* Level nadi Normal *AND* Level SpO2 Normal *AND* Usia Tua *THEN* Rawat Jalan dengan perbandingan (387:0).
10. *IF* Level nadi Normal *AND* Level SpO2 Rendah *AND* Tingkat Hipertensi Normal *THEN* Rawat Jalan dengan perbandingan (6:0).
11. *IF* Level nadi Normal *AND* Level SpO2 Rendah *AND* Tingkat Hipertensi *Stage 1 AND* Level BMI *Healthy weight THEN* Rawat Jalan dengan perbandingan (1:0).
12. *IF* Level nadi Normal *AND* Level SpO2 Rendah *AND* Tingkat Hipertensi *Stage 1 AND* Level BMI *Overweight THEN* Rawat Inap dengan perbandingan (1:0).
13. *IF* Level nadi Normal *AND* Level SpO2 Rendah *AND* Tingkat Hipertensi *Stage 1 AND* Level BMI *Underweight THEN* Rawat Inap dengan perbandingan (1:0).
14. *IF* Level nadi Normal *AND* Level SpO2 Rendah *AND* Tingkat Hipertensi *Stage 1 THEN* Rawat Jalan dengan perbandingan (8:2).
15. *IF* Level nadi Normal *AND* Level SpO2 Sedang *AND* Tingkat Hipertensi Normal *THEN* Rawat Jalan dengan perbandingan (1:0).
16. *IF* Level nadi Normal *AND* Level SpO2 Sedang *AND* Tingkat Hipertensi *Pra-hipertensi THEN* Rawat Jalan dengan perbandingan (3:0).
17. *IF* Level nadi Normal *AND* Level SpO2 Sedang *AND* Tingkat Hipertensi *Stage 1 THEN* Rawat Jalan dengan perbandingan (23:0).
18. *IF* Level nadi Normal *AND* Level SpO2 Sedang *AND* Tingkat Hipertensi *Stage 2 AND* Jenis Kelamin Laki-laki *THEN* Rawat Inap dengan perbandingan (1:0).
19. *IF* Level nadi Normal *AND* Level SpO2 Sedang *AND* Tingkat Hipertensi *Stage 2 AND* Jenis Kelamin Perempuan *THEN* Rawat Jalan dengan perbandingan (1:0).
20. *IF* Level Nadi Takikardia *THEN* Rawat Jalan dengan perbandingan (16:0)

Tabel 4. 17 CM Sebelum SMOTE

	Aktual: Rawat Inap	Aktual: Rawat Jalan	class precision
pred. Rawat Inap	TP = 2	FP = 6	25%
pred. Rawat Jalan	FN = 8	TN = 821	99.03%
class recall	20%	99.27%	

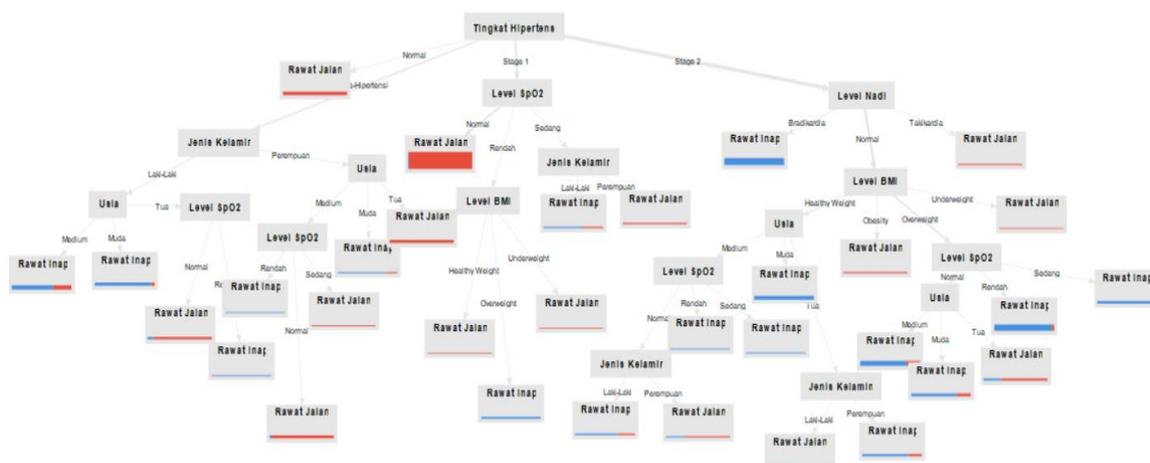
Berdasarkan hasil *confusion matrix* diperoleh bahwa:

1. Jumlah prediksi *plan perawatan* untuk rawat inap dan data realita yang sesuai atau benar pasien menjalani rawat inap adalah sebanyak 2. Kondisi yang sesuai antara hasil prediksi dengan data realita *plan* perawatan di lapangan disebut *true positive*.
2. Jumlah prediksi *plan* perawatan untuk jenis rawat inap menunjukkan ketidaksesuaian antara prediksi dan data realita di lapangan yang menunjukkan data aktual yaitu rawat jalan adalah sebanyak 6. Kondisi ini disebut sebagai *false positive*.
3. Jumlah prediksi *plan* perawatan untuk jenis rawat jalan namun tidak sesuai dengan data aktual yang menunjukkan rawat inap yaitu sebanyak 8. Kondisi ini disebut sebagai *false negative*.
4. Jumlah prediksi *plan* perawatan untuk jenis rawat jalan dan data realita di lapangan benar pasien menunjukkan rawat jalan, yaitu sebanyak 821. Kondisi ini disebut juga sebagai *true negative*.

Berikut ini merupakan perhitungan dari *confusion matrix*:

1. $Accuracy = \frac{TP+TN}{Total} = \frac{827}{837} = 98.33\% \approx 99\%$
2. $Misclassification Rate: \frac{FP+FN}{Total} = \frac{14}{837} = 0.01 = 1\%$
3. $Recall/true\ positive\ rate\ (rawat\ inap): \frac{TP}{TP+FN} = \frac{2}{10} = 20\%$
4. $Recall/true\ negative\ rate\ (rawat\ jalan): \frac{TN}{TN+FP} = \frac{821}{827} = 99.27\%$
5. $Precision\ rawat\ inap: \frac{TP}{TP+FP} = \frac{2}{8} = 25\%$
6. $Precision\ rawat\ jalan: \frac{TN}{TN+FN} = \frac{821}{821+8} = 99.03\%$
7. $Prevalance: \frac{TP+FN}{total} = \frac{2+8}{837} = 0.011 = 1.1\%$
8. $Area\ under\ curve\ (AUC) = 0.557$

4.4.2 Decision Tree C4.5 Setelah SMOTE



Gambar 4. 3 Decision tree setelah SMOTE

Pada model decision tree setelah penanganan ketimpangan data (*imbalanced data*), variabel yang memiliki nilai *gain ratio* tertinggi akan menjadi *root node* dalam *decision tree*. Dalam penelitian ini, *decision tree* setelah diterapkannya SMOTE, yang bertindak sebagai *root node* adalah variabel Tingkat hipertensi. Kemudian variabel lainnya akan mengikuti sebagai *internal node* lainnya. Jumlah seluruh aturan atau *rules* yang terbentuk adalah sebanyak 33 *rules*. Berikut ini merupakan *rules* yang terbentuk.

1. *IF* Tingkat Hipertensi Normal *THEN* Rawat Jalan dengan perbandingan (69:0).
2. *IF* Tingkat Hipertensi Pra-Hipertensi *AND* Jenis Kelamin Laki-laki *AND* Usia Medium *THEN* Rawat Inap dengan perbandingan (64:28).
3. *IF* Tingkat Hipertensi Pra-Hipertensi *AND* Jenis Kelamin Laki-laki *AND* Usia Muda *THEN* Rawat Inap dengan perbandingan (60:4).
4. *IF* Tingkat Hipertensi Pra-Hipertensi *AND* Jenis Kelamin Laki-laki *AND* Usia Tua *AND* Level SpO2 Normal *THEN* Rawat Inap dengan perbandingan (60:4).
5. *IF* Tingkat Hipertensi Pra-Hipertensi *AND* Jenis Kelamin Laki-laki *AND* Usia Tua *AND* Level SpO2 Rendah *THEN* Rawat Inap dengan perbandingan (4:0).
6. *IF* Tingkat Hipertensi Pra-Hipertensi *AND* Jenis Kelamin Perempuan *AND* Usia Medium *AND* Level SpO2 Rendah *THEN* Rawat Inap dengan perbandingan (1:0).
7. *IF* Tingkat Hipertensi Pra-Hipertensi *AND* Jenis Kelamin Perempuan *AND* Usia Medium *AND* Level SpO2 Normal *THEN* Rawat Inap dengan perbandingan (52:2).

8. *IF* Tingkat Hipertensi Pra-Hipertensi *AND* Jenis Kelamin Perempuan *AND* Usia Medium *AND* Level SpO2 Sedang *THEN* Rawat Jalan dengan perbandingan (2:0).
9. *IF* Tingkat Hipertensi Pra-Hipertensi *AND* Jenis Kelamin Perempuan *AND* Usia Muda *THEN* Rawat Inap dengan perbandingan (9:2).
10. *IF* Tingkat Hipertensi Pra-Hipertensi *AND* Jenis Kelamin Perempuan *AND* Usia Tua *THEN* Rawat Jalan dengan perbandingan (56:0).
11. *IF* Tingkat Hipertensi *Stage 1* *AND* Level SpO2 Normal *THEN* Rawat Jalan dengan perbandingan (428:5).
12. *IF* Tingkat Hipertensi *Stage 1* *AND* Level SpO2 Rendah *AND* Level BMI *Healthy weight* *THEN* Rawat Jalan dengan perbandingan (1:0).
13. *IF* Tingkat Hipertensi *Stage 1* *AND* Level SpO2 Rendah *AND* Level BMI *Overweight* *THEN* Rawat Inap dengan perbandingan (32:0).
14. *IF* Tingkat Hipertensi *Stage 1* *AND* Level SpO2 Rendah *AND* Level BMI *Underweight* *THEN* Rawat Jalan dengan perbandingan (1:0).
15. *IF* Tingkat Hipertensi *Stage 1* *AND* Level SpO2 Sedang *AND* Jenis Kelamin Laki-laki *THEN* Rawat Inap dengan perbandingan (11:6).
16. *IF* Tingkat Hipertensi *Stage 1* *AND* Level SpO2 Sedang *AND* Jenis Kelamin Perempuan *THEN* Rawat Jalan dengan perbandingan (17:0).
17. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Takikardia *THEN* Rawat Inap dengan perbandingan (9:0).
18. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Takikardia *THEN* Rawat Inap dengan perbandingan (9:0).
19. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Healthy weight* *AND* Usia Muda *THEN* Rawat Inap dengan perbandingan (82:0).
20. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Healthy weight* *AND* Usia Medium *AND* Level SpO2 Normal *AND* Jenis Kelamin Laki-laki *THEN* Rawat Inap dengan perbandingan (16:6).
21. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Healthy weight* *AND* Usia Medium *AND* Level SpO2 Normal *AND* Jenis Kelamin Perempuan *THEN* Rawat Inap dengan perbandingan (9:4).

22. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Healthy weight* *AND* Usia Medium *AND* Level SpO2 Rendah *THEN* Rawat Inap dengan perbandingan (6:0).
23. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Healthy weight* *AND* Usia Medium *AND* Level SpO2 Sedang *THEN* Rawat Inap dengan perbandingan (6:0).
24. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Healthy weight* *AND* Usia Tua *AND* Jenis Kelamin Laki-laki *THEN* Rawat Jalan dengan perbandingan (2:1).
25. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Healthy weight* *AND* Usia Tua *AND* Jenis Kelamin Perempuan *THEN* Rawat Inap dengan perbandingan (28:8).
26. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Obesity* *THEN* Rawat Jalan dengan perbandingan (15:0).
27. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Overweight* *AND* Level SpO2 Normal *AND* Usia Medium *THEN* Rawat Inap dengan perbandingan (85:25).
28. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Overweight* *AND* Level SpO2 Normal *AND* Usia Muda *THEN* Rawat Inap dengan perbandingan (40:12).
29. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Overweight* *AND* Level SpO2 Normal *AND* Usia Tua *THEN* Rawat Inap dengan perbandingan (27:10).
30. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Overweight* *AND* Level SpO2 Rendah *THEN* Rawat Inap dengan perbandingan (141:5).
31. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Overweight* *AND* Level SpO2 Sedang *THEN* Rawat Inap dengan perbandingan (49:1).
32. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Normal *AND* Level BMI *Underweight* *THEN* Rawat Jalan dengan perbandingan (2:0).
33. *IF* Tingkat Hipertensi *Stage 2* *AND* Level Nadi Takikardia *THEN* Rawat Jalan dengan perbandingan (9:0)

Tabel 4. 18 *Confusion matrix* setelah SMOTE

Aktual: Rawat Inap	Aktual: Rawat Jalan	class precision
--------------------	---------------------	-----------------

pred. Rawat Inap	TP = 797	FP = 101	88.75%
pred. Rawat Jalan	FN = 30	TN = 726	96.03%
class recall	96.37%	87.79%	

Berdasarkan hasil *confusion matrix* diperoleh bahwa:

1. Jumlah prediksi *plan* perawatan untuk rawat inap dan data realita yang sesuai atau benar pasien menjalani rawat inap adalah sebanyak 797. Kondisi yang sesuai antara hasil prediksi dengan data realita *plan* perawatan di lapangan disebut *true positive*.
2. Jumlah prediksi *plan* perawatan untuk jenis rawat inap menunjukkan ketidaksesuaian antara prediksi dan data realita di lapangan yang menunjukkan data aktual yaitu rawat jalan adalah sebanyak 101. Kondisi ini disebut sebagai *false positive*.
3. Jumlah prediksi *plan* perawatan untuk jenis rawat jalan namun tidak sesuai dengan data aktual yang menunjukkan rawat inap yaitu sebanyak 30. Kondisi ini disebut sebagai *false negative*.
4. Jumlah prediksi *plan* perawatan untuk jenis rawat jalan dan data realita di lapangan benar pasien menunjukkan rawat jalan, yaitu sebanyak 726. Kondisi ini disebut juga sebagai *true negative*.

Berikut ini merupakan perhitungan dari *confusion matrix* dari model yang telah dihasilkan.

1. $Accuracy = \frac{TP+TN}{Total} = \frac{797+726}{1654} = 92.08\%$
2. $Misclassification Rate: \frac{FP+FN}{Total} = \frac{101+30}{1654} = 0.08 = 8\%$
3. $Recall/true positive rate$ (rawat inap): $\frac{TP}{TP+FN} = \frac{797}{797+30} = 96.37\%$
4. $Recall/true negative rate$ (rawat jalan): $\frac{TN}{TN+FP} = \frac{726}{726+101} = 87.79\%$
5. $Precision$ rawat inap: $\frac{TP}{TP+FP} = \frac{797}{797+101} = 88.75\%$
6. $Precision$ rawat jalan: $\frac{TN}{TN+FN} = \frac{726}{726+30} = 96.03\%$
7. $Prevelance: \frac{TP+FN}{total} = \frac{718+109}{1654} = 0.5 = 50\%$
8. $Area under curves$ (AUC) = 0.965

4.4.2.1 Perhitungan Manual Decision Tree C4.5

Langkah pertama yang harus dilakukan dalam melakukan kalkulasi *decision tree* secara manual adalah menghitung nilai *entropy* masing-masing variable. Berikut ini merupakan contoh

perhitungan untuk beberapa variable yang digunakan dalam penelitian ini, yaitu level nadi dan SpO2 dengan menggunakan data sebelum dilakukannya proses SMOTE.

1. *Entropy* keseluruhan data

Tabel 4. 19 Jumlah data keseluruhan

Jenis Rawat	Jumlah
Rawat Inap	10
Rawat Jalan	827
Total	837

$$\begin{aligned}
 Entropy(S) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{827}{837} \times \log_2 \left(\frac{827}{837} \right) \right) + \left(-\frac{10}{837} \times \log_2 \left(\frac{10}{837} \right) \right) = 0.73
 \end{aligned}$$

2. Level Nadi

1) *Entropy instance* Level Nadi

Tabel 4. 20 *record instance* level nadi

Level nadi	Rawat Inap	Rawat Jalan	Grand Total
Bradikardia	2	1	3
Normal	8	810	818
Takikardia	0	16	16
Grand Total	10	827	837

- Bradikardia

Berikut ini merupakan perhitungan *entropy instance* Bradikardia.

$$\begin{aligned}
 Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{2}{3} \times \log_2 \left(\frac{2}{3} \right) \right) + \left(-\frac{1}{3} \times \log_2 \left(\frac{1}{3} \right) \right) = 0.98
 \end{aligned}$$

- Normal

Berikut ini merupakan perhitungan *entropy instance* Normal.

$$\begin{aligned}
 Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{8}{818} \times \log_2 \left(\frac{8}{818} \right) \right) + \left(-\frac{810}{818} \times \log_2 \left(\frac{810}{818} \right) \right) = 0.72
 \end{aligned}$$

- Takikardia

Berikut ini merupakan perhitungan *entropy instance* Takikardia.

$$\begin{aligned} Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{0}{16} \times \log_2 \left(\frac{0}{16} \right) \right) + \left(-\frac{16}{16} \times \log_2 \left(\frac{16}{16} \right) \right) = 0 \end{aligned}$$

- 2) *Entropy* Atribut Level Nadi

$$\begin{aligned} Entropy(A) &= \sum_{i=1}^m \frac{|S_i|}{|S|} \times Entropy(S_i) \\ Entropy(A) &= \left(\left(\frac{3}{837} \times 0.98 \right) + \left(\frac{818}{837} \times 0.72 \right) + \left(\frac{16}{837} \times 0.98 \right) \right) = 0.71 \end{aligned}$$

- 3) *Information gain* atribut level nadi

$$\begin{aligned} Gain(S, A) &= Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(A) \\ Gain(S, A) &= 0.73 - 0.71 = 0.02 \end{aligned}$$

- 4) *Split information* level nadi

$$\begin{aligned} split\ info &= \sum_{i=1}^c -\frac{S_i}{S} \times \log_2 \frac{S_i}{S} \\ split\ info &= \left(-\frac{3}{837} \times \log_2 \left(\frac{3}{837} \right) \right) - \left(\frac{818}{837} \times \log_2 \left(\frac{818}{837} \right) \right) \\ &\quad - \left(\frac{16}{837} \times \log_2 \left(\frac{16}{837} \right) \right) \\ split\ info &= 0.042 \end{aligned}$$

- 5) *Gain ratio* level nadi

$$Gain\ ratio = \frac{0.02}{0.04} = 0.40$$

3. Level SpO2

1) Entropy instance Level SpO2

Tabel 4. 21 record instance level SpO2

Level SpO2	Rawat Inap	Rawat Jalan	Grand Total
Cukup	0	1	1
Normal	6	780	786
Rendah	3	18	21
Sedang	1	28	29
Grand Total	10	827	837

- Cukup

Berikut ini merupakan perhitungan *entropy instance* Cukup

$$\begin{aligned}
 Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{0}{1} \times \log_2 \left(\frac{0}{1} \right) \right) + \left(-\frac{1}{1} \times \log_2 \left(\frac{1}{1} \right) \right) = 0
 \end{aligned}$$

- Normal

Berikut ini merupakan perhitungan *entropy instance* Normal

$$\begin{aligned}
 Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{6}{786} \times \log_2 \left(\frac{6}{786} \right) \right) + \left(-\frac{780}{786} \times \log_2 \left(\frac{780}{786} \right) \right) = 0.72
 \end{aligned}$$

- Rendah

Berikut ini merupakan perhitungan *entropy instance* Rendah

$$\begin{aligned}
 Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{3}{21} \times \log_2 \left(\frac{3}{21} \right) \right) + \left(-\frac{18}{21} \times \log_2 \left(\frac{18}{21} \right) \right) = 0.88
 \end{aligned}$$

- Sedang

Berikut ini merupakan perhitungan *entropy instance* Sedang

$$Entropy(S_i) = \sum_{i=1}^n -p_i \times \log_2 p_i$$

$$= \left(-\frac{1}{29} \times \log_2 \left(\frac{1}{29} \right) \right) + \left(-\frac{28}{29} \times \log_2 \left(\frac{28}{29} \right) \right) = 0.76$$

2) *Entropy* Atribut Level SpO2

$$\begin{aligned} \text{Entropy}(A) &= \sum_{i=1}^m \frac{|S_i|}{|S|} \times \text{Entropy}(S_i) \\ &= \left(\left(\frac{1}{837} \times 0 \right) + \left(\frac{786}{837} \times 0.72 \right) + \left(\frac{21}{837} \times 0.88 \right) + \left(\frac{29}{837} \times 0.76 \right) \right) = 0.72 \end{aligned}$$

3) *Information gain* atribut level SpO2

$$\begin{aligned} \text{Gain}(S, A) &= \text{Entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{Entropy}(A) \\ \text{Gain}(S, A) &= 0.73 - 0.72 = 0.02 \end{aligned}$$

4) *Split information* level SpO2

$$\begin{aligned} \text{split info} &= \sum_{i=1}^c -\frac{S_i}{S} \times \log_2 \frac{S_i}{S} \\ \text{split info} &= \left(-\frac{1}{837} \times \log_2 \left(\frac{1}{837} \right) \right) - \left(\frac{786}{837} \times \log_2 \left(\frac{786}{837} \right) \right) \\ &\quad - \left(\frac{21}{837} \times \log_2 \left(\frac{21}{837} \right) \right) - \left(\frac{29}{837} \times \log_2 \left(\frac{29}{837} \right) \right) \\ \text{split info} &= 0.040 \end{aligned}$$

5) *Gain ratio* level SpO2

$$\text{Gain ratio} = \frac{0.02}{0.04} = 0.01$$

4. Usia

1) *Entropy instance* Usia

Tabel 4. 22 record instance Usia

Usia	Rawat Inap	Rawat Jalan	Grand Total
------	------------	-------------	-------------

Medium	4	363	367
Muda	3	43	46
Tua	3	421	424
Grand Total	10	827	837

- Medium

Berikut ini merupakan perhitungan *entropy instance* Medium

$$\begin{aligned}
 Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{4}{367} \times \log_2 \left(\frac{4}{367} \right) \right) + \left(-\frac{363}{367} \times \log_2 \left(\frac{363}{367} \right) \right) = 0.73
 \end{aligned}$$

- Muda

Berikut ini merupakan perhitungan *entropy instance* Muda

$$\begin{aligned}
 Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{3}{46} \times \log_2 \left(\frac{3}{46} \right) \right) + \left(-\frac{43}{46} \times \log_2 \left(\frac{43}{46} \right) \right) = 0.80
 \end{aligned}$$

- Tua

Berikut ini merupakan perhitungan *entropy instance* Tua

$$\begin{aligned}
 Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 Entropy(S_i) &= \left(-\frac{3}{46} \times \log_2 \left(\frac{3}{46} \right) \right) + \left(-\frac{43}{46} \times \log_2 \left(\frac{43}{46} \right) \right) = 0.80
 \end{aligned}$$

2) Entropy Atribut Usia

$$\begin{aligned}
 Entropy(A) &= \sum_{i=1}^m \frac{|S_i|}{|S|} \times Entropy(S_i) \\
 &= \left(\left(\frac{367}{837} \times 0 \right) + \left(\frac{46}{837} \times 0.72 \right) + \left(\frac{424}{837} \times 0.88 \right) \right) = 0.725
 \end{aligned}$$

3) Information gain atribut Usia

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(A)$$

$$Gain(S, A) = 0.727 - 0.725 = 0.002$$

4) *Split information* Usia

$$split\ info = \sum_{i=1}^c -\frac{S_i}{S} \times \log_2 \frac{S_i}{S}$$

$$split\ info = \left(-\frac{367}{837} \times \log_2 \left(\frac{367}{837} \right) \right) - \left(\frac{46}{837} \times \log_2 \left(\frac{46}{837} \right) \right) \\ - \left(\frac{424}{837} \times \log_2 \left(\frac{424}{837} \right) \right)$$

$$split\ info = 0.24$$

5) *Gain ratio* Usia

$$Gain\ ratio = \frac{Gain(S, A)}{split\ info} = \frac{0.002}{0.24} = 0.01$$

5. Tingkat hipertensi

1) *Entropy instance* Tingkat Hipertensi

Tabel 4. 23 *Record instance* tingkat hipertensi

Row Labels	Rawat Inap	Rawat Jalan	Grand Total
Normal	0	69	69
Pra-Hipertensi	2	184	186
Stage 1	1	453	454
Stage 2	7	121	128
Grand Total	10	827	837

- Normal

Berikut ini merupakan perhitungan *entropy instance* Normal

$$Entropy(S_i) = \sum_{i=1}^n -p_i \times \log_2 p_i \\ = \left(-\frac{0}{69} \times \log_2 \left(\frac{0}{69} \right) \right) + \left(-\frac{69}{69} \times \log_2 \left(\frac{69}{69} \right) \right) = 0$$

- Pra-Hipertensi

Berikut ini merupakan perhitungan *entropy instance* Pra-Hipertensi

$$\begin{aligned} Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{2}{186} \times \log_2 \left(\frac{2}{186} \right) \right) + \left(-\frac{184}{186} \times \log_2 \left(\frac{184}{186} \right) \right) = 0.72 \end{aligned}$$

- Stage 1

Berikut ini merupakan perhitungan *entropy instance stage 1*

$$\begin{aligned} Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{1}{454} \times \log_2 \left(\frac{1}{454} \right) \right) + \left(-\frac{453}{454} \times \log_2 \left(\frac{453}{454} \right) \right) = 0.71 \end{aligned}$$

- Stage 2

Berikut ini merupakan perhitungan *entropy instance stage 2*

$$\begin{aligned} Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{7}{128} \times \log_2 \left(\frac{7}{128} \right) \right) + \left(-\frac{121}{128} \times \log_2 \left(\frac{121}{128} \right) \right) = 0.79 \end{aligned}$$

2) Entropy Atribut Tingkat hipertensi

$$\begin{aligned} Entropy(A) &= \sum_{i=1}^m \frac{|S_i|}{|S|} \times Entropy((S_i)) \\ &= \left(\left(\frac{69}{837} \times 0 \right) + \left(\frac{186}{837} \times 0.72 \right) + \left(\frac{454}{837} \times 0.71 \right) + \left(\frac{128}{837} \times 0.79 \right) \right) = 0.66 \end{aligned}$$

3) Information gain atribut Tingkat hipertensi

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(A)$$

$$Gain(S, A) = 0.73 - 0.66 = 0.06$$

4) *Split information* Tingkat hipertensi

$$\begin{aligned} \text{split info} &= \sum_{i=1}^c -\frac{S_i}{S} \times \log_2 \frac{S_i}{S} \\ \text{split info} &= \left(-\frac{69}{837} \times \log_2 \left(\frac{69}{837} \right) \right) - \left(\frac{186}{837} \times \log_2 \left(\frac{186}{837} \right) \right) \\ &\quad - \left(\frac{454}{837} \times \log_2 \left(\frac{454}{837} \right) \right) - \left(\frac{128}{837} \times \log_2 \left(\frac{128}{837} \right) \right) \\ \text{split info} &= 1.67 \end{aligned}$$

5) *Gain ratio* Usia

$$\text{Gain ratio} = \frac{\text{Gain}(S, A)}{\text{split info}} = \frac{0.06}{1.67} = 0.04$$

6. Kesadaran

Tabel 4. 24 *Record instance* kesadaran

Kesadaran	Rawat Inap	Rawat Jalan	Grand Total
Compos Mentis	10	827	837
Grand Total	10	827	837

1) *Entropy instance* Kesadaran

$$\begin{aligned} \text{Entropy}(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{10}{837} \times \log_2 \left(\frac{10}{837} \right) \right) + \left(-\frac{827}{837} \times \log_2 \left(\frac{827}{837} \right) \right) = 0.73 \end{aligned}$$

2) *Entropy atribut* Kesadaran

$$\begin{aligned} \text{Entropy}(A) &= \sum_{i=1}^m \frac{|S_i|}{|S|} \times \text{Entropy}(S_i) \\ &= \left(\left(\frac{837}{837} \times 0.73 \right) \right) = 0.73 \end{aligned}$$

3) *Information gain* atribut Tingkat hipertensi

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(A)$$

$$Gain(S, A) = 0.73 - 0.73 = 0$$

4) *Split information* Tingkat hipertensi

$$split\ info = \sum_{i=1}^c -\frac{S_i}{S} \times \log_2 \frac{S_i}{S}$$

$$split\ info = \left(-\frac{837}{837} \times \log_2 \left(\frac{837}{837} \right) \right)$$

$$split\ info = 0$$

5) *Gain ratio* Usia

$$Gain\ ratio = \frac{Gain(S, A)}{split\ info} = \frac{0}{0} = 0$$

7. Level BMI

1) *Entropy instance* Level BMI

Tabel 4. 25 *Record instance* tingkat hipertensi

Row Labels	Rawat Inap	Rawat Jalan	Grand Total
Healthy Weight	3	73	76
Obesity	0	35	35
Overweight	7	713	720
Underweight	0	6	6
Grand Total	10	827	837

- *Healthy weight*

Berikut ini merupakan perhitungan *entropy instance* Normal

$$Entropy(S_i) = \sum_{i=1}^n -p_i \times \log_2 p_i$$

$$= \left(-\frac{3}{76} \times \log_2 \left(\frac{3}{76} \right) \right) + \left(-\frac{73}{76} \times \log_2 \left(\frac{73}{76} \right) \right) = 0.77$$

- *Obesity*

Berikut ini merupakan perhitungan *entropy instance* Pra-Hipertensi

$$\begin{aligned} Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{0}{35} \times \log_2 \left(\frac{0}{35} \right) \right) + \left(-\frac{35}{35} \times \log_2 \left(\frac{35}{35} \right) \right) = 0 \end{aligned}$$

- *Overweight*

Berikut ini merupakan perhitungan *entropy instance stage 1*

$$\begin{aligned} Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{7}{720} \times \log_2 \left(\frac{7}{720} \right) \right) + \left(-\frac{713}{720} \times \log_2 \left(\frac{713}{720} \right) \right) = 0.72 \end{aligned}$$

- *Underweight*

Berikut ini merupakan perhitungan *entropy instance stage 2*

$$\begin{aligned} Entropy(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\ &= \left(-\frac{0}{6} \times \log_2 \left(\frac{0}{6} \right) \right) + \left(-\frac{6}{6} \times \log_2 \left(\frac{6}{6} \right) \right) = 0 \end{aligned}$$

2) *Entropy* Atribut Level BMI

$$\begin{aligned} Entropy(A) &= \sum_{i=1}^m \frac{|S_i|}{|S|} \times Entropy((S_i)) \\ &= \left(\left(\frac{76}{837} \times 0.77 \right) + \left(\frac{35}{837} \times 0 \right) + \left(\frac{720}{837} \times 0.72 \right) + \left(\frac{6}{837} \times 0 \right) \right) = 0.69 \end{aligned}$$

3) *Information gain* atribut Level BMI

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times Entropy(A)$$

$$Gain(S, A) = 0.73 - 0.66 = 0.04$$

4) *Split information* Level BMI

$$\begin{aligned}
 \text{split info} &= \sum_{i=1}^c -\frac{S_i}{S} \times \log_2 \frac{S_i}{S} \\
 \text{split info} &= \left(-\frac{76}{837} \times \log_2 \left(\frac{76}{837} \right) \right) - \left(\frac{35}{837} \times \log_2 \left(\frac{35}{837} \right) \right) \\
 &\quad - \left(\frac{720}{837} \times \log_2 \left(\frac{720}{837} \right) \right) - \left(\frac{6}{837} \times \log_2 \left(\frac{6}{837} \right) \right) \\
 \text{split info} &= 0.74
 \end{aligned}$$

5) *Gain ratio* Usia

$$\text{Gain ratio} = \frac{\text{Gain}(S, A)}{\text{split info}} = \frac{0.04}{0.74} = 0.05$$

8. Jenis Kelamin

1) *Entropy instance* Jenis Kelamin

Tabel 4. 26 Record instance Jenis kelamin

Row Labels	Rawat Inap	Rawat Jalan	Grand Total
Laki-Laki	6	323	329
Perempuan	4	504	508
Total	10	827	837

- Laki-laki

Berikut ini merupakan perhitungan *entropy instance* Laki-laki

$$\begin{aligned}
 \text{Entropy}(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{6}{329} \times \log_2 \left(\frac{6}{329} \right) \right) + \left(-\frac{323}{329} \times \log_2 \left(\frac{323}{329} \right) \right) = 0.74
 \end{aligned}$$

- Perempuan

Berikut ini merupakan perhitungan *entropy instance* Perempuan

$$\begin{aligned}
 \text{Entropy}(S_i) &= \sum_{i=1}^n -p_i \times \log_2 p_i \\
 &= \left(-\frac{4}{508} \times \log_2 \left(\frac{4}{508} \right) \right) + \left(-\frac{504}{508} \times \log_2 \left(\frac{504}{508} \right) \right) = 0.72
 \end{aligned}$$

6) *Entropy* Atribut Jenis kelamin

$$\begin{aligned} \text{Entropy}(A) &= \sum_{i=1}^m \frac{|S_i|}{|S|} \times \text{Entropy}((S_i)) \\ &= \left(\left(\frac{329}{837} \times 0.74 \right) + \left(\frac{508}{837} \times 0.72 \right) \right) = 0.72 \end{aligned}$$

7) *Information gain* atribut Jenis kelamin

$$\begin{aligned} \text{Gain}(S, A) &= \text{Entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{Entropy}(A) \\ \text{Gain}(S, A) &= 0.72 - 0.72 = 0 \end{aligned}$$

8) *Split information* Jenis kelamin

$$\begin{aligned} \text{split info} &= \sum_{i=1}^c -\frac{S_i}{S} \times \log_2 \frac{S_i}{S} \\ \text{split info} &= \left(-\frac{329}{837} \times \log_2 \left(\frac{329}{837} \right) \right) - \left(\frac{508}{837} \times \log_2 \left(\frac{508}{837} \right) \right) \\ \text{split info} &= 0.97 \end{aligned}$$

9) *Gain ratio* Jenis kelamin

$$\text{Gain ratio} = \frac{\text{Gain}(S, A)}{\text{split info}} = \frac{0}{0.97} = 0$$

4.4.2.2 Rekapitulasi perhitungan root node

Berikut ini merupakan rekapitulasi hasil perhitungan *decision tree* dalam penentuan *root node*.

Tabel 4. 27 Rekapitulasi perhitungan manual

		Tingkat Hipertensi					
Tingakt Hipertensi	Hasil	Juml ah	Entropy instance	Entropy Atribut	Information Gain	Split info	Gain ratio
Normal	Rawat Inap	0	0.00	0.66	0.06	1.67	0.04

	Rawat Jalan	69					
Pra-Hipertensi	Rawat Inap	2	0.72				
	Rawat Jalan	184					
Stage 1	Rawat Inap	1	0.71				
	Rawat Jalan	453					
Stage 2	Rawat Inap	7	0.79				
	Rawat Jalan	121					
Level Nadi							
Level Nadi	Hasil	Jumlah	Entropy instance	Entropy Atribut	Information Gain	Split info	Gain ratio
Bradikardia	Rawat Inap	2	0.98				
	Rawat Jalan	1					
Normal	Rawat Inap	8	0.72	0.71	0.02	0.04	0.40
	Rawat Jalan	810					
Takikardia	Rawat Inap	0	0.00				
	Rawat Jalan	16					
Level SpO2							
Level SpO2	Hasil	Jumlah	Entropy instance	Entropy Atribut	Information Gain	Split info	Gain ratio

Cukup	Rawat Inap	0	0.00				
	Rawat Jalan	1					
Normal	Rawat Inap	6	0.72				
	Rawat Jalan	780					
				0.72	0	0.40	0.01
Rendah	Rawat Inap	3	0.88				
	Rawat Jalan	18					
Sedang	Rawat Inap	1	0.76				
	Rawat Jalan	28					
Usia							
Usia	Hasil	Jumlah	Entropy instance	Entropy Atribut	Information Gain	Split info	Gain ratio
Medium	Rawat Inap	4	0.73				
	Rawat Jalan	363					
Muda	Rawat Inap	3	0.80	0.73	0	0.24	0.01
	Rawat Jalan	43					
Tua	Rawat Inap	3	0.72				
	Rawat Jalan	421					
Kesadaran							

Kesadaran	Hasil	Jumlah	Entropy instance	Entropy Atribut	Information Gain	Split info	Gain ratio
Compos Mentis	Rawat Inap	10	0.73	0.73	0	0	0
	Rawat Jalan	827					
Jenis Kelamin							
Jenis Kelamin	Hasil	Jumlah	Entropy instance	Entropy Atribut	Information Gain	Split info	Gain ratio
Laki-laki	Rawat Inap	6	0.74	0.73	0	0.97	0
	Rawat Jalan	323					
Perempuan	Rawat Inap	4	0.72	0.73	0	0.97	0
	Rawat Jalan	504					
Level BMI							
Level BMI	Hasil	Jumlah	Entropy instance	Entropy Atribut	Information Gain	Split info	Gain ratio
Underweight	Rawat Inap	0	0.00	0.73	0.04	0.74	0.05
	Rawat Jalan	6					
Healthy weight	Rawat Inap	3	0.77	0.73	0.04	0.74	0.05
	Rawat Jalan	73					
Overweight	Rawat Inap	7	0.72	0.73	0.04	0.74	0.05
	Rawat Jalan	713					

	Rawat	0	
	Inap		
Obesity	<hr/>		0.00
	Rawat	35	
	Jalan		

Berdasarkan tabel 4.25, dapat diketahui bahwa atribut yang memiliki nilai *gain ratio* tertinggi adalah atribut level nadi dengan nilai sebesar 0,40. Sehingga atribut level nadi dalam model *decision tree* yang dikembangkan akan bertindak sebagai *root node*. Perhitungan untuk menentukan internal node memiliki proses yang sama dengan perhitungan dalam menentukan *root node* hanya saja terdapat perbedaan dalam data yang digunakan. Pada perhitungan selanjutnya, diperlukan penyaringan data pada level nadi bradikardia. Sedangkan level nadi takikardia akan langsung memberikan keputusan (*leaf node*) yaitu rawat jalan dengan jumlah data rawat jalan yang *direcord* sebanyak 16 data. Proses akan berhenti apabila masing-masing *instance* dalam tiap atribut telah menghasilkan *entropy* berjumlah 0. Perhitungan manual yang dilakukan menunjukkan level nadi sebagai *root node* selaras dengan model yang berhasil dihasilkan oleh *rapidminer*.

BAB V

PEMBAHASAN

5.1 Analisis Rules Analisis Klasifikasi Algoritma *Decision Tree* C4.5

Dalam penelitian yang dilakukan, diperoleh 2 hasil aturan/*rules decision tree* yang berbeda. Perbedaan yang terjadi akibat pemberian dua *treatment* yang berbeda pada data yang digunakan dalam penelitian ini. Pada percobaan pertama, *decision tree* dilakukan dengan kondisi data yang mengalami ketimpangan (*imbalanced*) pada label data yaitu data rawat inap dan rawat jalan. Tercatat, ketimpangan yang terjadi pada label data yang digunakan sangat signifikan, yang mana data rawat jalan hanya menyentuh angka 1% dari total keseluruhan data dengan jumlah data rawat jalan berjumlah 827 data dan data rawat inap berjumlah 10 data. Pada percobaan kedua, ketimpangan data yang terjadi, telah diatasi dengan menggunakan metode SMOTE (*synthetic minority over-sampling technique*). Setelah ketimpangan data yang terjadi diatasi dengan SMOTE, perbandingan antara data rawat jalan dengan data rawat inap menjadi seimbang dengan perbandingan 50:50 yang membuat total jumlah data menjadi 1654 data.

Aturan/*rules* yang terbentuk antara 2 percobaan yang dilakukan menunjukkan hasil yang berbeda. Perbedaan signifikan yang terjadi terletak pada atribut yang bertindak menjadi *root node* dalam model yang dihasilkan, baik percobaan pertama tanpa SMOTE dan percobaan kedua dengan menggunakan metode SMOTE. Pada percobaan pertama tanpa menggunakan metode SMOTE, atribut yang menjadi *root node* adalah variable level nadi dengan jumlah *rules* yang terbentuk sebanyak 20 *rules*. Sedangkan pada percobaan yang menggunakan SMOTE untuk mengatasi ketimpangan data yang terjadi, atribut yang bertindak sebagai *root node* yaitu atribut tingkat hipertensi dengan jumlah *rules* yang terbentuk sebanyak 33 *rules*. Perbedaan *root node* yang terjadi disebabkan oleh bertambahnya jumlah *instance* masing-masing atribut setelah diterapkannya metode SMOTE. Bertambahnya jumlah *instance* dari masing-masing atribut menyebabkan berubahnya nilai *split information*. Semakin kecil nilai *split information* suatu atribut maka semakin besar pula nilai *gain ratio* dari suatu atribut, sehingga membuat atribut tersebut menjadi atribut yang paling baik dalam melakukan klasifikasi. Berdasarkan hasil wawancara dengan dokter, diketahui bahwa tanda-tanda vital yang paling penting dari pasien dengan penyakit hipertensi adalah tingkat hipertensi, sehingga bisa dikatakan bahwa model dengan integrasi metode SMOTE menghasilkan model yang valid menurut ilmu medis.

5.2 Evaluasi Performa Klasifikasi

Pada analisis klasifikasi, performa yang dihasilkan ditinjau melalui sebuah alat bernama *confusion matrix*. Berikut ini merupakan evaluasi performa klasifikasi tanpa menggunakan metode SMOTE.

Tabel 5. 1 Performa klasifikasi

Parameter	Nilai
<i>Accuracy</i>	99%
<i>Misclassification rate</i>	1%
<i>Recall/true positive rate</i>	20%
<i>Recall/ true negative rate</i>	99.27%
<i>Precision rawat inap</i>	25%
<i>Precision rawat jalan</i>	99.03%
<i>Prevelance</i>	1.1%
<i>Area under curve</i>	55.7%

Pada hasil evaluasi analisis klasifikasi yang telah ditampilkan pada Tabel 5.1, memiliki nilai keakuratan sebesar 99% yang artinya kebenaran model dalam melakukan prediksi, baik memprediksi kelas data rawat jalan dan rawat inap dari total keseluruhan data, sebanyak 99% prediksi benar. Dengan tingkat kesalahan model dalam melakukan prediksi yaitu sebesar 1%. *Recall true positive rate* (rawat inap) menunjukkan nilai seberapa sering model memprediksi atau memberikan label rawat inap jika data aktual adalah rawat inap, Dalam hal ini, model dapat memprediksi sebesar 20% dari total keseluruhan data rawat inap. Sedangkan *Recall/ true negative rate* (rawat jalan) memperoleh nilai 99.27% yang berarti menunjukkan seberapa sering model memberikan label rawat jalan jika data actual adalah data rawat jalan. Dalam hal presisi, model menghasilkan presisi untuk label kelas data rawat inap menghasilkan nilai sebesar 25% dan untuk label kelas data rawat jalan menghasilkan nilai sebesar 99.03%, yang artinya ketika model mengeluarkan prediksi rawat inap maka kemungkinan model benar adalah 25% dan apabila model mengeluarkan prediksi berupa rawat jalan, maka kemungkinan model benar dalam melakukan prediksi adalah sebesar 99.03%. kemampuan model dalam memisahkan antara kelas data positif dan negative pada model tanpa integrasi dengan metode SMOTE, menghasilkan performa yang buruk. Hal ini dibuktikan dengan nilai AUC yang hanya sebesar

55.7% Untuk nilai *prevelance* merupakan nilai yang menunjukkan seberapa banyak rawat inap yang terjadi secara actual yang dalam hal ini sebesar 1,1%.

besarnya *gap* nilai nilai *precision* dan *recall* antara data rawat inap dan rawat jalan menunjukkan bahwa model yang dihasilkan tidak mampu memprediksi data rawat inap yang disebabkan karena model hanya menghafal data yang digunakan, bukan memahami pola data yang ada. Sehingga ketika model diberikan data rawat inap yang belum pernah dilihat oleh model sebelumnya atau selain *rules*/aturan tersebut, model hanya mencoba untuk mengepaskan data baru yang belum pernah dilihat oleh model dengan data yang sudah biasa dilihat yang dalam hal ini adalah data rawat jalan. Dalam istilah *machine learning* tingkah laku model ini disebut sebagai *overfitting*. Selain itu, *overfitting* yang terjadi dibuktikan dengan nilai akurasi pada data *training* yang lebih besar dibandingkan dengan *data testing* dengan selisih 0.6%. Selain itu, kurangnya data dengan label rawat inap membuat model tidak memiliki sumber daya yang banyak untuk belajar dan memahami pola data rawat inap sehingga menghasilkan performa yang buruk dalam memprediksi kelas data minoritas (rawat inap). Perbedaan nilai *recall* dan tingginya nilai *precision* yang signifikan disebabkan oleh ketimpangan data rawat jalan dan rawat inap yang membuat model kurang terlatih dalam melakukan prediksi data rawat inap sehingga menghasilkan model yang *overfit* yang menyebabkan model cenderung mengabaikan kelas data minoritas, yang dalam hal ini merupakan data dengan label rawat inap.

5.3 Perbandingan Performa Model SMOTE dan Non-SMOTE

Berikut ini merupakan perbandingan performa model klasifikasi yang dihasilkan dengan dan tanpa mengatasi ketimpangan data yang terjadi.

Tabel 5. 2 Perbandingan performa SMOTE dan Non-SMOTE

Parameter	Tanpa SMOTE	SMOTE	Random Undersampling
<i>Accuracy</i>	99%	92.08%	80%
<i>Misclassification rate</i>	1%	8%	20%
<i>Recall/true positive rate</i>	20%	96.37%	80%
<i>Recall/ true negative rate</i>	99.27%	87.79%	80%
<i>Precision rawat inap</i>	25%	88.75%	80%
<i>Precision rawat jalan</i>	99.03%	96.03%	80%
<i>Prevelance</i>	1.1%	50%	50%

<i>Area under curve</i>	55.7%	96.5%	40%
-------------------------	-------	-------	-----

Berdasarkan Tabel 5.2, diketahui perbedaan performa dari model *decision tree* sebelum dan sesudah ketimpangan data diatasi, terjadi pada seluruh parameter. penurunan performa terdapat pada beberapa parameter, yaitu nilai keakuratan/*accuracy* model dengan menggunakan SMOTE mengalami penurunan 7% dibandingkan dengan model tanpa menggunakan SMOTE, nilai *recall true negative rate* yang menurun 12% dan nilai *precision* rawat jalan turun 3%. Namun peningkatan terjadi di beberapa parameter, tepatnya pada parameter yang mengukur performa model dalam memprediksi label kelas data minoritas atau rawat inap. Peningkatan terjadi pada *recall true positive rate* sebesar 76%, nilai *precision* untuk kelas data rawat inap mengalami peningkatan sebesar 63% dan data actual menunjukkan jumlah data kelas minoritas meningkat menjadi 50% dari total keseluruhan data. Selain itu, model dengan menggunakan integrasi metode SMOTE menghasilkan model yang baik dalam memisahkan antara kelas data positif dengan kelas data negative dengan selisih nilai AUC adalah 40.8%. Nilai AUC sendiri digunakan untuk mengetahui sebaik apa model klasifikasi yang dihasilkan dalam mengidentifikasi, membedakan dan memisahkan antara kelas data positif dan kelas data negatif. Semakin nilai AUC mendekati 1 maka semakin baik pula model dalam memisahkan kelas data positif dan negative. Peningkatan performa yang dialami model disebabkan oleh bertambahnya data dengan label kelas data rawat inap. Peningkatan jumlah data rawat inap membuat model dapat belajar lebih banyak untuk memahami pola data rawat inap dan rawat jalan, sehingga model akan lebih berpengalaman dalam membedakan pola data rawat inap atau rawat jalan.

Penggunaan metode SMOTE untuk menangani ketimpangan data yang terjadi menunjukkan hal positif pada model, karena setelah penerapan SMOTE, model menunjukkan hasil yang baik secara menyeluruh. Pada model sebelum diterapkannya metode SMOTE, terjadi kecenderungan model memprediksi label kelas data mayor dan mengabaikan label kelas data minor dengan akibat klasifikasi menjadi kurang akurat. Namun setelah diterapkannya metode SMOTE, masalah tersebut berhasil diatasi yang ditandai dengan nilai *recall* dan *precision* baik untuk kelas data mayoritas dan minoritas memiliki perbedaan yang tidak signifikan. Hasil ini sejalan dengan penelitian yang dilakukan oleh Sofyan & Prasetyo (2021) yang menyatakan bahwa, Penerapan SMOTE pada data yang mengalami ketimpangan memberikan hasil yang baik secara keseluruhan parameter. berbeda dengan model tanpa menggunakan metode SMOTE, model yang diintegrasikan dengan SMOTE berhasil menangani *overfitting* yang

terjadi. Menurut penelitian yang dilakukan oleh Wijayanti et al (2021) model yang menggunakan SMOTE memiliki kelebihan dalam mengatasi kemungkinan informasi yang hilang, menghindarkan model dari *overfitting* dan berhasil meningkatkan akurasi prediksi untuk kelas data minoritas, yang dalam hal ini adalah data dengan label rawat inap. Sehingga, berdasarkan analisis yang telah dilakukan, diketahui bahwa model *decision tree* yang diintegrasikan dengan Teknik *oversampling* SMOTE memiliki performa yang lebih baik dan positif dibandingkan dengan model yang menggunakan *imbalanced data*.

Perbandingan antara SMOTE dan penggunaan Teknik *undersampling* memiliki perbedaan hasil yang signifikan ditinjau dari berbagai parameter. dilihat dari parameter yang digunakan untuk mengukur performa model yang dihasilkan, diketahui bahwa, penggunaan metode SMOTE unggul di semua parameter yang digunakan. Hasil yang ditunjukkan oleh metode *random undersampling* adalah konsisten di angka 80%. Hal ini disebabkan oleh kurangnya sumber daya data untuk melatih dan membangun model sehingga kombinasi aturan yang tercipta menjadi sangat terbatas yang membuat seluruh nilai pada masing-masing parameter menunjukkan nilai yang sama baik untuk kelas data mayoritas maupun kelas data minoritas. Untuk mengatasi hal tersebut dilakukanlah pembobotan saat Pembangunan model. Pembobotan (*upweighted*) dilakukan agar data yang mengalami *downsampling* memiliki bobot setara dengan kelas minoritas. Penggunaan Teknik *undersampling* untuk menangani ketimpangan kelas data dalam penelitian ini dinilai kurang tepat karena kelas data minoritas dari *dataset* yang digunakan sangat sedikit, sehingga ketika menggunakan Teknik *undersampling* data minoritas menjadi sangat sedikit yang membuat hilangnya informasi yang terkandung dalam *dataset*.

5.4 Analisis Hasil Penggunaan Tipe Data Berbeda

Menurut Kamagi & Hansun (2016), secara teori, algoritma decision tree C4.5 memiliki kelebihan yang fleksibel dalam penggunaan tipe data baik tipe data numerik maupun data diskrit. Dalam penelitian ini ditemukan terdapat perbedaan performa yang cukup signifikan saat model menggunakan tipe data numerik dan tipe data kategori. Perbedaan terjadi pada *rules* yang dibentuk oleh model menghasilkan sebanyak 9 aturan/*rules*, jauh lebih sedikit dibandingkan dengan model yang menggunakan data bertipe diskrit yang menghasilkan aturan sebanyak 20 aturan/*rules*. Penggunaan data diskrit dalam *decision tree* membuat model menghasilkan keputusan yang lebih *general* (umum) sehingga memangkas kemungkinan terjadinya

overfitting dan menghasilkan aturan yang lebih banyak dibandingkan dengan menggunakan data bertipe numerik. Selain itu, pada data dengan menggunakan data numerik, indikasi terjadinya model mengalami *overfitting* semakin jelas. Hal ini dibuktikan dengan melihat nilai akurasi pada saat menggunakan data *training* dan data *testing*. Apabila model ketika menggunakan data latih menunjukkan akurasi yang lebih baik dibandingkan dengan data uji maka, model mengalami *overfitting*. Pada model dengan menggunakan tipe data numerik, Diketahui bahwa akurasi data latih lebih besar dengan skor akurasi sebesar 100% sedangkan ketika model diuji dengan data *testing* model menghasilkan skor sebesar 99.40%. berbeda dengan ketika model menggunakan data dengan tipe diskrit, model menghasilkan nilai akurasi 99% baik dengan menggunakan data *training* maupun data *testing*. Sehingga, penggunaan data dengan tipe numerik dalam kasus ini, memperjelas kondisi *overfitting* yang dialami oleh model sebelum integrasi dengan metode SMOTE. Dalam hal memisahkan antara label kelas data positif dengan label kelas data negative, model dengan menggunakan data diskrit menghasilkan performa yang lebih baik. Hal ini ditunjukkan dengan nilai AUC pada model menggunakan data diskrit menyentuh angka 55.7% sedangkan model dengan menggunakan data bertipe numerik hanya menghasilkan skor 25%.

5.5 Analisis Perbaikan Masalah

Permasalahan yang dialami oleh rumah sakit As-Syifa Manna adalah permasalahan yang umum di alami oleh rumah sakit lain, yaitu masalah waktu tunggu bagi pasien rawat inap. Permasalahan yang terjadi adalah system kerja rumah sakit saat ini memberikan waktu tunggu yang cukup lama bagi pasien yang telah divonis rawat inap oleh dokter. Hal ini disebabkan oleh ketika pasien telah mendapatkan vonis rawat inap, pasien harus kembali mengurus administrasi (mendaftar) agar semua kebutuhan bagi pasien tersebut dapat dipersiapkan oleh pihak rumah sakit. Proses inilah yang memberikan waktu tunggu bagi pasien rawat inap. Melalui penelitian yang dilakukan, penulis membbuktikan algoritma yang dapat melakukan prediksi terhadap *output* pasien apakah pasien akan diklasifikasikan sebagai pasien rawat inap atau rawat jalan. Algoritma *decision tree* C4.5 yang digunakan terbukti efektif dalam melakukan prediksi, sehingga untuk mengatasi permasalahan waktu tunggu yang dialami oleh pasien rawat inap, pihak rumah sakit dapat menyematkan algoritma *decision tree* C4.5 ke sistem informasi manajemen rumah sakit (SIMRS) atau mengembangkan sebuah aplikasi baru yang dikhususkan untuk melakukan prediksi terhadap *output* pasien agar terhindar dari waktu tunggu. Penerapan

aplikasi prediksi ini dapat diterapkan pada bagian *nurse station*, yang mana bagian ini adalah bagian dimana tanda-tanda vital pasien akan diperiksa. Sehingga pihak perawat telah dapat memperkirakan *output* pasien dan bisa segera mempersiapkan seluruh kemungkinan yang dibutuhkan oleh pasien. Dengan sistem kerja seperti ini, sesaat setelah pasien menerima vonis rawat inap oleh dokter, pasien sudah langsung mendapatkan seluruh kebutuhan pengobatan seperti kamar, rencana perawatan, pendaftaran dan administrasi.

Perbaikan masalah yang ditawarkan melalui penelitian ini sudah sejalan dengan saran pada penelitian yang dilakukan oleh Dewi (2015) yang menyatakan bahwa, salah satu ide perbaikan untuk mengatasi waktu tunggu yang dialami pasien rawat inap adalah dengan melakukan optimalisasi sistem rumah sakit. Melalui penelitian ini, penerapan algoritma *decision tree* untuk prediksi *plan* perawatan pasien merupakan salah satu langkah optimalisasi terhadap sistem rumah sakit saat ini, sehingga diharapkan dapat membantu permasalahan waktu tunggu yang dialami oleh pasien rawat inap.

BAB VI

PENUTUP

6.1 Kesimpulan

Berdasarkan penelitian yang telah dilakukan, dapat disimpulkan bahwa:

1. Dari 2 model yang berhasil terbentuk menunjukkan perbedaan dari sisi *root node* yang terbentuk dan jumlah *rules*. Pada model dengan menggunakan data yang tidak seimbang menghasilkan *root node* level nadi, sedangkan model yang diintegrasikan dengan SMOTE untuk menyeimbangkan kelas data, menghasilkan *root node* tingkat hipertensi. Dari sisi jumlah aturan/*rules* yang terbentuk, model tanpa menggunakan SMOTE menghasilkan sebanyak 20 aturan/*rules* dan model dengan menggunakan SMOTE menghasilkan sebanyak 33 aturan/*rules*. Sesuai dengan validasi dokter, variable tingkat hipertensi merupakan variable yang paling berpengaruh terhadap penilaian pasien pengidap penyakit hipertensi, sehingga hasil model yang diintegrasikan dengan SMOTE dapat dijadikan pola untuk menentukan pasien rawat inap/jalan.
2. Performa model klasifikasi tanpa menggunakan SMOTE menunjukkan performa yang sangat baik dalam memprediksi kelas data rawat jalan, namun sangat buruk dalam memprediksi kelas data rawat inap. Secara keseluruhan, akurasi model tanpa menggunakan SMOTE adalah sebesar 99.04%.
3. Penggunaan metode *oversampling*, yaitu SMOTE memiliki tingkat akurasi yang lebih rendah dibandingkan dengan model tanpa SMOTE yaitu sebesar 92.08%. Namun model dengan menggunakan SMOTE menghasilkan performa yang sangat baik dalam memprediksi kelas data minoritas (rawat inap) dan kelas data mayoritas (rawat jalan). Dengan kata lain model hasil integrasi dengan SMOTE memiliki performa yang lebih baik secara menyeluruh. Sehingga dapat disimpulkan bahwa, model hasil integrasi *decision tree* dan SMOTE memiliki kemampuan yang sama baiknya dalam memprediksi pasien rawat inap maupun rawat jalan
4. Rumah sakit dapat melakukan perbaikan dengan cara menerapkan algoritma klasifikasi *decision tree* C4.5 yang diintegrasikan dengan metode SMOTE ke dalam sistem rumah sakit atau mengembangkan sebuah aplikasi prediksi *plan* perawatan pasien agar pasien rawat inap dapat terhindar dari waktu tunggu.

6.2 Saran

Mengacu kepada hasil penelitian dan kesimpulan yang telah diperoleh dalam penelitian ini, penulis memberikan saran baik untuk pihak rumah sakit maupun penelitian yang akan dilakukan pada masa mendatang:

1. Pihak rumah sakit sebaiknya menggunakan penelitian ini untuk menjadi acuan agar rumah sakit bisa menerapkan system automasi dan mengatasi permasalahan waktu tunggu yang terjadi pada pasien rawat inap.
2. Untuk menghindari terbentuknya pohon keputusan yang bias, sebaiknya menggunakan metode lanjutan seperti *forecasting* dalam menangani data yang hilang. Sehingga dapat menghasilkan pohon keputusan yang lebih akurat dan tidak bias.
3. Dalam menangani data yang tidak seimbang, melalui penelitian ini, metode SMOTE terbukti ampuh untuk mengatasi permasalahan tersebut dengan menunjukkan hasil yang positif dan performa yang baik secara keseluruhan parameter.

DAFTAR PUSTAKA

- Alham, S. R. J. I. (2021). Sistem Diagnosis Penyakit Jantung Koroner Dengan Menggunakan Algoritma C4.5 Berbasis Website (Studi Kasus: RSUD Dr. Soedarso Pontianak). *Petir*, *14*(2), 214–222. <https://doi.org/10.33322/petir.v14i2.1338>
- Auble, T. E., Hsieh, M., Gardner, W., Cooper, G. F., Stone, R. A., McCausland, J. B., & Yealy, D. M. (2005). A prediction rule to identify low-risk patients with heart failure. *Academic Emergency Medicine*, *12*(6), 514–521. <https://doi.org/10.1197/j.aem.2004.11.026>
- Ayudhitama, A. P., Pujianto, U., Elektro, T., Teknik, F., Malang, U. N., Tree, D., & Network, N. (2020). Analisa 4 Algoritma Dalam Klasifikasi Penyakit Liver. *Jurnal Informatika Polinema*, *6*, 1–9.
- Azwanti, N., & Elisa, E. (2019). Analisis Pola Penyakit Hipertensi Menggunakan Algoritma C4.5. *InfoTekJar (Jurnal Nasional Informatika Dan Teknologi Jaringan)*, *3*(2), 116–123. <https://doi.org/10.30743/infotekjar.v3i2.944>
- Chang, W., Liu, Y., Xiao, Y., Yuan, X., Xu, X., & Zhang, S. (2019). A Machine-Learning-Based Prediction Method for Hypertension Outcomes Based on Medical Data. *Diagnostics* *2019*, *9*, 178. <https://doi.org/10.3390/diagnostics9040178>
- Chawla, N. V., Kevin W. Bowyer, Hall, L. O., & W. Philip Kegelmeyer. (2002). SMOTE: Synthetic Minority Over-sampling Technique. *Journal of Artificial Intelligence Research*, *3016*(2), 321–357. <https://doi.org/10.1002/eap.2043>
- Chu, H. Y., Huang, H. C., Huang, C. Y., Chu, C. C., Su, C. T., Tsai, I. L., Hu, H. L. S., & Guo, S. L. (2021). A predictive model for identifying low medication adherence among older adults with hypertension: A classification and regression tree model. *Geriatric Nursing*, *42*(6), 1309–1315. <https://doi.org/10.1016/j.gerinurse.2021.08.011>
- Defiyanti, S. and D. L. C. P. (2019). PERBANDINGAN KINERJA ALGORITMA ID3 DAN C4.5 DALAM KLASIFIKASI SPAM-MAIL Sofi. *Gunadarma University Repository*, *9*(2), 62–68.
- Delen, D., Kuzey, C., & Uyar, A. (2013). Measuring firm performance using financial ratios: A decision tree approach. *Expert Systems with Applications*, *40*(10), 3970–3983. <https://doi.org/10.1016/j.eswa.2013.01.012>
- Dewi, P. K. (2015). Analisis Alur Proses Penerimaan Pasien Rawat Inap di Rumah Sakit “X” Tahun 2015 dengan Pendekatan Lean Hospital. *Jurnal Administrasi Rumah Sakit*

- Indonesia*, 2(1), 1–16. <https://doi.org/10.7454/arsi.v2i1.2185>
- Dini, M. Z., Rakhmatsyah, A., & Wardana, A. A. (2022). Detection of Oxygen Levels (SpO₂) and Heart Rate Using a Pulse Oximeter for Classification of Hypoxemia Based on Fuzzy Logic. *Jurnal Ilmiah Teknik Elektro Komputer Dan Informatika*, 8(1), 17. <https://doi.org/10.26555/jiteki.v8i1.22139>
- Dwi Untari, Hastuti, K., Hidayat, E. Y., Dwi Untari, Limão, N., & Gaol, N. Y. L. (2018). Data Mining untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4.5. *Fakultas Ilmu Komputer Universitas Dian Nuswantoro*, 2(November), 31–48.
- Dwiasnati, S., & Devianto, Y. (2019). Optimasi Prediksi Keputusan Calon Nasabah Potensial menggunakan Algoritma C 4.5 berbasis Particle Swarm Optimization. *Jurnal Informatika*, 6(2), 286–292. <https://doi.org/10.31311/ji.v6i2.5771>
- Dyas, A. R., Henderson, W. G., Madsen, H. J., Bronsert, M. R., Colborn, K. L., Lambert-Kerzner, A., McIntyre, R. C., & Meguid, R. A. (2022). Development and validation of a prediction model for conversion of outpatient to inpatient surgery. *Surgery (United States)*, 172(1), 249–256. <https://doi.org/10.1016/j.surg.2022.01.025>
- Erwansyah, K., Andika, B., & Gunawan, R. (2021). Implementasi Data Mining Menggunakan Asosiasi Dengan Algoritma Apriori Untuk Mendapatkan Pola Rekomendasi Belanja Produk Pada Toko Avis Mobile. *J-SISKO TECH (Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD)*, 4(1), 148. <https://doi.org/10.53513/jsk.v4i1.2628>
- Fadlilah, S., Hamdani Rahil, N., & Lanni, F. (2020). Analisis Faktor Yang Mempengaruhi Tekanan Darah Dan Saturasi Oksigen Perifer (Spo₂). *Jurnal Kesehatan Kusuma Husada*, Spo 2, 21–30. <https://doi.org/10.34035/jk.v11i1.408>
- Fernández, A., García, S., Herrera, F., & Chawla, N. V. (2018). SMOTE for Learning from Imbalanced Data: Progress and Challenges, Marking the 15-year Anniversary. *Journal of Artificial Intelligence Research*, 61, 863–905. <https://doi.org/10.1613/jair.1.11192>
- Fiandra, Y. A., Defit, S., & Yuhandri, Y. (2017). Penerapan Algoritma C4.5 untuk Klasifikasi Data Rekam Medis berdasarkan International Classification Diseases (ICD-10). *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 1(2), 82–89. <https://doi.org/10.29207/resti.v1i2.48>
- Fiarni, C., Sipayung, E. M., & Tumundo, P. B. T. (2019). Academic Decision Support System for Choosing Information Systems Sub Majors Programs using Decision Tree Algorithm.

- Journal of Information Systems Engineering and Business Intelligence*, 5(1), 57.
<https://doi.org/10.20473/jisebi.5.1.57-66>
- Gumelar, G., Norlaila, Ain, Q., Marsuciati, R., Bambang, S. A., Sunyoto, A., & Mustafa, M. S. (2021). Kombinasi Algoritma Sampling dengan Algoritma Klasifikasi untuk Meningkatkan Performa Klasifikasi Dataset Imbalance. *Prosiding Seminar Nasional Sistem Informasi Dan Teknologi (SISFOTEK) Ke 5*, 250–255.
- Gupta, B. (2017). *Analysis of Various Decision Tree Algorithms for Classification in Data Mining*. 163(8), 15–19.
- Hamonangan Nasution, M., & Iqbal Sofyan, M. (2020). Penerapan Metode Decision Tree Untuk Memprediksi Prestasi Siswa Kelas XII Dilihat dari Nilai Akhir Semester di SMK Negeri 1 Selong Tahun Pelajaran 2017/2018. *Infotek : Jurnal Informatika Dan Teknologi*, 3(1), 58–65. <https://doi.org/10.29408/jit.v3i1.1825>
- Hartanto, A. S., Basuki, A., & Juli, C. (2019). Correlation of Glasgow Coma Scale Score at Hospital Admission with Stroke Hemorrhagic Patient Mortality at Hasan Sadikin Hospital. *Journal of Medicine and Health*, 2(4), 938–944. <https://doi.org/10.28932/jmh.v2i4.1139>
- Insani, A., Fatimah, S., & Wardani, W. (2022). Senam dan Penyuluhan Kesehatan: Kegiatan Pencegahan Hipertensi Pada Masyarakat Kecamatan Meranti. *Shihatuna: Jurnal Pengabdian Kesehatan Masyarakat*, 2(2), 57–62.
- Ishaq, A., Sadiq, S., Umer, M., Ullah, S., Mirjalili, S., Rupapara, V., & Nappi, M. (2021). Improving the Prediction of Heart Failure Patients' Survival Using SMOTE and Effective Data Mining Techniques. *IEEE Access*, 9, 39707–39716. <https://doi.org/10.1109/ACCESS.2021.3064084>
- Islam, M. M., & Shamsuddin, R. (2021). Machine learning to promote health management through lifestyle changes for hypertension patients. *Array*, 12, 100090. <https://doi.org/10.1016/j.array.2021.100090>
- Jayasri, N. P., & Aruna, R. (2022). Big data analytics in health care by data mining and classification techniques. *ICT Express*, 8(2), 250–257. <https://doi.org/10.1016/j.ict.2021.07.001>
- Jin, C., Li, F., Ma, S., & Wang, Y. (2022). Sampling scheme-based classification rule mining method using decision tree in big data environment. *Knowledge-Based Systems*, 244, 108522. <https://doi.org/10.1016/j.knosys.2022.108522>
- Kamagi, D. H., & Hansun, S. (2016). Implementasi Data Mining dengan Algoritma C4.5 untuk

- Memprediksi Tingkat Kelulusan Mahasiswa. *ULTIMATICS, Vol. VI, No. 1, VI(1)*, 254–260. <https://doi.org/10.1109/EPEPEMC.2016.7752007>
- Kaur, H., Chaudhary, S., Mohanty, S., Sharma, G., Kumaran, S. S., Ghata, N., Bhatia, R., Nehra, A., & Pandey, R. M. (2022). Comparing cognition, coping skills and vedic personality of individuals practicing yoga, physical exercise or sedentary lifestyle: a cross-sectional fMRI study. *Integrative Medicine Research, 11(1)*, 100750. <https://doi.org/10.1016/j.imr.2021.100750>
- Kugelman, D., Huang, S., Teo, G., Doran, M., Singh, V., Buchalter, D., & Long, W. J. (2022). A Novel Machine Learning Predictive Tool Assessing Outpatient or Inpatient Designation for Medicare Patients Undergoing Total Knee Arthroplasty. *Arthroplasty Today, 13*, 120–124. <https://doi.org/10.1016/j.artd.2021.12.004>
- Kugelman, D. N., Teo, G., Huang, S., Doran, M. G., Singh, V., & Long, W. J. (2021). A Novel Machine Learning Predictive Tool Assessing Outpatient or Inpatient Designation for Medicare Patients Undergoing Total Hip Arthroplasty. *Arthroplasty Today, 8*(January 2020), 194–199. <https://doi.org/10.1016/j.artd.2021.03.001>
- Lesmana, I. P. D. (2012). Pengembangan Decision Tree J48 Untuk Diagnosis Penyakit Diabetes Mellitus. *Seminar Nasional Teknologi Informasi & Komunikasi Terapan, 2012*(Semantik), 189–193.
- Lestari, A. (2020). Increasing Accuracy of C4.5 Algorithm Using Information Gain Ratio and Adaboost for Classification of Chronic Kidney Disease. *Journal of Soft Computing Exploration, 1(1)*, 32–38. <https://doi.org/10.52465/josce.v1i1.6>
- Luo, J., Xu, J., Aldosari, O., Althubiti, S. A., & Deebani, W. (2022). Design and Implementation of an Efficient Electronic Bank Management Information System Based Data Warehouse and Data Mining Processing. *Information Processing and Management, 59(6)*, 103086. <https://doi.org/10.1016/j.ipm.2022.103086>
- Mahfud Imam, Gumantan Aditya, F. E. (2020). Analisis Imt (Indeks Massa Tubuh) Atlet Ukm Sepakbola. *Journal Of "Sports Athleticism in Teaching and Recreation on Interdisciplinary Analysis, 3*, 9–13.
- Martinez-Ríos, E., Montesinos, L., Alfaro-Ponce, M., & Pecchia, L. (2021). A review of machine learning in hypertension detection and blood pressure estimation based on clinical and physiological data. *Biomedical Signal Processing and Control, 68*(June), 102813. <https://doi.org/10.1016/j.bspc.2021.102813>

- Mienye, I. D., & Sun, Y. (2021). Performance analysis of cost-sensitive learning methods with application to imbalanced medical data. *Informatics in Medicine Unlocked*, 25, 100690. <https://doi.org/10.1016/j.imu.2021.100690>
- Mitrofanov, S. A., & Semenkin, E. S. (2021). Tree retraining in the decision tree learning algorithm. *IOP Conference Series: Materials Science and Engineering*, 1047(1). <https://doi.org/10.1088/1757-899X/1047/1/012082>
- Mohammed, A. J. (2020). Improving Classification Performance for a Novel Imbalanced Medical Dataset using SMOTE Method. *International Journal of Advanced Trends in Computer Science and Engineering*, 9(3), 3161–3172. <https://doi.org/10.30534/ijatcse/2020/104932020>
- Mubaroroh, H. H., Yasin, H., & Rusgiyono, A. (2022). Analisis Sentimen Data Ulasan Aplikasi Ruangguru Pada Situs Google Play Menggunakan Algoritma Naïve Bayes Classifier Dengan Normalisasi Kata Levenshtein Distance. *Jurnal Gaussian*, 11(2), 248–257. <https://doi.org/10.14710/j.gauss.v11i2.35472>
- Munthe, I. R., & Sihombing, V. (2018). Klasifikasi Algoritma Iterative Dichotomizer (ID3) untuk Tingkat kepuasan pada Sarana Laboratorium Komputer. *Jurnal Teknologi Dan Ilmu Komputer Prima (JUTIKOMP)*, 1(2), 27–34. <https://doi.org/10.34012/jutikomp.v1i2.237>
- Mutmainah, S. (2021). Penanganan Imbalance Data Pada Klasifikasi. *SNATi*, 1, 10–16.
- Nejadgholi, I., Sadreazami, H., Rajan, S., & Bolic, M. (2019). Classification of doppler radar reflections as preprocessing for breathing rate monitoring. *IET Signal Processing*, 13(1), 21–28. <https://doi.org/10.1049/iet-spr.2018.5245>
- Ningsih, S. R., Damanik, I. S., Windarto, A. P., Tambunan, H. S., Jalaluddin, J., & Wanto, A. (2019). Analisis K-Medoids Dalam Pengelompokan Penduduk Buta Huruf Menurut Provinsi. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 1(September), 721. <https://doi.org/10.30645/senaris.v1i0.78>
- Nisa', C., Puspaningrum, E. Y., & Maulana, H. (2020). Penerapan Metode Convolutional Neural Network untuk Klasifikasi Penyakit Daun Apel pada Imbalanced Data. *Prosiding Seminar Nasional Informatika Bela Negara*, 1, 169–175. <https://doi.org/10.33005/santika.v1i0.46>
- Nogay, N. H., & Nogay, H. S. (2022). Estimation of Survival According to Body Mass Index (BMI), Hypertension , Diabetes and Heart Disease with Optimizable Decision Trees. *10(1)*, 47–52. <https://doi.org/10.17694/bajece.1037645>

- Nour, M., & Polat, K. (2020). Automatic Classification of Hypertension Types Based on Personal Features by Machine Learning Algorithms. *Mathematical Problems in Engineering*, 2020, 1–14. <https://doi.org/10.1155/2020/2742781>
- Nurhopipah, A., & Hasanah, U. (2020). Dataset Splitting Techniques Comparison For Face Classification on CCTV Images. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, 14(4), 341. <https://doi.org/10.22146/ijccs.58092>
- Nurlia, E., & Enri, U. (2021). Penerapan Fitur Seleksi Forward Selection Untuk Menentukan Kematian Akibat Gagal Jantung Menggunakan Algoritma C4.5. *Jurnal Teknik Informatika Musirawas) Elin Nurlia*, 6(1), 42.
- Nurmasani, A., & Pristyanto, Y. (2021). Algoritme Stacking Untuk Klasifikasi Penyakit Jantung Pada Dataset Imbalanced Class. *Pseudocode*, 8(1), 21–26. <https://doi.org/10.33369/pseudocode.8.1.21-26>
- Ozcan, M., & Peker, S. (2023). *Healthcare Analytics A classification and regression tree algorithm for heart disease modeling and prediction*. 3(December 2022). <https://doi.org/10.1016/j.health.2022.100130>
- Pambudi, R. E., Sriyanto, S., & Firmansyah, F. (2022). Klasifikasi Penyakit Stroke Menggunakan Algoritma Decision Tree C.45. *Teknika*, 16(2), 221 – 226–221 – 226. <https://jurnal.polsri.ac.id/index.php/teknika/article/view/4914>
- Pamuji, F. Y., & Ramadhan, V. P. (2021). Komparasi Algoritma Random Forest dan Decision Tree untuk Memprediksi Keberhasilan Immunotherapy. *Jurnal Teknologi Dan Manajemen Informatika*, 7(1), 46–50. <https://doi.org/10.26905/jtmi.v7i1.5982>
- Quesado, I., Duarte, J., Silva, Á., Manuel, M., & Quintas, C. (2022). Data Mining Models for Automatic Problem Identification in Intensive Medicine. *Procedia Computer Science*, 210, 218–223. <https://doi.org/10.1016/j.procs.2022.10.140>
- Rafsunjani, S., Safa, R. S., Imran, A. Al, Rahim, S., & Nandi, D. (2019). An Empirical Comparison of Missing Value Imputation Techniques on APS Failure Prediction. *International Journal of Information Technology and Computer Science*, 11(2), 21–29. <https://doi.org/10.5815/ijitcs.2019.02.03>
- Ratniasih, N. L. (2016). Konversi Data Training Tentang Pemilihan Kelas. *Eksplora Informatika*, 145–154.
- Reskiaddin, L. O., Anhar, V. Y., Sholikhah, & Wartono. (2020). Challenges and Barrier on Community Empowerment in Communicable Disease on Semi Rural Area : A Evidence

- Based Practice in Padukuhan Samirono, Sleman Yogyakarta. *Jurnal Kesmas Jambi (JKMJ)*, 4(2), 43–49.
- Richa Rachmawati, Y. Y., Ayu Sanjaya, Y. P., & Edilia, S. (2022). Web-Based Temperature, Oxygen Saturation, and Heart Rate Monitoring System. *IAIC Transactions on Sustainable Digital Innovation (ITSDI)*, 4(1), 38–45. <https://doi.org/10.34306/itsdi.v4i1.567>
- Sanjeev, C., Dash, K., Kumar, A., Dehuri, S., & Ghosh, A. (2023). An outliers detection and elimination framework in classification task of data mining. *Decision Analytics Journal*, 6(January), 100164. <https://doi.org/10.1016/j.dajour.2023.100164>
- Santoni, M. M., Chamidah, N., & Matondang, N. (2020). Prediksi Hipertensi menggunakan Decision Tree, Naïve Bayes dan Artificial Neural Network pada software KNIME. *Techno.Com*, 19(4), 353–363. <https://doi.org/10.33633/tc.v19i4.3872>
- Santos, L. I., Camargos, M. O., D'Angelo, M. F. S. V., Mendes, J. B., Medeiros, E. E. C. de, Guimarães, A. L. S., & Palhares, R. M. (2022). Decision tree and artificial immune systems for stroke prediction in imbalanced data. *Expert Systems with Applications*, 191(June 2021), 116221. <https://doi.org/10.1016/j.eswa.2021.116221>
- Saputri, R. P. (2018). *Penerapan Data Mining Technique sebagai Evaluasi Ketepatan Akurasi terhadap Klasifikasi Mushroom Data Set Pendahuluan Metodologi Penelitian*. 06211540000040.
- Saputro, E., & Rosiyadi, D. (2022). Penerapan Metode Random Over-Under Sampling Pada Algoritma Klasifikasi Penentuan Penyakit Diabetes. *Bianglala Informatika*, 10(1), 42–47. <https://doi.org/10.31294/bi.v10i1.11739>
- Sari Ikhsan, L. (2019). Rancang Bangun Alat Ukur Frekuensi Pernapasan Manusia Berbasis Sensor Serat Optik. *Jurnal Fisika Unand*, 8(4), 301–307. <http://jfu.fmipa.unand.ac.id/index.php/jfu/article/view/429>
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Sinam, I. I., & Lawan, A. (2019). An improved C4.5 model classification algorithm based on Taylor's series. *Jordanian Journal of Computers and Information Technology*, 5(1), 34–42. <https://doi.org/10.5455/jjcit.71-1546551963>
- Sir, Y. A., & Soepranoto, A. H. H. (2022). Pendekatan Resampling Data Untuk Menangani Masalah Ketidakseimbangan Kelas. *Jurnal Komputer Dan Informatika*, 10(1), 31–38.

<https://doi.org/10.35508/jicon.v10i1.6554>

- Sofyan, S., & Prasetyo, A. (2021). Penerapan Synthetic Minority Oversampling Technique (SMOTE) Terhadap Data Tidak Seimbang Pada Tingkat Pendapatan Pekerja Informal Di Provinsi D.I. Yogyakarta Tahun 2019. *Seminar Nasional Official Statistics, 2021*(1), 868–877. <https://doi.org/10.34123/semnasoffstat.v2021i1.1081>
- Sun, J., Lang, J., Fujita, H., & Li, H. (2018). Imbalanced enterprise credit evaluation with DTE-SBD: Decision tree ensemble based on SMOTE and bagging with differentiated sampling rates. *Information Sciences, 425*, 76–91. <https://doi.org/10.1016/j.ins.2017.10.017>
- Tayefi, M., Esmaeili, H., Saberi Karimian, M., Amirabadi Zadeh, A., Ebrahimi, M., Safarian, M., Nematy, M., Parizadeh, S. M. R., Ferns, G. A., & Ghayour-Mobarhan, M. (2017). The application of a decision tree to establish the parameters associated with hypertension. *Computer Methods and Programs in Biomedicine, 139*, 83–91. <https://doi.org/10.1016/j.cmpb.2016.10.020>
- Truesdale, K. P., & Stevens, J. (2008). Do the Obese Know They Are Obese? *North Carolina Medical Journal, 69*(3), 188–194.
- Wang, L., Porter, B., Maynard, C., Bryson, C., Sun, H., Lowy, E., McDonnell, M., Frisbee, K., Nielson, C., & Fihn, S. D. (2012). Predicting risk of hospitalization or death among patients with heart failure in the veterans health administration. *American Journal of Cardiology, 110*(9), 1342–1349. <https://doi.org/10.1016/j.amjcard.2012.06.038>
- Wijayanti, N. P. Y. T., Eka N Kencana, & I Wayan Sumarjaya. (2021). Smote: Potensi Dan Kekurangannya Pada Survei. *E-Jurnal Matematika, 10*(4), 235. <https://doi.org/10.24843/mtk.2021.v10.i04.p348>
- Wu, J. H., Kao, H. Y., & Sambamurthy, V. (2016). The integration effort and E-health compatibility effect and the mediating role of E-health synergy on hospital performance. *International Journal of Information Management, 36*(6), 1288–1300. <https://doi.org/10.1016/j.ijinfomgt.2016.09.002>
- Yamamoto, T., Todani, M., Oda, Y., Kaneko, T., Kaneda, K., Fujita, M., Miyauchi, T., & Tsuruta, R. (2015). Predictive factors for hospitalization of patients with heat illness in Yamaguchi, Japan. *International Journal of Environmental Research and Public Health, 12*(9), 11770–11780. <https://doi.org/10.3390/ijerph120911770>
- Yang, J., Li, Y., Liu, Q., Li, L., Feng, A., Wang, T., Zheng, S., Xu, A., & Lyu, J. (2020). Brief introduction of medical database and data mining technology in big data era. *Journal of*

- Evidence-Based Medicine*, 13(1), 57–69. <https://doi.org/10.1111/jebm.12373>
- Yuli Mardi. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika*. *Jurnal Edik Informatika*, 2(2), 213–219.
- Yuniarti, H., Sigit, R., & Rofiq, M. A. (2020). Penerapan Fuzzy Tsukamoto pada Alat Deteksi Penyakit Hipoksemia, Hipotermia, Hipertensi, dan Diabetes untuk Health Care Kiosk. *Journal of Applied Informatics and Computing*, 4(2), 163–173. <https://doi.org/10.30871/jaic.v4i2.2643>
- Zhou, B., Carrillo-Larco, R. M., Danaei, G., Riley, L. M., Paciorek, C. J., Stevens, G. A., Gregg, E. W., Bennett, J. E., Solomon, B., Singleton, R. K., Sophiea, M. K., Iurilli, M. L. C., Lhoste, V. P. F., Cowan, M. J., Savin, S., Woodward, M., Balanova, Y., Cifkova, R., Damasceno, A., ... Zuñiga Cisneros, J. (2021). Worldwide trends in hypertension prevalence and progress in treatment and control from 1990 to 2019: a pooled analysis of 1201 population-representative studies with 104 million participants. *The Lancet*, 398(10304), 957–980. [https://doi.org/10.1016/S0140-6736\(21\)01330-1](https://doi.org/10.1016/S0140-6736(21)01330-1)
- Zhou, B., Perel, P., Mensah, G. A., & Ezzati, M. (2021). Global epidemiology, health burden and effective interventions for elevated blood pressure and hypertension. *Nature Reviews Cardiology*, 18(11), 785–802. <https://doi.org/10.1038/s41569-021-00559-8>
- Zulfa, I. M., Rahmawati, Y. A., & Anggraini, P. F. (2022). Potensi Interaksi Antar Obat dalam Persepan Rawat Jalan Pasien Penyakit Jantung Akibat Hipertensi. *Jurnal Farmasi Indonesia*, 19(1), 90–97. <https://doi.org/10.31001/jfi.v19i1.1253>

LAMPIRAN

Coding Pre-processing

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

In [2]: dfid = pd.read_excel('id_ranap_new.xlsx', sheet_name = 'ranap')

In [3]: dfid

Out[3]:
```

	ID	Umur	Jenis Kelamin	Kode
0	19513	66 Th	Perempuan	I11
1	12242	90 Th	Laki-Laki	I11
2	74099	68 Th	Perempuan	I11
3	74409	80 Th	Perempuan	I11
4	75575	46 Th	Laki-Laki	I11
5	75769	50 Th	Perempuan	I11
6	6503	54 Th	Laki-Laki	I11

```
In [4]: df = pd.read_excel('ranap_new.xlsx')

In [5]: df

Out[5]:
```

	ID	Suhu(C)	Systolic	Dyastolic	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(Cm)	Berat(Kg)	SpO2	GCS(E,V,M)	Kesadaran
0	59349	36.4	NaN	NaN	110.0	24.0	NaN	10.6	99.0	15	Compos Mentis
1	59349	36.4	NaN	NaN	110.0	24.0	NaN	10.6	99.0	15	Compos Mentis
2	59349	36.4	NaN	NaN	110.0	24.0	NaN	10.6	99.0	15	Compos Mentis
3	25491	36.4	170	100	88.0	21.0	156	75	96.0	15	Compos Mentis
4	25491	36.4	170	100	88.0	21.0	NaN	NaN	96.0	15	Compos Mentis

```
In [6]: result = df.merge(dfid, on='ID', how='inner')

In [7]: result

Out[7]:
```

	ID	Suhu(C)	Systolic	Dyastolic	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(Cm)	Berat(Kg)	SpO2	GCS(E,V,M)	Kesadaran	Umur	Jenis Kelamin	Kode
0	6503	37.7	130	90	80.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
1	6503	37.7	120	90	80.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
2	6503	37.7	120	90	80.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
3	6503	36.9	125	66	80.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
4	6503	36.7	120	70	82.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
5	6503	36.7	110	70	82.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
6	6503	36.8	110	80	82.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
7	6503	36.4	120	80	80.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
8	6503	36.5	140	90	76.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
9	6503	36.6	113	90	78.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
10	75769	36.5	170	100	82.0	20.0	NaN	NaN	99.0	15	Compos Mentis	50 Th	Perempuan	I11
11	75769	36.9	140	90	82.0	21.0	NaN	NaN	99.0	15	Compos Mentis	50 Th	Perempuan	I11
12	75769	36.4	140	90	80.0	20.0	NaN	NaN	99.0	15	Compos Mentis	50 Th	Perempuan	I11
13	75769	36.4	140	90	80.0	20.0	NaN	NaN	99.0	15	Compos Mentis	50 Th	Perempuan	I11
14	75769	36.5	96	63	63.0	20.0	NaN	NaN	99.0	15	Compos Mentis	50 Th	Perempuan	I11
15	75769	36.9	96	63	80.0	21.0	NaN	NaN	99.0	15	Compos Mentis	50 Th	Perempuan	I11
16	75575	36.6	190	100	80.0	20.0	NaN	NaN	99.0	15	Compos Mentis	46 Th	Laki-Laki	I11
17	75575	36.6	110	90	80.0	20.0	NaN	NaN	99.0	15	Compos Mentis	46 Th	Laki-Laki	I11
18	75575	36.6	110	80	80.0	20.0	NaN	NaN	99.0	15	Compos Mentis	46 Th	Laki-Laki	I11
19	74409	37.1	194	89	97.0	20.0	-	-	98.0	15	Compos Mentis	80 Th	Perempuan	I11
20	74409	37	189	85	90.0	20.0	-	-	99.0	15	Compos Mentis	80 Th	Perempuan	I11

```
In [8]: result = result.drop_duplicates(subset=['ID'])
```

```
In [9]: result
```

	ID	Suhu(C)	Systolic	Dyastolic	Nadi/(mnt)	Respirasi/(mnt)	Tinggi(Cm)	Berat(Kg)	SpO2	GCS(E,V,M)	Kesadaran	Umur	Jenis Kelamin	Kode
0	6503	37.7	130	90	80.0	20.0	NaN	NaN	98.0	15	Compos Mentis	54 Th	Laki-Laki	I11
10	75769	36.5	170	100	82.0	20.0	NaN	NaN	99.0	15	Compos Mentis	50 Th	Perempuan	I11
16	75575	36.6	190	100	80.0	20.0	NaN	NaN	99.0	15	Compos Mentis	46 Th	Laki-Laki	I11
19	74409	37.1	194	89	97.0	20.0	-	-	98.0	15	Compos Mentis	80 Th	Perempuan	I11
25	74099	38	140	70	80.0	20.0	NaN	NaN	99.0	15	Compos Mentis	68 Th	Perempuan	I11
33	12242	36.5	130	90	80.0	22.0	150	60	99.0	15	Compos Mentis	90 Th	Laki-Laki	I11
49	19513	36.5	153	70	66.0	21.0	150	48	NaN	e4v5m6	Compos Mentis	66 Th	Perempuan	I11

```
In [10]: result['Umur'] = result['Umur'].str.replace("Th","")
result['Umur'] = result['Umur'].astype("float")

C:\Users\ASUS\AppData\Local\Temp\ipykernel_2800\1309687125.py:1: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
result['Umur'] = result['Umur'].str.replace("Th","")
C:\Users\ASUS\AppData\Local\Temp\ipykernel_2800\1309687125.py:2: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy
result['Umur'] = result['Umur'].astype('float')
```

```
In [11]: result = result.drop('GCS(E,V,M)', axis=1)
```

```
In [12]: result=result.reindex(columns=['ID','Jenis Kelamin','Kode','Umur','Suhu(C)','Systolic','Dyastolic','Nadi/(mnt)','Respirasi/(mnt)','Tinggi(Cm)','Berat(Kg)','SpO2','Kesadaran'])
```

```
In [13]: result
```

```
In [14]: df1 = pd.read_excel('rajal jan-mar.xlsx',sheet_name = 'Final Jan-mar')
df2 = pd.read_excel('rajal apr-jun.xlsx',sheet_name = 'Rekap Final')
df3 = pd.read_excel('rajal jul-sep.xlsx',sheet_name = 'csv rawat jalan Jul-Sep')
df4 = pd.read_excel('rajal okt-des.xlsx',sheet_name = 'Final')
```

```
In [15]: rekap = [df1, df2, df3, df4]
df_rekap = pd.concat(rekap)
df_rekap
```

	ID	Jenis Kelamin	Umur	Kode
0	44656.0	Laki-Laki	31 Th	I11
1	65588.0	Laki-Laki	69 Th	I11
2	23941.0	Laki-Laki	70 Th	I11
3	1715.0	Laki-Laki	62 Th	I11
4	15154.0	Perempuan	72 Th	I11
...
118	45975.0	Perempuan	65 Th	I11
119	77124.0	Laki-Laki	39 Th	I11
120	36448.0	Laki-Laki	68 Th	I11
121	50650.0	Laki-Laki	37 Th	I11
122	60008.0	Perempuan	85 Th	I11

501 rows x 4 columns

```
In [16]: df_rekap['Kode'].unique()
Out[16]: array(['I11'], dtype=object)
```

```
In [17]: df_periksa1 = pd.read_excel('jan mar.xlsx')
df_periksa2 = pd.read_excel('apr mei.xlsx')
df_periksa3 = pd.read_excel('jun des.xlsx')
```

```
In [17]: df_periksa1 = pd.read_excel('jan mar.xlsx')
df_periksa2 = pd.read_excel('apr mei.xlsx')
df_periksa3 = pd.read_excel('jun des.xlsx')
```

```
In [18]: rekap_periksa = [df_periksa1, df_periksa2, df_periksa3]
df_periksa = pd.concat(rekap_periksa)
```

```
In [19]: df_periksa
```

	ID	Nama Pasien	Tgl.Rawat	Jam	Suhu(C)	Sistolik	Diastolik	Nadi/(mnt)	Respirasi/(mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat
0	1028.0	RINDA HAYANI	19/10/2022	07:36:03	36.7	120	70	89.0	20.0	1.59	55	99.0	Compos Mentis	Rawat Jalan
1	74488.0	H. AMALLUDDIN S. S.Ag	19/10/2022	08:39:19	36.5	120	80	80.0	20.0	NaN	NaN	99.0	Compos Mentis	Rawat Jalan
2	53997.0	RAZIQU HANAN ILVAS	19/10/2022	17:37:48	36.6	NaN	NaN	20.0	99.0	1.10	20	99.0	Compos Mentis	Rawat Jalan
3	58951.0	KAIDA BINTI DAVAT	19/10/2022	14:40:26	36	90	70	90.0	20.0	NaN	NaN	89.0	Compos Mentis	Rawat Jalan
4	20509.0	YULIAN EFENDI	2022-08-02 00:00:00	17:25:18	36	160	90	90.0	18.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
...
19292	41054.0	SINGGI JULIAWAN	2022-01-06 00:00:00	00:51:36	36.7	120	80	87.0	20.0	1.65	50	NaN	Compos Mentis	Rawat Jalan
19293	41342.0	METRI HANITA	2022-01-06 00:00:00	00:10:37	36.7	130	80	80.0	20.0	1.50	62	NaN	Compos Mentis	Rawat Jalan
19294	69955.0	FITRIA RAHMADHANI	2022-02-06 00:00:00	17:01:24	36	NaN	NaN	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
19295	8977.0	BERNADETA SAMINEM	2022-02-06 00:00:00	17:15:45	36	140	80	90.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
19296	17537.0	SITUSMAWATI BINTI BUYUNG DRAHMAN	2022-03-06 00:00:00	16:21:10	36	110	70	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan

63932 rows x 14 columns

```
In [20]: result_rajal = df_periksa.merge(df_rekap, on='ID', how='inner')
```

```
In [21]: result_rajal
```

	ID	Nama Pasien	Tgl.Rawat	Jam	Suhu(C)	Sistolik	Diastolik	Nadi/(mnt)	Respirasi/(mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat	Jenis Kelamin	Umur	Kode
0	20509.0	YULIAN EFENDI	2022-08-02 00:00:00	17:25:18	36	160	90	90.0	18.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan	Laki-Laki	59 Th	I11
4	20509.0	YULIAN EFENDI	2022-08-02 00:00:00	17:25:18	36	160	90	90.0	18.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan	Laki-Laki	60 Th	I11

```
In [22]: result_rajal["Umur"] = result_rajal["Umur"].str.replace("Th","")
result_rajal["Umur"] = result_rajal["Umur"].astype('float')
```

```
In [23]: result_rajal = result_rajal.drop_duplicates(subset=['ID', 'Nama Pasien'])
```

```
In [24]: result_rajal
```

Out[24]:

ID	Nama Pasien	Tgl.Rawat	Jam	Suhu(C)	Sistolik	Diastolik	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat	Jenis Kelamin	Umur	Kode	
0	20509.0	YULIAN EFENDI	2022-08-02 00:00:00	17:25:18	36	160	90	90.0	18.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan	Laki-Laki	59.0	I11
17	76432.0	TAUFIK JAMBAK	26/01/2023	00:40:19	36.5	150	90	87.0	24.0	1.65	80	99.0	Compos Mentis	Rawat Jalan	Laki-Laki	49.0	I11
31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
33	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1137	NaN	SARI	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
...
19153	57759.0	HAMKA KUPSIN	2022-10-02 00:00:00	16:03:18	36.5	120	80	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan	Laki-Laki	61.0	I11
19155	67061.0	NARMAWATI	2022-10-02 00:00:00	09:53:22	36.5	120	70	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan	Perempuan	36.0	I11
19157	65588.0	M. YAMIN	2022-03-02 00:00:00	14:55:04	36	110	70	82.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan	Laki-Laki	69.0	I11
19161	14434.0	MUHAMMAD YAMIN	27/01/2022	15:54:44	35.4	160	80	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan	Laki-Laki	57.0	I11
19164	23941.0	AZWART	23/01/2022	16:09:16	38.4	130	80	98.0	20.0	1.60	50	NaN	Compos Mentis	Rawat Jalan	Laki-Laki	70.0	I11

168 rows x 17 columns

```
In [25]: result_rajal = result_rajal.drop(columns=['Nama Pasien', 'Tgl.Rawat', 'Jam'])
```

```
In [26]: result_rajal
```

Out[26]:

ID	Suhu(C)	Sistolik	Diastolik	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat	Jenis Kelamin	Umur	Kode	
0	20509.0	36	160	90	90.0	18.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan	Laki-Laki	59.0	I11
17	76432.0	36.5	150	90	87.0	24.0	1.65	80	99.0	Compos Mentis	Rawat Jalan	Laki-Laki	49.0	I11
31	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

```
In [27]: result_rajal=result_rajal.reindex(columns=['ID','Jenis Kelamin','Kode','Umur','Suhu(C)','Sistolik','Diastolik','Nadi(/mnt)','Respirasi(/mnt)','Tinggi(m)',
'Berat(Kg)','SpO2','Kesadaran','Jenis Rawat'])
```

```
In [28]: result_rajal
```

Out[28]:

ID	Jenis Kelamin	Kode	Umur	Suhu(C)	Sistolik	Diastolik	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat	
0	20509.0	Laki-Laki	I11	59.0	36	160	90	90.0	18.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
17	76432.0	Laki-Laki	I11	49.0	36.5	150	90	87.0	24.0	1.65	80	99.0	Compos Mentis	Rawat Jalan
31	NaN	Perempuan	I11	60.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Rawat Jalan
33	NaN	Perempuan	I11	60.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Rawat Jalan
1137	NaN	Perempuan	I11	60.0	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	Rawat Jalan
...
19153	57759.0	Laki-Laki	I11	61.0	36.5	120	80	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
19155	67061.0	Perempuan	I11	36.0	36.5	120	70	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
19157	65588.0	Laki-Laki	I11	69.0	36	110	70	82.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
19161	14434.0	Laki-Laki	I11	57.0	35.4	160	80	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
19164	23941.0	Laki-Laki	I11	70.0	38.4	130	80	98.0	20.0	1.60	50	NaN	Compos Mentis	Rawat Jalan

168 rows x 14 columns

```
In [30]: df_ranap = pd.read_excel('result_ranap.xlsx')
```

```
In [31]: df_ranap
```

Out[31]:

ID	Jenis Kelamin	Kode	Umur	Suhu(C)	Sistolik	Diastolik	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat	
0	6503	Laki-Laki	I11	54	37.7	145	92	20	NaN	NaN	90.0	Somnolence	Rawat Inap	
1	75769	Perempuan	I11	50	36.5	170	100	82	20	NaN	NaN	80.0	Somnolence	Rawat Inap
2	75575	Laki-Laki	I11	46	36.6	190	100	58	20	NaN	NaN	99.0	Sopor	Rawat Inap
3	74409	Perempuan	I11	80	37.1	194	94	97	20	NaN	NaN	82.0	Somnolence	Rawat Inap
4	74099	Perempuan	I11	68	38.0	148	96	80	20	NaN	NaN	99.0	Somnolence	Rawat Inap
5	12242	Laki-Laki	I11	90	36.5	189	84	80	22	1.5	60.0	76.0	Sopor	Rawat Inap
6	19513	Perempuan	I11	66	36.5	166	90	55	21	1.5	48.0	NaN	Somnolence	Rawat Inap

```
In [32]: df_ranap = df_ranap.rename({'Systolic':'Sistolik', 'Dyastolic':'Diastolik'}, axis=1)
```

```
In [33]: df_ranap
```

Out[33]:

ID	Jenis Kelamin	Kode	Umur	Suhu(C)	Sistolik	Diastolik	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat	
0	6503	Laki-Laki	I11	54	37.7	145	92	20	NaN	NaN	90.0	Somnolence	Rawat Inap	
1	75769	Perempuan	I11	50	36.5	170	100	82	20	NaN	NaN	80.0	Somnolence	Rawat Inap
2	75575	Laki-Laki	I11	46	36.6	190	100	58	20	NaN	NaN	99.0	Sopor	Rawat Inap
3	74409	Perempuan	I11	80	37.1	194	94	97	20	NaN	NaN	82.0	Somnolence	Rawat Inap
4	74099	Perempuan	I11	68	38.0	148	96	80	20	NaN	NaN	99.0	Somnolence	Rawat Inap
5	12242	Laki-Laki	I11	90	36.5	189	84	80	22	1.5	60.0	76.0	Sopor	Rawat Inap
6	19513	Perempuan	I11	66	36.5	166	90	55	21	1.5	48.0	NaN	Somnolence	Rawat Inap

```
In [34]: final = [df_ranap, result_rajal]
```

```
In [35]: final = pd.concat(final)
```

```
In [36]: final.info()
<class 'pandas.core.frame.DataFrame'>
Int64Index: 175 entries, 0 to 19164
Data columns (total 14 columns):
# Column      Non-Null Count  Dtype
---  ---
0 ID           127 non-null    float64
1 Jenis Kelamin 175 non-null    object
2 Kode         175 non-null    object
3 Umur         175 non-null    float64
4 Suhu(C)      124 non-null    object
5 Sistolik     127 non-null    object
6 Diastolik    127 non-null    object
7 Nadi(/mnt)   127 non-null    float64
8 Respirasi(/mnt) 127 non-null    float64
9 Tinggi(m)    29 non-null     float64
10 Berat(Kg)   34 non-null     object
11 SpO2        105 non-null    float64
12 Kesadaran   127 non-null    object
13 Jenis Rawat 174 non-null    object
dtypes: float64(6), object(8)
memory usage: 20.5+ KB

In [37]: final
Out[37]:
```

	ID	Jenis Kelamin	Kode	Umur	Suhu(C)	Sistolik	Diastolik	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat
0	6503.0	Laki-Laki	111	54.0	37.7	145	92	80.0	20.0	NaN	NaN	90.0	Somnolence	Rawat Inap
1	75769.0	Perempuan	111	50.0	36.5	170	100	82.0	20.0	NaN	NaN	80.0	Somnolence	Rawat Inap
2	75575.0	Laki-Laki	111	46.0	36.6	190	100	58.0	20.0	NaN	NaN	99.0	Sopor	Rawat Inap
3	74409.0	Perempuan	111	80.0	37.1	194	94	97.0	20.0	NaN	NaN	82.0	Somnolence	Rawat Inap
4	74099.0	Perempuan	111	68.0	38.0	148	96	80.0	20.0	NaN	NaN	99.0	Somnolence	Rawat Inap
...
19153	57759.0	Laki-Laki	111	61.0	36.5	120	80	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
19155	67061.0	Perempuan	111	36.0	36.5	120	70	80.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan
19157	65588.0	Laki-Laki	111	69.0	36	110	70	82.0	20.0	NaN	NaN	NaN	Compos Mentis	Rawat Jalan

Handling Missing value

```
In [38]: final = pd.read_excel('Final TA.xlsx')

In [39]: from sklearn.impute import SimpleImputer
imputer = SimpleImputer(missing_values = np.nan, strategy = 'mean')
imputer.fit(final.iloc[:, 3:12])
final.iloc[:, 3:12] = imputer.transform(final.iloc[:, 3:12])

In [40]: final
Out[40]:
```

	ID	Jenis Kelamin	Kode	Umur	Suhu(C)	Sistolik	Diastolik	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat
0	6503	Laki-Laki	111	54.0	37.7	145.0	92.0	80.0	20.0	1.593929	66.382812	90.000000	Somnolence	Rawat Inap
1	75769	Perempuan	111	50.0	36.5	170.0	100.0	82.0	20.0	1.593929	66.382812	80.000000	Somnolence	Rawat Inap
2	75575	Laki-Laki	111	46.0	36.6	190.0	100.0	58.0	20.0	1.593929	66.382812	99.000000	Sopor	Rawat Inap
3	74409	Perempuan	111	80.0	37.1	200.0	94.0	97.0	20.0	1.593929	66.382812	82.000000	Somnolence	Rawat Inap
4	74099	Perempuan	111	68.0	38.0	148.0	96.0	80.0	20.0	1.593929	66.382812	99.000000	Somnolence	Rawat Inap
...
832	14434	Laki-Laki	111	57.0	36.0	130.0	70.0	88.0	20.0	1.593929	66.382812	98.284658	Compos Mentis	Rawat Jalan
833	14434	Laki-Laki	111	57.0	36.0	130.0	70.0	87.0	20.0	1.593929	66.382812	98.284658	Compos Mentis	Rawat Jalan
834	23941	Laki-Laki	111	70.0	38.4	130.0	80.0	98.0	20.0	1.600000	50.000000	98.284658	Compos Mentis	Rawat Jalan
835	23941	Laki-Laki	111	70.0	36.0	130.0	70.0	87.0	20.0	1.593929	66.382812	98.284658	Compos Mentis	Rawat Jalan
836	23941	Laki-Laki	111	70.0	36.0	110.0	70.0	87.0	20.0	1.593929	66.382812	98.284658	Compos Mentis	Rawat Jalan

837 rows × 14 columns

```
In [41]: #Perhitungan BMI
def BMI(final):
    df = pd.DataFrame(final)
    return final['Berat(Kg)']/final['Tinggi(m)']**2
final['BMI'] = final.apply(BMI, axis = 1)
final

Out[41]:
```

	ID	Jenis Kelamin	Kode	Umur	Suhu(C)	Sistolik	Diastolik	Nadi(/mnt)	Respirasi(/mnt)	Tinggi(m)	Berat(Kg)	SpO2	Kesadaran	Jenis Rawat	BMI
0	6503	Laki-Laki	111	54.0	37.7	145.0	92.0	80.0	20.0	1.593929	66.382812	90.000000	Somnolence	Rawat Inap	26.128708
1	75769	Perempuan	111	50.0	36.5	170.0	100.0	82.0	20.0	1.593929	66.382812	80.000000	Somnolence	Rawat Inap	26.128708
2	75575	Laki-Laki	111	46.0	36.6	190.0	100.0	58.0	20.0	1.593929	66.382812	99.000000	Sopor	Rawat Inap	26.128708
3	74409	Perempuan	111	80.0	37.1	200.0	94.0	97.0	20.0	1.593929	66.382812	82.000000	Somnolence	Rawat Inap	26.128708
4	74099	Perempuan	111	68.0	38.0	148.0	96.0	80.0	20.0	1.593929	66.382812	99.000000	Somnolence	Rawat Inap	26.128708
...
832	14434	Laki-Laki	111	57.0	36.0	130.0	70.0	88.0	20.0	1.593929	66.382812	98.284658	Compos Mentis	Rawat Jalan	26.128708
833	14434	Laki-Laki	111	57.0	36.0	130.0	70.0	87.0	20.0	1.593929	66.382812	98.284658	Compos Mentis	Rawat Jalan	26.128708
834	23941	Laki-Laki	111	70.0	38.4	130.0	80.0	98.0	20.0	1.600000	50.000000	98.284658	Compos Mentis	Rawat Jalan	19.531250
835	23941	Laki-Laki	111	70.0	36.0	130.0	70.0	87.0	20.0	1.593929	66.382812	98.284658	Compos Mentis	Rawat Jalan	26.128708
836	23941	Laki-Laki	111	70.0	36.0	110.0	70.0	87.0	20.0	1.593929	66.382812	98.284658	Compos Mentis	Rawat Jalan	26.128708

837 rows × 15 columns

```
In [42]: # kategorisasi Denyut nadi
Level_nadi = [
    (final['Nadi(/mnt)'] <=60),
    (final['Nadi(/mnt)'] > 60) & (final['Nadi(/mnt)'] <=100),
    (final['Nadi(/mnt)'] >100),
]
values = ['Bradikardia', 'Normal', 'Takikardia',]
final['Level Nadi'] = np.select(Level_nadi, values)
```

```

In [43]: #Kategorisasi suhu
Level_suhu = [
    (final['Suhu(C)'] < 36),
    (final['Suhu(C)'] >= 36 & (final['Suhu(C)'] <=37.7),
    (final['Suhu(C)'] >37.7) & (final['Suhu(C)'] <= 40),
    (final['Suhu(C)'] > 40),
    ]
values = ['Hipotermia', 'Normal', 'Demam', 'Hipereksia']
final['Level Suhu'] = np.select(Level_suhu, values)

In [44]: #Kategorisasi SpO2
Level_SpO2 = [
    (final['SpO2'] >= 95 & (final['SpO2'] <= 100),
    (final['SpO2'] >= 91 & (final['SpO2'] <=94),
    (final['SpO2'] >=86 & (final['SpO2'] <= 90),
    (final['SpO2'] <=85),
    ]
values = ['Normal', 'Cukup', 'Sedang', 'Rendah']
final['Level SpO2'] = np.select(Level_SpO2, values)

In [45]: # Kategorisasi Respirasi
Level_Respirasi = [
    (final['Respirasi/(mnt)'] >= 12 & (final['Respirasi/(mnt)'] <= 20),
    (final['Respirasi/(mnt)'] <12),
    (final['Respirasi/(mnt)'] > 20)
    ]
values = ['Normal', 'Tidak Normal', 'Tidak Normal']
final['Level respirasi'] = np.select(Level_Respirasi, values)

In [46]: conditions = [
    (final['Sistolik'] < 120) & (final['Diastolik']< 80),
    (final['Sistolik'] >=120 & (final['Diastolik']<= 80),
    (final['Sistolik'] >=130 & (final['Sistolik']<=140) | (final['Diastolik'] >=80) & (final['Diastolik']<= 89),
    (final['Sistolik'] >140) & (final['Diastolik']>=90)
    ]
values = ['Normal', 'Pra-Hipertensi', 'Stage 1', 'Stage 2']
final['Tingkat Hipertensi'] = np.select(conditions, values)

In [47]: #kateogri Level BMI
Level_BMI = [
    (final['BMI'] < 18.5),
    (final['BMI'] >=18.5 & (final['BMI'] <=24.9),
    (final['BMI'] >=25 & (final['BMI']<= 29.9),
    (final['BMI'] >=30)
    ]
values = ['Underweight', 'Healthy weight', 'Overweight', 'Obesity']
final['Level BMI'] = np.select(Level_BMI, values)

In [48]: final

```

ID	Jenis Kelamin	Kode	Umur	Suhu(C)	Sistolik	Diastolik	Nadi/(mnt)	Respirasi/(mnt)	Tinggi(m)	...	SpO2	Kesadaran	Jenis Rawat	BMI	Level Nadi	Level Suhu	Level SpO2	Level respirasi	Tingkat Hipertensi	Lev	
0	6503	Laki-Laki	111	54.0	37.7	145.0	92.0	80.0	20.0	1.593929	...	90.000000	Compos Mentis	Rawat Inap	26.128708	Normal	Normal	Sedang	Normal	Stage 2	Over
1	75769	Perempuan	111	50.0	36.5	170.0	100.0	82.0	20.0	1.593929	...	80.000000	Compos Mentis	Rawat Inap	26.128708	Normal	Normal	Rendah	Normal	Stage 2	Over
2	75575	Laki-Laki	111	46.0	36.6	190.0	100.0	58.0	20.0	1.593929	...	99.000000	Compos Mentis	Rawat Inap	26.128708	Bradikardia	Normal	Normal	Normal	Stage 2	Over
3	74409	Perempuan	111	80.0	37.1	200.0	94.0	97.0	20.0	1.593929	...	82.000000	Compos Mentis	Rawat Inap	26.128708	Normal	Normal	Rendah	Normal	Stage 2	Over
4	74099	Perempuan	111	68.0	38.0	148.0	96.0	80.0	20.0	1.593929	...	99.000000	Compos Mentis	Rawat Inap	26.128708	Normal	Demam	Normal	Normal	Stage 2	Over
...
832	14434	Laki-Laki	111	57.0	36.0	130.0	70.0	88.0	20.0	1.593929	...	98.284658	Compos Mentis	Rawat Jalan	26.128708	Normal	Normal	Normal	Normal	Pra-Hipertensi	Over
833	14434	Laki-Laki	111	57.0	36.0	130.0	70.0	87.0	20.0	1.593929	...	98.284658	Compos Mentis	Rawat Jalan	26.128708	Normal	Normal	Normal	Normal	Pra-Hipertensi	Over
834	23941	Laki-Laki	111	70.0	38.4	130.0	80.0	98.0	20.0	1.600000	...	98.284658	Compos Mentis	Rawat Jalan	19.531250	Normal	Demam	Normal	Normal	Stage 1	Over

```
In [54]: final.drop(['ID', 'Kode', 'Suhu(C)', 'Sistolik', 'Diastolik', 'Nadi(/mnt)', 'Respirasi(/mnt)', 'Tinggi(m)', 'Berat(Kg)', 'SpO2', 'BMI' ], axis = 1, inplace = True)
final
```

```
Out[54]:
```

	Jenis Kelamin	Umur	Kesadaran	Jenis Rawat	Level Nadi	Level Suhu	Level SpO2	Level respirasi	Tingkat Hipertensi	Level BMI
0	Laki-Laki	54.0	Compos Mentis	Rawat Inap	Normal	Normal	Sedang	Normal	Stage 2	Overweight
1	Perempuan	50.0	Compos Mentis	Rawat Inap	Normal	Normal	Rendah	Normal	Stage 2	Overweight
2	Laki-Laki	46.0	Compos Mentis	Rawat Inap	Bradikardia	Normal	Normal	Normal	Stage 2	Overweight
3	Perempuan	80.0	Compos Mentis	Rawat Inap	Normal	Normal	Rendah	Normal	Stage 2	Overweight
4	Perempuan	68.0	Compos Mentis	Rawat Inap	Normal	Demam	Normal	Normal	Stage 2	Overweight
...
832	Laki-Laki	57.0	Compos Mentis	Rawat Jalan	Normal	Normal	Normal	Normal	Pra-Hipertensi	Overweight
833	Laki-Laki	57.0	Compos Mentis	Rawat Jalan	Normal	Normal	Normal	Normal	Pra-Hipertensi	Overweight
834	Laki-Laki	70.0	Compos Mentis	Rawat Jalan	Normal	Demam	Normal	Normal	Stage 1	Healthy Weight
835	Laki-Laki	70.0	Compos Mentis	Rawat Jalan	Normal	Normal	Normal	Normal	Pra-Hipertensi	Overweight
836	Laki-Laki	70.0	Compos Mentis	Rawat Jalan	Normal	Normal	Normal	Normal	Normal	Overweight

837 rows x 10 columns

```
In [55]: final = final.reindex(columns=['Jenis Kelamin', 'Umur', 'Level BMI', 'Tingkat Hipertensi', 'Level Nadi', 'Level Suhu', 'Level SpO2', 'Level respirasi', 'Kesadaran', 'Jenis Rawat'])
```

```
In [56]: final
```

```
Out[56]:
```

	Jenis Kelamin	Umur	Level BMI	Tingkat Hipertensi	Level Nadi	Level Suhu	Level SpO2	Level respirasi	Kesadaran	Jenis Rawat
0	Laki-Laki	54.0	Overweight	Stage 2	Normal	Normal	Sedang	Normal	Compos Mentis	Rawat Inap
1	Perempuan	50.0	Overweight	Stage 2	Normal	Normal	Rendah	Normal	Compos Mentis	Rawat Inap
2	Laki-Laki	46.0	Overweight	Stage 2	Bradikardia	Normal	Normal	Normal	Compos Mentis	Rawat Inap
836	Laki-Laki	70.0	Compos Mentis	Rawat Jalan	Normal	Normal	Normal	Normal	Normal	Overweight

837 rows x 10 columns

```
In [55]: final = final.reindex(columns=['Jenis Kelamin', 'Umur', 'Level BMI', 'Tingkat Hipertensi', 'Level Nadi', 'Level Suhu', 'Level SpO2', 'Level respirasi', 'Kesadaran', 'Jenis Rawat'])
```

```
In [56]: final
```

```
Out[56]:
```

	Jenis Kelamin	Umur	Level BMI	Tingkat Hipertensi	Level Nadi	Level Suhu	Level SpO2	Level respirasi	Kesadaran	Jenis Rawat
0	Laki-Laki	54.0	Overweight	Stage 2	Normal	Normal	Sedang	Normal	Compos Mentis	Rawat Inap
1	Perempuan	50.0	Overweight	Stage 2	Normal	Normal	Rendah	Normal	Compos Mentis	Rawat Inap
2	Laki-Laki	46.0	Overweight	Stage 2	Bradikardia	Normal	Normal	Normal	Compos Mentis	Rawat Inap
3	Perempuan	80.0	Overweight	Stage 2	Normal	Normal	Rendah	Normal	Compos Mentis	Rawat Inap
4	Perempuan	68.0	Overweight	Stage 2	Normal	Demam	Normal	Normal	Compos Mentis	Rawat Inap
...
832	Laki-Laki	57.0	Overweight	Pra-Hipertensi	Normal	Normal	Normal	Normal	Compos Mentis	Rawat Jalan
833	Laki-Laki	57.0	Overweight	Pra-Hipertensi	Normal	Normal	Normal	Normal	Compos Mentis	Rawat Jalan
834	Laki-Laki	70.0	Healthy Weight	Stage 1	Normal	Demam	Normal	Normal	Compos Mentis	Rawat Jalan
835	Laki-Laki	70.0	Overweight	Pra-Hipertensi	Normal	Normal	Normal	Normal	Compos Mentis	Rawat Jalan
836	Laki-Laki	70.0	Overweight	Normal	Normal	Normal	Normal	Normal	Compos Mentis	Rawat Jalan

837 rows x 10 columns

```
In [57]: final.to_excel('final mr 1.xlsx')
```

```
In [ ]:
```

Model Rapidminer menggunakan data original

The screenshot shows the RapidMiner Studio interface. The main workspace displays a process flow with three operators: 'Retrieve Data baru bgt', 'SMOTE Upsampling', and 'Cross Validation'. The 'Cross Validation' operator is highlighted, and its parameters are shown in the right-hand pane. The parameters include 'leave one out' (unchecked), 'number of folds' (10), and 'sampling type' (automatic). The 'Help' pane for 'Cross Validation' is also visible, providing a synopsis and description of the operator.

Model Rapidminer menggunakan data SMOTE

The screenshot shows the RapidMiner Studio interface. The main workspace displays a process flow with three operators: 'Retrieve Data baru bgt', 'SMOTE Upsampling', and 'Cross Validation'. The 'Process' operator is highlighted, and its parameters are shown in the right-hand pane. The parameters include 'logverbosity' (init) and 'logfile' (empty). The 'Help' pane for 'Process' is also visible, providing a synopsis and description of the operator.