

LAPORAN TUGAS BESAR DATA MINING

“Klasifikasi Rekam Medis Pasien Pada Rumah Sakit Islam Ibnu Sina Menggunakan Algoritma Decision Tree”

Dosen Pengampu:

Aina Hubby Aziira, M.Eng

Dwi Welly Sukma Nirad, M.T.



Disusun Oleh:

Kelompok 3

Nadia Deari Hanifah (2211521004)

Najla Nadiva (2211521006)

Naufal Adli Dhiaurrahman (2211521008)

Azhra Meisa Khairani (2211523010)

Rania Shofi Malika (2211523014)

PROGRAM STUDI SISTEM INFORMASI

FAKULTAS TEKNOLOGI INFORMASI

UNIVERSITAS ANDALAS

2024

KATA PENGANTAR

Tugas Besar Data Mining ini dibuat untuk menyelesaikan tugas mata kuliah “Data Mining”, Departemen Sistem Informasi, Fakultas Teknologi Informasi, Universitas Andalas, semester genap. Serta untuk menjelaskan konsep yang telah kami rencanakan dari hasil diskusi yang dilakukan.

Kelompok penyusun mengucapkan banyak terimakasih kepada Allah SWT karena tanpa nikmat yang diberikan tentunya makalah ini tidak akan selesai dengan sempurna. Terima Kasih pula kepada Ibu Aina Hubby Aziira, M.Eng dan Ibu Dwi Welly Sukma Nirad, M.T selaku dosen pengampu dari mata kuliah "Data Mining" yang telah membimbing dalam pembuatan *Laporan Tugas Besar Data Mining* ini. Tidak lupa penyusun mengucapkan terimakasih kepada orang tua yang telah membantu dan menyemangati penyusun dalam pembuatan laporan ini.

Penyusun menyadari bahwa tugas besar ini memiliki banyak kekurangan serta kelemahan baik dari segi penulisan maupun penyusunan kata. Oleh karena itu, diharapkan kritik dan saran dari pembaca yang dapat membangun kesempurnaan proposal ini. Semoga dapat bermanfaat bagi semua pihak yang terkait.

Padang, 24 Mei 2024

Kelompok 3

DAFTAR ISI

KATA PENGANTAR.....	ii
DAFTAR ISI.....	iii
DAFTAR GAMBAR.....	iv
DAFTAR TABEL.....	v
BAB I PENDAHULUAN.....	1
1.1. Latar Belakang.....	1
1.2. Deskripsi.....	2
1.3. Tujuan.....	3
1.4. Sasaran.....	3
1.5. Manfaat.....	3
1.6. Ruang Lingkup.....	4
BAB II METODOLOGI PENELITIAN.....	6
2.1. Dasar Teori.....	6
2.1.1. Rumah Sakit Islam Ibnu Sina.....	6
2.1.2. Rekam Medis.....	7
2.1.3. Data Mining.....	7
2.1.4. Tahap - Tahap Data Mining.....	10
2.1.5. Dataset.....	11
2.1.6. Klasifikasi.....	11
2.1.7. Algoritma Decision Tree.....	12
2.1.8. Cara Masuk Pasien.....	13
2.2. Metode Penelitian.....	14
2.2.1. Latar Belakang.....	14
2.2.2. Alasan Digunakan Metode Klasifikasi.....	14
2.2.3. Alasan Digunakan Algoritma Decision Tree.....	14
2.3. Sumber Data.....	15
2.4. Timeline Pengerjaan.....	16
BAB III HASIL DAN PEMBAHASAN.....	18
3.1. Pengumpulan Data.....	18
3.2. Pre-Processing Data.....	20
3.2.1. Data Reduction.....	20
3.2.2. Data Cleaning.....	21
3.3. Analisis Data Menggunakan Perhitungan Manual.....	23
3.4. Implementasi Data Mining Menggunakan Python.....	42
3.4.1. Pengolahan Data.....	42
3.4.2. Hasil Pemodelan.....	55
3.5. Perbandingan Hasil.....	56
3.6. Lampiran Kode.....	56

BAB IV PENUTUP.....	58
4.1. Kesimpulan.....	58
4.2. Saran.....	58
DAFTAR PUSTAKA.....	60

DAFTAR GAMBAR

Gambar 2.1 Kajian Umum Data Mining.....	8
Gambar 3.1 Dataset Rekam Medis Rumah Sakit Ibnu Sina Bulan Januari 2024.....	20
Gambar 3.2 Pohon Keputusan Berdasarkan Atribut Opname.....	29
Gambar 3.3 Pohon Keputusan Berdasarkan Atribut Kel.Umur.....	33
Gambar 3.4 Pohon Keputusan Berdasarkan Atribut Jenis Penyakit Umur 25-44 Tahun.....	36
Gambar 3.5 Pohon Keputusan dari Hasil Perhitungan Manual.....	40
Gambar 3.6 Kode Import Library.....	41
Gambar 3.7 Kode Import CSV.....	41
Gambar 3.8 Tipe Data Dataset Rekam Medis.....	42
Gambar 3.9 Dataset Rekam Medis.....	43
Gambar 3.10 Kode Membaca Tipe Data.....	43
Gambar 3.11 Tipe Data Dataset Rekam Medis.....	44
Gambar 3.12 Kode Menghapus Kolom.....	45
Gambar 3.13 Tampilan Kolom yang diperlukan.....	45
Gambar 3.14 Kode Data Shape.....	45
Gambar 3.15 Output Data Shape.....	46
Gambar 3.16 Kode Menghitung Nilai Setiap Kolom.....	46
Gambar 3.17 Output Menghitung Nilai Setiap Kolom.....	46
Gambar 3.18 Kode Menghapus Data Duplikat.....	46
Gambar 3.19 Output Data yang Telah Dibersihkan.....	47
Gambar 3.20 Kode Menghitung Jumlah Nilai Setiap Kolom.....	47
Gambar 3.21 Output Menghitung Nilai Setiap Kolom.....	48
Gambar 3.22 Kode Mencetak Data yang Bernilai Null.....	48
Gambar 3.23 Output Data yang Bernilai Null.....	48
Gambar 3.24 Kode Menghapus Data yang Bernilai Null.....	49
Gambar 3.25 Output Menghitung Nilai Setiap Kolom.....	49
Gambar 3.26 Kode Memeriksa Jumlah Data yang Bernilai Null.....	49
Gambar 3.27 Output Jumlah Data yang Bernilai Null.....	50
Gambar 3.28 Kode Menggunakan Label Encoder.....	50
Gambar 3.29 Output Pengubahan Nilai Menggunakan Label Encoder.....	51
Gambar 3.30 Kode Menentukan Data Atribut Dan data Target.....	51
Gambar 3.31 Kode Pembuatan Model Decision Tree.....	52
Gambar 3.32 Kode Akurasi Pemodelan.....	52
Gambar 3.33 Output Akurasi Pemodelan.....	52
Gambar 3.34 Kode Visualisasi Decision Tree.....	52
Gambar 3.35 Visualisasi Decision Tree Rekam Medis.....	53

DAFTAR TABEL

Tabel 2.1 Timeline.....	16
Tabel 3.1 Sampel dari Dataset.....	23
Tabel 3.2 Perhitungan Entropi dan Gain 1.....	27
Tabel 3.3 Pengambilan keputusan dari Atribut Opname.....	28
Tabel 3.4 Atribut Opname Bernilai False.....	29
Tabel 3.5 Perhitungan Entropi dan Gain 2.....	31
Tabel 3.6 Pengambilan Keputusan dari Atribut Kel.Umur.....	32
Tabel 3.7 Atribut Kel.Umur Berumur 25-44 Tahun.....	33
Tabel 3.8 Perhitungan Entropi dan Gain 3.....	35
Tabel 3.9 Pengambilan Keputusan dari Atribut Jenis Penyakit.....	36
Tabel 3.10 Atribut Kel.Umur Berumur 45-64 Tahun.....	37
Tabel 3.11 Perhitungan Entropi dan Gain 4.....	38
Tabel 3.12 Pengambilan Atribut Jenis Kelamin sebagai Node berikutnya.....	39
Tabel 3.13 Pengambilan Atribut Jenis Penyakit sebagai Node berikutnya.....	39

BAB I

PENDAHULUAN

1.1. Latar Belakang

Dalam era digital yang terus berkembang, penggunaan teknologi informasi dalam berbagai sektor telah menjadi kebutuhan yang mendesak. Salah satu sektor yang sangat diuntungkan dari penggunaan teknologi ini adalah sektor kesehatan. Rumah Sakit Islam Ibnu Sina, sebagai salah satu institusi kesehatan terkemuka, telah mengadopsi sistem rekam medis elektronik (Electronic Medical Records - EMR) untuk menyimpan dan mengelola data medis pasien. Penggunaan EMR memungkinkan pengelolaan data pasien yang lebih efisien dan akurat, serta mendukung pengambilan keputusan klinis yang lebih baik (Nguyen, Bellucci, & Nguyen, 2014).

Namun, dengan meningkatnya jumlah data medis yang tersimpan, muncul tantangan baru dalam hal analisis dan pemanfaatan data tersebut. Salah satu metode yang dapat digunakan untuk mengatasi tantangan ini adalah dengan menggunakan teknik data mining. Data mining memungkinkan ekstraksi informasi yang berharga dari dataset yang besar dan kompleks (García, Luengo, & Herrera, 2015). Dalam konteks rekam medis, data mining dapat digunakan untuk mengidentifikasi pola dan tren yang dapat membantu dalam diagnosis dan perawatan pasien (Jothi, Rashid, & Husain, 2015). Selain membantu dalam diagnosis dan perawatan pasien, teknik data mining juga dapat digunakan untuk menentukan cara masuk pasien berdasarkan beberapa kriteria yang ada pada rekam medis. Penentuan cara masuk pasien, apakah melalui IGD, rawat jalan, atau rawat inap, merupakan aspek penting dalam manajemen rumah sakit. Keputusan ini harus didasarkan pada kondisi medis pasien, tingkat keparahan penyakit, dan faktor lainnya yang tercatat dalam rekam medis (Lavrac, 1999). Dengan analisis yang tepat, rumah sakit dapat memastikan bahwa pasien mendapatkan perawatan yang sesuai dengan kebutuhannya, serta mengoptimalkan penggunaan sumber daya rumah sakit.

Algoritma decision tree merupakan salah satu teknik data mining yang populer digunakan untuk tujuan klasifikasi. Algoritma ini memiliki keunggulan dalam hal interpretabilitas dan kemudahan penggunaan (Kotsiantis, Zaharakis, & Pintelas, 2010). Dengan menggunakan decision tree, kita dapat membangun model prediktif yang dapat mengklasifikasikan cara masuk pasien berdasarkan data rekam medis mereka. Misalnya, berdasarkan gejala yang dilaporkan, riwayat medis, dan hasil pemeriksaan awal, algoritma dapat memprediksi apakah pasien perlu masuk melalui IGD, rawat jalan, atau

rawat inap (Patil & Parmar, 2016). Penelitian ini bertujuan untuk mengklasifikasikan rekam medis pasien di Rumah Sakit Islam Ibnu Sina menggunakan algoritma decision tree guna menentukan cara masuk yang paling tepat bagi pasien. Dengan demikian, diharapkan dapat memberikan kontribusi signifikan dalam peningkatan kualitas pelayanan kesehatan di rumah sakit tersebut. Selain itu, hasil dari penelitian ini dapat menjadi referensi bagi rumah sakit lain yang ingin mengimplementasikan teknologi serupa (Maroco et al., 2011).

1.2. Deskripsi

Penelitian ini memiliki fokus utama pada identifikasi dan klasifikasi pasien ke dalam tiga kategori perawatan utama: rawat inap, rawat jalan, atau rawat darurat. Penelitian ini memanfaatkan data rekam medis yang kaya akan informasi mengenai kondisi kesehatan pasien dan, melalui analisis data mining, mengeksplorasi pola-pola tersembunyi dalam data tersebut. Dengan demikian, diharapkan dapat membantu pengambilan keputusan terkait jenis perawatan yang paling sesuai untuk setiap pasien.

Melalui analisis pola-pola dalam data rekam medis, penelitian ini bertujuan untuk mengidentifikasi faktor-faktor kunci yang mempengaruhi jenis perawatan yang diperlukan oleh pasien. Misalnya, jenis penyakit, usia, gejala yang dialami, dan jenis kelamin dapat menjadi indikator penting dalam menentukan apakah pasien memerlukan perawatan rawat inap, rawat jalan, atau darurat. Dengan informasi ini, rumah sakit dapat meningkatkan efisiensi pelayanan dengan mengarahkan pasien ke jenis perawatan yang paling sesuai dengan kebutuhan mereka. Selain itu, optimalisasi penggunaan sumber daya rumah sakit dapat dicapai, sehingga meningkatkan kualitas perawatan pasien secara keseluruhan.

Hasil dari penelitian ini diharapkan dapat memberikan wawasan baru yang berguna dalam pengelolaan pasien dan pelayanan kesehatan di Rumah Sakit Islam Ibnu Sina Padang Panjang Yarsi Sumbar. Implementasi sistem klasifikasi berbasis Decision Tree ini dapat membantu rumah sakit dalam membuat keputusan yang lebih cepat dan tepat mengenai jenis perawatan yang diperlukan oleh pasien, yang pada gilirannya dapat meningkatkan kepuasan pasien dan efisiensi operasional rumah sakit. Selain itu, penelitian ini juga diharapkan dapat memberikan kontribusi dalam pengembangan metode analisis data medis yang lebih efektif dan efisien, yang dapat diaplikasikan di rumah sakit lain untuk tujuan serupa.

1.3. Tujuan

Penelitian ini bertujuan untuk membangun suatu model berbasis algoritma Decision Tree untuk mengklasifikasikan rekam medis pasien di Rumah Sakit Islam Ibnu Sina Padang Panjang Yarsi Sumbar. Tujuan spesifiknya meliputi:

1. Mengembangkan Algoritma yang Efektif

Menciptakan algoritma Decision Tree yang mampu mengidentifikasi jenis perawatan yang diperlukan pasien (rawat inap, rawat jalan, atau rawat darurat) berdasarkan data rekam medis.

2. Menganalisis dan Memilih Variabel Klinis

Melakukan analisis untuk memilih variabel-variabel klinis yang signifikan dalam memprediksi jenis perawatan yang tepat.

3. Membangun dan Menilai Model Klasifikasi

Membangun model klasifikasi yang akurat dan mengevaluasi manfaat penerapannya dalam meningkatkan efisiensi operasional dan kualitas perawatan pasien.

1.4. Sasaran

Penelitian ini bertujuan untuk memberikan pemahaman mendalam tentang pola-pola dalam layanan perawatan yang diberikan kepada pasien. Melalui analisis data rekam medis yang komprehensif, penelitian ini berupaya mengungkap insight yang berguna untuk mengoptimalkan keputusan terkait jenis perawatan yang paling sesuai, baik itu rawat inap, rawat jalan, atau rawat darurat.

Laporan ini ditujukan kepada berbagai pemangku kepentingan, termasuk manajemen rumah sakit, staf medis, dan pihak administrasi, yang berpotensi mendapatkan manfaat signifikan dari hasil analisis yang diperoleh. Dengan memanfaatkan data mining dan algoritma Decision Tree, diharapkan dapat ditingkatkan efisiensi operasional, pengelolaan sumber daya, dan kualitas pelayanan kesehatan secara keseluruhan. Pemahaman yang lebih baik tentang kebutuhan perawatan pasien juga dapat mendukung pengambilan keputusan yang lebih tepat dan berbasis data, sehingga mampu memberikan pelayanan yang lebih responsif dan tepat sasaran.

1.5. Manfaat

Penelitian ini memberikan manfaat signifikan bagi pembaca dan penulis. Bagi pembaca, terutama manajemen rumah sakit dan staf medis, penelitian ini menawarkan wawasan untuk meningkatkan kualitas pelayanan kesehatan, efisiensi operasional, dan

pengambilan keputusan berbasis data. Bagi penulis, penelitian ini merupakan kesempatan untuk mengembangkan keahlian dalam analisis data dan penerapan algoritma Decision Tree, memberikan kontribusi ilmiah dalam metode analisis data medis, serta meningkatkan kompetensi profesional di bidang data mining dan analisis data kesehatan. Penelitian ini tidak hanya mendukung peningkatan pelayanan pasien tetapi juga memperkaya pengetahuan dan keterampilan di bidang analisis data.

1.6. Ruang Lingkup

Ruang lingkup dari penelitian klasifikasi data mining ini mencakup:

1. Pengumpulan Data

Penelitian ini melibatkan pengumpulan data dari berbagai sumber, termasuk sistem informasi rumah sakit dan catatan medis. Data ini mencakup informasi tentang demografi pasien, jenis layanan yang diterima, dan hasil pengobatan.

2. Pemrosesan Data

Sebelum menggunakan algoritma decision tree, data yang dikumpulkan perlu diproses untuk memastikan bahwa data tersebut dalam format yang sesuai dan siap untuk analisis. Ini mencakup pembersihan data, normalisasi, dan transformasi data.

3. Penerapan Algoritma Decision Tree

Penelitian ini menggunakan algoritma decision tree untuk mengklasifikasikan data berdasarkan berbagai faktor yang mempengaruhi layanan perawatan. Algoritma ini akan membantu dalam mengidentifikasi pola-pola dalam data, yang dapat digunakan untuk membuat keputusan yang lebih baik dalam pelayanan kesehatan.

4. Analisis dan Interpretasi

Setelah algoritma decision tree diterapkan, hasilnya perlu dianalisis dan diinterpretasikan untuk memahami pola-pola dalam layanan perawatan. Ini mencakup evaluasi terhadap efisiensi penggunaan sumber daya dan identifikasi area yang memerlukan peningkatan.

5. Pengembangan Kebijakan dan Strategi

Hasil dari analisis data digunakan untuk mengembangkan kebijakan dan strategi baru dalam pelayanan kesehatan, termasuk peningkatan dalam hal kualitas perawatan, efisiensi penggunaan sumber daya, dan kepuasan pasien.

6. Pendidikan dan Penelitian

Penelitian ini juga berkontribusi pada pendidikan dan penelitian, seperti pengembangan program pelatihan untuk staf medis dan penelitian tentang tren kesehatan dan pola penyakit yang berpotensi menjadi fokus penelitian lebih lanjut.

7. Peningkatan Kinerja dan Reputasi Rumah Sakit

Tujuan dari penelitian ini mencakup peningkatan kinerja dan reputasi rumah sakit melalui peningkatan kualitas layanan dan efisiensi penggunaan sumber daya.

BAB II

METODOLOGI PENELITIAN

Penelitian ini menggunakan metode klasifikasi dengan pendekatan kuantitatif untuk menganalisis data rekam medis dari Rumah Sakit Islam Ibnu Sina Padang Panjang Yarsi Sumbar yang dikumpulkan pada bulan Januari 2024. Algoritma Decision Tree dipilih karena kemampuan interpretabilitas dan efektivitasnya dalam menangani variabel non-linier. Proses penelitian dimulai dengan pengumpulan data rekam medis, diikuti oleh tahap pembersihan (data cleansing) dan transformasi untuk memastikan kualitas dan konsistensi data. Model Decision Tree kemudian dibangun untuk mengidentifikasi pola dan hubungan dalam data, dengan tujuan mengklasifikasikan jenis perawatan yang paling sesuai untuk setiap pasien.

2.1. Dasar Teori

2.1.1. Rumah Sakit Islam Ibnu Sina

Rumah sakit merupakan suatu organisasi yang dijalankan oleh tenaga medis profesional, dilengkapi dengan fasilitas kedokteran yang permanen, serta menyediakan layanan medis, perawatan berkelanjutan, diagnosis, dan pengobatan penyakit bagi pasien. Rumah sakit mempunyai berbagai macam jenis pelayanan kesehatan yang dapat diunggulkan untuk mempertahankan loyalitas pasien diantaranya rawat darurat, rawat jalan, dan rawat inap (Supartiningsih, 2017). Rawat Darurat beroperasi 24 jam sehari tanpa henti untuk menangani pasien yang membutuhkan perawatan segera (*emergency*) dan korban kecelakaan (*casualty*). Fokus utama Rawat Darurat adalah memberikan bantuan secepat mungkin guna menyelamatkan nyawa pasien, sementara Rawat Jalan adalah fasilitas di rumah sakit yang melayani pasien yang datang untuk merencanakan operasi atau melanjutkan pengobatan, seperti kemoterapi, tanpa memerlukan pemeriksaan fisik atau pemeriksaan penunjang yang mendalam (Erfavira, A, 2012). Rawat Inap adalah layanan yang disediakan oleh perawat di ruang rawat inap untuk pasien yang memerlukan perawatan intensif dan pengawasan terus-menerus di rumah sakit (Anjaryani, 2009). Sama halnya dengan rumah sakit pada umumnya, Rumah Sakit Islam Ibnu Sina yang berlokasi di Jl. Soekarno Hatta No. 17, Kota Padang Panjang juga mempunyai 3 jenis pelayanan kesehatan yakni rawat darurat, rawat jalan, dan rawat inap.

2.1.2. Rekam Medis

Pasal 46 ayat (1) UU Praktik Kedokteran menjelaskan bahwa rekam medis merupakan dokumen yang memuat informasi tentang identitas pasien, pemeriksaan, pengobatan, tindakan, dan layanan lain yang diberikan. Sementara itu, Peraturan Menteri Kesehatan Nomor 749a/Menkes/Per/XII/1989 dan Permenkes Nomor 269/MenKes/Per/III/2008 menguraikan bahwa rekam medis berisi catatan dan dokumen seputar pasien, termasuk identitas, pemeriksaan, pengobatan, dan tindakan medis, khususnya di sarana pelayanan kesehatan seperti rawat jalan dan rawat inap, baik yang dikelola oleh pemerintah maupun swasta. Perbedaan utamanya adalah penekanan Permenkes terhadap sarana pelayanan kesehatan. Rekam medis mencatat fakta terkait kondisi pasien, riwayat penyakit, pengobatan masa lalu, dan pengobatan saat ini yang dicatat oleh tenaga kesehatan yang memberikan layanan kepada pasien tersebut (Abduh, R., 2021).

Rekam medis memiliki peran yang signifikan dalam praktik kedokteran, sebagaimana yang dikutip oleh Fransiska Novita (2017) . Beberapa peran utamanya meliputi:

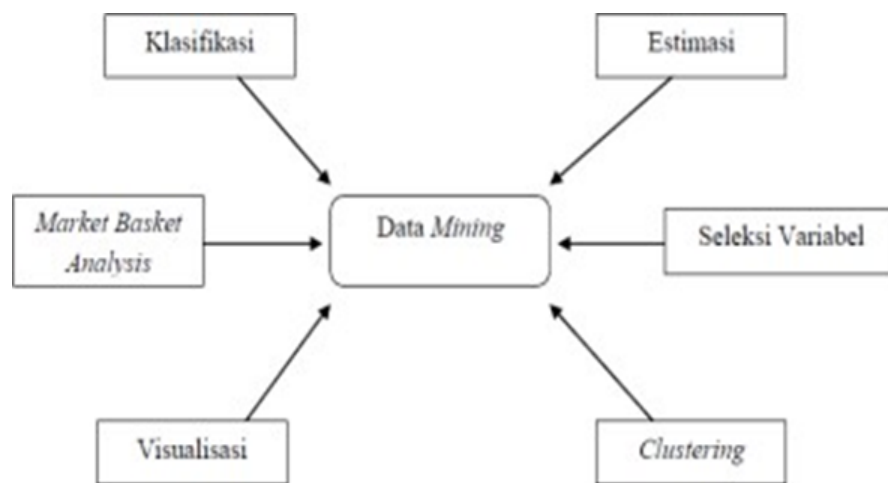
1. Sebagai alat komunikasi antara dokter dan tenaga kesehatan.
2. Fondasi dalam merencanakan pengobatan atau perawatan pasien.
3. Bukti tertulis atas layanan dan pengobatan yang diberikan kepada pasien.
4. Dasar analisis untuk studi dan evaluasi mutu pelayanan kepada pasien.
5. Melindungi kepentingan hukum pasien, rumah sakit, serta dokter dan tenaga kesehatan lainnya.
6. Menyediakan data spesifik yang bermanfaat untuk penelitian.
7. Fondasi perhitungan biaya pelayanan medis pasien.
8. Sumber informasi dan pertanggungjawaban terkait pelayanan medis.

2.1.3. Data Mining

Data mining adalah serangkaian proses untuk mengeksplorasi nilai tambah dari kumpulan data dalam bentuk pengetahuan yang tidak diketahui secara artifisial (Moertini, V. S., 2002). Harus diingat bahwa kata mining itu sendiri berkonotasi dengan upaya mengekstraksi sejumlah kecil komoditas berharga dari sejumlah besar bahan dasar. Karena data mining sebenarnya berasal dari bidang ilmiah seperti kecerdasan buatan, pembelajaran mesin, statistik, dan basis data (Larose, D. T., 2005). Dengan kata lain, data mining adalah proses penggalian pola dari data.

Data mining menjadi alat yang semakin penting untuk mengubah data menjadi informasi. Ini sering digunakan dalam berbagai praktik analitis seperti pemasaran, pengawasan, deteksi penipuan, dan penemuan ilmiah. Selama bertahun-tahun, bisnis, ilmuwan, dan pemerintah telah menggunakan penambangan data untuk mengekstrak data dalam jumlah besar, seperti catatan perjalanan penumpang maskapai, data sensus, dan data pemindai supermarket, untuk menghasilkan laporan riset pasar.

Data mining menjadi pusat dari beberapa kajian (Santosa, B., 2007). Diantaranya adalah estimasi, seleksi variabel, clustering, visualisasi, market basket analysis dan klasifikasi. Berikut gambaran pada gambar 2..



Gambar 2.1 Kajian Umum Data Mining

Adapun beberapa masalah yang sesuai dan dapat diselesaikan dengan penerapan teknik data mining yaitu:

1. Memerlukan keputusan yang bersifat knowledge-based
2. Mempunyai lingkungan yang berubah
3. Metode yang ada sekarang bersifat suboptimal.
4. Tersedia data yang bisa diakses, cukup dan relevan.
5. Memberikan keuntungan yang tinggi jika keputusan yang diambil tepat [9] .

Salah satu tujuan utama data mining adalah untuk mendapatkan hubungan atau pola yang dilihat antara dua atau lebih atribut dan dua atau lebih objek yang mungkin memberikan indikasi yang bermanfaat. Menurut Tan (2004) pemanfaatan dari data mining sendiri bisa dilihat dari dua sudut pandang, baik sudut pandang

komersial dan sudut pandang keilmuan. Dari sudut pandang komersial, data mining bisa digunakan untuk menangani adanya ledakan dari volume data. Dengan melihat bagaimana menyimpannya, mengekstraknya dan memanfaatkannya. Tentunya berbagai ilmu komputasi dapat untuk menghasilkan informasi yang dibutuhkan.

Data mining juga digunakan untuk membantu dalam melakukan analisis koleksi pengamatan. Data mining memiliki manfaat yang signifikan dalam berbagai bidang, mulai dari bidang bisnis, kesehatan, keuangan, pemasaran, dan lain-lain. Pertama, data mining memungkinkan pengambilan keputusan yang lebih baik dan berbasis fakta. Dengan menganalisis data historis dan mengidentifikasi pola-pola yang tersembunyi, data mining dapat membantu mengungkap wawasan dan tren yang berharga. Hal ini memungkinkan organisasi untuk membuat keputusan yang lebih informasional, memperbaiki efisiensi operasional, mengoptimalkan strategi pemasaran, dan meningkatkan kepuasan pelanggan. Selain itu, data mining juga dapat digunakan untuk deteksi penipuan, pengelolaan risiko, prediksi pasar, dan pengembangan produk yang lebih baik. Dengan demikian, data mining berperan penting dalam mendukung pengambilan keputusan yang lebih cerdas dan berdampak positif pada keberhasilan organisasi.

Fungsi umum dalam data mining yaitu :

1. Association, adalah proses untuk menemukan aturan asosiasi Antara suatu kombinasi item dalam suatu waktu.
2. Sequence, adalah proses untuk menentukan aturan asosiasi Antara suatu kombinasi item dalam suatu waktu dan diterapkan lebih dari satu periode.
3. Clustering, adalah proses pengelompokan sejumlah data/objek ke dalam kelompok data sehingga setiap kelompok berisi data yang mirip.
4. Classification, adalah proses penemuan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk dapat memperkirakan kelas dari suatu objek yang labelnya tidak diketahui.
5. Regression, adalah proses pemetaan data dalam suatu nilai prediksi.
6. Forecasting, adalah proses pengestimasian nilai prediksi berdasarkan pola-pola di dalam sekumpulan data.

7. Solution, adalah proses penemuan akar masalah dan problem solving dari persoalan bisnis yang diharapkan atau paling tidak sebagai informasi dalam pengambilan keputusan.

2.1.4. Tahap - Tahap Data Mining

Menurut Han, Kamber, dan Pei (2012) data mining memiliki beberapa tahapan dalam implementasinya, yaitu:

1. Seleksi Data

Tahap ini melibatkan pemilihan data yang relevan untuk analisis dari berbagai sumber. Data yang dipilih harus berkaitan langsung dengan masalah atau tujuan yang ingin dicapai melalui data mining. Seleksi yang tepat dapat meningkatkan efisiensi dan akurasi proses berikutnya.

2. Pra- Pemrosesan

Tahap ini mencakup pembersihan dan pengorganisasian data yang telah dipilih. Proses ini melibatkan penanganan data yang hilang, menghapus data yang tidak relevan, dan mengoreksi ketidakkonsistenan dalam data. Tujuannya adalah untuk meningkatkan kualitas data sehingga hasil analisis menjadi lebih akurat.

3. Transformasi

Pada tahap transformasi, data diubah ke dalam format yang sesuai untuk proses data mining. Ini bisa melibatkan normalisasi data, agregasi, atau pembuatan fitur baru yang dapat membantu dalam analisis. Transformasi data bertujuan untuk mempermudah pengenalan pola dan model.

4. Data Mining

Tahap inti dari seluruh proses, di mana teknik dan algoritma statistik serta machine learning diterapkan untuk menemukan pola atau model yang bermanfaat dari data. Teknik yang digunakan bisa bervariasi tergantung pada tujuan, seperti klasifikasi, klusterisasi, regresi, atau asosiasi.

5. Evaluasi/ Interpretasi

Setelah model data mining diterapkan, hasilnya dievaluasi untuk menentukan keakuratan dan kegunaannya. Interpretasi hasil sangat penting untuk memastikan bahwa temuan yang diperoleh dapat memberikan wawasan yang berarti dan dapat diimplementasikan untuk pengambilan keputusan. Hasil yang dievaluasi juga dapat menunjukkan apakah perlu dilakukan iterasi ulang dari proses data mining.

2.1.5. Dataset

Dataset dalam data mining adalah kumpulan data yang terstruktur atau terorganisir yang digunakan untuk analisis dan penambangan informasi. Dataset dapat berupa tabel, spreadsheet, database, atau format data lainnya yang menyimpan informasi tentang entitas atau objek tertentu. Setiap baris dalam dataset mewakili satu contoh atau instansi data, sedangkan setiap kolom mewakili atribut atau fitur yang terkait dengan setiap contoh. Dataset dapat digunakan untuk melatih model prediktif, mengidentifikasi pola atau tren, dan mengekstrak pengetahuan yang berharga melalui teknik analisis data mining seperti clustering, klasifikasi, regresi, dan lainnya.

2.1.6. Klasifikasi

Klasifikasi adalah salah satu teknik utama dalam data mining yang digunakan untuk mengidentifikasi kategori atau kelas dari data baru berdasarkan data yang sudah diketahui sebelumnya (Han, Kamber, & Pei, 2012). Teknik ini memanfaatkan algoritma pembelajaran mesin untuk membuat model prediktif dari data yang sudah diberi label. Proses klasifikasi biasanya melibatkan beberapa tahap, antara lain:

1. Pemilihan Fitur

Memilih atribut atau fitur yang relevan dari data yang akan digunakan dalam model klasifikasi.

2. Pembentukan Model

Menggunakan data pelatihan yang sudah diberi label untuk melatih algoritma klasifikasi, sehingga model dapat mengenali pola dalam data.

3. Evaluasi Model

Menggunakan data uji untuk menilai performa model dengan mengukur akurasi, presisi, recall, dan metrik lainnya.

4. Penerapan Model

Setelah model dievaluasi dan dianggap cukup akurat, model diterapkan pada data baru untuk memprediksi kelas atau kategori.

2.1.7. Algoritma *Decision Tree*

Algoritma *Decision Tree* adalah salah satu teknik populer dalam data mining dan machine learning yang digunakan untuk tugas klasifikasi dan regresi. Algoritma ini membangun model prediktif dalam bentuk pohon keputusan, yang terdiri dari node dan cabang. Setiap node mewakili keputusan berdasarkan nilai atribut tertentu, dan setiap cabang mewakili hasil keputusan tersebut (Han, Kamber, & Pei, 2012).

Algoritma *Decision Tree* membuat keputusan berdasarkan pembagian data menjadi subset yang semakin kecil hingga mencapai kondisi terminal. Dua metode umum yang digunakan untuk membagi data pada setiap node adalah Information Gain (berbasis Entropy) dan Gini Index.

1. Information Gain (Berbasis Entropy)

Entropy mengukur ketidakpastian atau impurity dalam data. Information Gain mengukur pengurangan ketidakpastian setelah membagi data berdasarkan atribut tertentu. Rumus Entropy dan Information Gain adalah sebagai berikut:

a. Rumus Entropy:

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i$$

Keterangan:

- S : himpunan kasus
- n : jumlah partisi S
- p_i : proporsi dari S_i terhadap S

b. Information Gain:

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * \log_2 p_i$$

Keterangan:

- S : himpunan
- A : atribut
- n : jumlah partisi atribut A
- | Si | : jumlah kasus pada partisi ke-i
- | S | : jumlah kasus dalam S

2. Gini Index

Gini Index adalah ukuran impurity atau ketidakmurnian dalam data. Semakin rendah nilai Gini Index, semakin bersih pembagian data. Rumus Gini Index adalah sebagai berikut:

Gini Index(D):

$$\text{Gini} = 1 - \sum_{i=1}^n (p_i)^2$$

Keterangan:

- D : set data
- n : jumlah kelas
- pi : proporsi dari Si terhadap S

2.1.8. Cara Masuk Pasien

"Cara Masuk Pasien" dalam rekam medis merujuk pada cara atau jalur yang dilalui oleh pasien ketika pertama kali masuk ke rumah sakit untuk mendapatkan perawatan. Ini adalah salah satu aspek penting dalam sistem manajemen rumah sakit karena membantu mengklasifikasikan kondisi awal dan kebutuhan medis pasien. Berdasarkan cara masuk, pasien dapat dikelompokkan ke dalam beberapa kategori, yaitu:

1. Rawat Jalan

Pasien yang datang ke rumah sakit untuk pemeriksaan, konsultasi, atau prosedur medis yang tidak memerlukan rawat inap. Setelah mendapatkan perawatan, pasien bisa langsung pulang.

2. Rawat Darurat

Pasien yang membutuhkan perawatan medis segera karena kondisi yang mengancam nyawa atau kesehatan mereka. Mereka biasanya masuk melalui unit gawat darurat (UGD).

3. Rawat Inap

Pasien yang memerlukan pengawasan dan perawatan intensif yang hanya bisa dilakukan dengan menginap di rumah sakit. Keputusan untuk rawat inap biasanya didasarkan pada kondisi medis yang memerlukan observasi jangka panjang atau perawatan yang tidak dapat dilakukan di rumah.

2.2. Metode Penelitian

2.2.1. Latar Belakang

Rumah Sakit Islam Ibnu Sina Padang Panjang Yarsi Sumbar merupakan salah satu lembaga kesehatan yang memberikan pelayanan medis kepada masyarakat di sekitarnya. Sebagai bagian dari upaya meningkatkan efisiensi dalam penanganan pasien, penting untuk memiliki sistem yang dapat mengidentifikasi jenis perawatan yang paling sesuai untuk setiap pasien. Dalam hal ini, analisis data mining menjadi alat yang sangat berguna untuk memahami pola-pola dan hubungan dalam data rekam medis pasien.

Pada bulan Januari 2024, rumah sakit tersebut mengumpulkan sejumlah besar data rekam medis dari pasien yang datang untuk mendapatkan perawatan. Data ini mencakup berbagai variabel, seperti gejala yang dirasakan, hasil tes laboratorium, riwayat penyakit, dan jenis perawatan yang diberikan (rawat inap, rawat jalan, atau rawat darurat). Tujuan dari analisis ini adalah untuk mengidentifikasi pola-pola dalam data ini yang dapat digunakan untuk klasifikasi jenis perawatan yang paling sesuai untuk setiap pasien.

2.2.2. Alasan Digunakan Metode Klasifikasi

Dalam analisis rekam medis tersebut, diperlukan metode classification untuk mengklasifikasikan data menjadi kategori atau kelas tertentu berdasarkan pola atau atribut tertentu yang ada dalam dataset. Dalam konteks rekam medis, klasifikasi memungkinkan rumah sakit untuk mengidentifikasi jenis perawatan yang paling sesuai untuk setiap pasien berdasarkan informasi medis yang tersedia. Selain itu, metode klasifikasi juga memungkinkan untuk menggali pola-pola yang mendasari dalam data rekam medis, yang dapat memberikan wawasan yang

berharga bagi pengambilan keputusan klinis dan perencanaan strategis dalam pelayanan kesehatan. Dengan demikian, penggunaan metode klasifikasi dalam analisis data mining pada rekam medis memiliki dampak positif yang signifikan dalam meningkatkan efektivitas, efisiensi, dan kualitas pelayanan kesehatan.

2.2.3. Alasan Digunakan Algoritma Decision Tree

Metode yang digunakan untuk analisis ini adalah algoritma Decision Tree. Algoritma ini merupakan salah satu metode klasifikasi yang paling populer dalam analisis data mining. Dengan memanfaatkan algoritma Decision Tree, dapat mengidentifikasi variabel-variabel yang paling berpengaruh dalam menentukan jenis perawatan yang sesuai, serta membangun model klasifikasi yang dapat memprediksi jenis perawatan berdasarkan informasi yang ada dalam rekam medis.

Penggunaan metode Decision Tree dalam analisis data rekam medis untuk mengidentifikasi jenis perawatan yang sesuai bagi pasien di Rumah Sakit Islam Ibnu Sina Padang Panjang Yarsi Sumbar didasarkan pada beberapa pertimbangan yang penting. Pertama, algoritma Decision Tree menawarkan tingkat interpretabilitas yang tinggi. Dalam konteks kesehatan, keterbacaan yang baik dari model adalah krusial karena memungkinkan praktisi medis dan manajer rumah sakit untuk memahami faktor-faktor apa yang mempengaruhi keputusan perawatan. Struktur pohon keputusan yang dihasilkan oleh algoritma ini dengan jelas menunjukkan variabel-variabel yang paling berpengaruh dalam menentukan jenis perawatan yang diperlukan bagi setiap pasien.

Selain itu, Decision Tree juga memiliki kemampuan untuk mengatasi variabel non-linier. Dalam data rekam medis, terdapat beragam informasi yang berhubungan dengan kesehatan pasien, seperti gejala yang dirasakan, hasil tes laboratorium, dan riwayat penyakit. Metode Decision Tree dapat secara efektif menangani kompleksitas hubungan antara variabel-variabel ini, bahkan ketika pola-pola yang mendasarinya bersifat non-linier. Hal ini memungkinkan pembangunan model klasifikasi yang akurat dalam menentukan jenis perawatan yang paling sesuai dengan kondisi setiap pasien.

2.3. Sumber Data

Sumber data untuk penelitian ini diperoleh dari rekam medis rumah sakit Ibnu Sina Padang Panjang, yang merupakan sumber data privat yang mencakup informasi medis pasien. Data ini dikumpulkan secara resmi dan sah dari rumah sakit tersebut pada bulan Januari 2024. Pengumpulan data dilakukan dengan mematuhi semua regulasi dan kebijakan privasi yang berlaku untuk memastikan kerahasiaan informasi medis pasien terjaga. Rumah sakit telah memberikan izin dan akses kepada tim penelitian untuk menggunakan data rekam medis tersebut dalam rangka pengembangan sistem klasifikasi untuk menentukan layanan perawatan pasien. Dengan memanfaatkan sumber data yang valid dan terpercaya ini, diharapkan penelitian dapat menghasilkan model klasifikasi yang akurat dan dapat diandalkan untuk meningkatkan efisiensi dan kualitas layanan perawatan di rumah sakit.

2.4. Timeline Pengerjaan

Pada pengerjaan tugas besar ini, kami telah merencanakan sebuah timeline pengerjaan yang meliputi periode selama 3 bulan. Proses dimulai dari tahap perancangan konsep, dilanjutkan dengan pengumpulan data, dan akhirnya mencapai tahap pembuatan laporan. Kami telah menyusun tabel timeline pengerjaan yang mendetail untuk memandu proses ini dengan lebih efisien dan efektif. Tabel timeline tersebut akan memberikan gambaran yang jelas mengenai waktu yang dialokasikan untuk setiap tahapan, serta memastikan bahwa penelitian berjalan sesuai dengan jadwal yang telah ditetapkan.

Tabel 2.1 Timeline

TIMELINE												
KEGIATAN	AKTIFITAS	Maret				April					Mei	
		1	8	15	22	29	5	12	19	26		
Perancangan Konsep Tubes	Menentukan Tema Tugas Besar	✓										
	Mencari Referensi Studi Literatur	✓										
	Mengidentifikasi Metode Data Mining		✓									

	yang Tepat											
	Mengidentifikasi Algoritma yang tepat		✓									
Pengumpulan Data	Mencari Dataset Sekunder		✓									
Perancangan Penelitian	Menentukan Metode Data Mining yang akan dieksekusi		✓									
	Mendapatkan sumber dataset rekam medis rs			✓								
	Pembuatan Proposal Penelitian				✓							
Pengolahan Awal Data	Menganalisa Atribut yang diperlukan pada Dataset				✓							
	Import Data dan Cleansing Data						✓					
	Melakukan Pengolahan Lebih Lanjut (Data Wrangling)								✓			
	Data Processing Berdasarkan Metodologi								✓			
Konsep Laporan	Penyusunan BAB I								✓			
	Mencari Daftar Pustaka (Referensi)									✓		
Data Processing	Mendapat Hasil Akurasi Data									✓		
	Mendapatkan Kesimpulan dari Data yang diolah									✓		
Laporan	Penyusunan BAB II										✓	

Lanjutan	dan BAB III												
	Pembuatan PPT											✓	
Pengumpulan Laporan	Draft Paper											✓	
Presentasi													

BAB III

HASIL DAN PEMBAHASAN

Pada penelitian ini, data rekam medis yang ada pada Rumah Sakit Islam Ibnu Sina telah dikumpulkan dan kemudian dianalisis menggunakan metode klasifikasi decision tree. Data yang diperoleh telah ditransformasikan untuk dapat diproses lebih lanjut dari bentuk asalnya ke dalam bentuk yang sesuai untuk klasifikasi. Tujuan dari analisis data ini adalah untuk menghasilkan model klasifikasi yang dapat memprediksi cara masuk pasien di Rumah Sakit Islam Ibnu Sina secara konsisten dan relevan serta memperbaiki kesalahan data. Dengan menggunakan metode decision tree, diharapkan data yang diperoleh dapat memberikan makna yang lebih mendalam dan akurat dalam memprediksi cara masuk pasien.

3.1. Pengumpulan Data

Sumber data utama yang digunakan dalam penelitian ini adalah dataset rekam medis pasien di Rumah Sakit Islam Ibnu Sina Pada Bulan Januari 2024 dengan jumlah keseluruhan data sebanyak 4267 record dan 46 kolom. Data yang diterima dalam bentuk format .xlsx atau Microsoft Excel sehingga mempermudah dalam melakukan pembersihan dan pemfilteran data. Data yang diperoleh sudah melalui perizinan kepada pihak Rumah Sakit Islam Ibnu Sina di bawah wewenang manajer rumah sakit tersebut.

Berikut adalah daftar atribut yang terdapat dalam dataset:

1. NOMOR REGISTRASI: Nomor registrasi unik untuk setiap pasien.
2. NOMOR REKAM MEDIS: Nomor rekam medis pasien.
3. UMUR: Usia pasien pada saat kunjungan.
4. KELOMPOK UMUR: Kelompok usia pasien.
5. JENIS KELAMIN: Jenis kelamin pasien.
6. GOLONGAN DARAH: Golongan darah pasien.
7. SUKU: Suku atau etnis pasien.
8. BAHASA: Bahasa yang digunakan oleh pasien.
9. STATUS: Status pernikahan pasien.
10. ASAL PASIEN: Lokasi geografis atau kota asal pasien.
11. KETERANGAN ASAL PASIEN: Keterangan tambahan mengenai asal pasien.
12. CARA MASUK: Cara masuk pasien ke rumah sakit.

13. POLIKLINIK/RUANG: Poliklinik atau ruang yang dikunjungi oleh pasien.
14. OPNAME: Status opname pasien.
15. NOMOR KAMAR/TT: Nomor kamar atau tempat tidur pasien.
16. KELAS: Kelas perawatan pasien.
17. CARA BAYAR: Metode pembayaran yang digunakan oleh pasien.
18. NOMOR JAMINAN: Nomor jaminan atau asuransi pasien.
19. COVID19_NO_SEP: Status kasus COVID-19 pasien.
20. TANGGAL MASUK: Tanggal masuk pasien ke rumah sakit.
21. TANGGAL PULANG: Tanggal pulang pasien dari rumah sakit.
22. LOS: Lama tinggal pasien di rumah sakit (Length of Stay).
23. CARA KELUAR: Cara pasien keluar dari rumah sakit.
24. KETERANGAN (CARA KELUAR): Keterangan tambahan mengenai cara keluar pasien.
25. KEADAAN KELUAR: Keadaan pasien saat keluar dari rumah sakit.
26. DX.MASUK: Diagnosis saat masuk.
27. DPJP: Dokter penanggung jawab pasien.
28. NO.DTD: Nomor dokumen terkait pasien.
29. KASUS: Kasus atau kondisi medis pasien.
30. DTD UTAMA: Diagnosis utama pasien.
31. ICD 10.01: Kode ICD-10 untuk diagnosis utama.
32. DIAGNOSIS 01: Deskripsi diagnosa utama.
33. JENIS PENYAKIT : Kelompok tingkah keparahan penyakit
34. ICD 10.EXTERNAL CAUSES: Kode ICD-10 untuk penyebab eksternal.
35. EXTERNAL CAUSES: Penyebab eksternal kondisi medis.
36. NO.DTD EXTERNAL CAUSES: Nomor dokumen terkait penyebab eksternal.
37. DTD EXTERNAL CAUSES: Tanggal dokumen terkait penyebab eksternal.
38. ICD 9CM.01: Kode ICD-9CM untuk diagnosis utama.
39. PROCEDURE 01: Prosedur medis yang dilakukan.
40. HAK KELAS: Hak kelas pasien.
41. KODE INACBG: Kode INACBG untuk jenis perawatan.
42. DESKRIPSI INACBG: Deskripsi INACBG untuk jenis perawatan.
43. BIAYA RS: Total biaya perawatan yang dibebankan pada rumah sakit.
44. TARIFF INA CBG (HAK): Tarif INA CBG berdasarkan hak kelas pasien.
45. BAYAR (IUR): Jumlah yang dibayarkan oleh pasien (In Up-Down Rule).

46. SELISIH: Selisih antara biaya perawatan dan pembayaran pasien.

Data ini akan menjadi dasar untuk membangun model klasifikasi menggunakan algoritma Decision Tree dalam penelitian ini.

Gambar 3.1 Dataset Rekam Medis Rumah Sakit Ibnu Sina Bulan januari 2024

(File : [Rekam Medis Jan 2024](#))

3.2. Pre-Processing Data

Pre-Processing Data adalah tahapan dari Data Mining yaitu suatu proses atau tahapan yang dilakukan untuk mengolah data mentah menjadi data yang berkualitas yang baik dilanjutkan ke proses selanjutnya. Tahapan preprocessing data sebagai berikut:

3.2.1. Data Reduction

Data Reduction adalah proses untuk mengurangi atau mereduksi sejumlah data yang tidak dibutuhkan. Data Reduction sangat berguna untuk mendapatkan item data yang akan digunakan dalam penelitian ini. Data rekam medis yang telah diperoleh akan difilter terlebih dahulu dan diambil beberapa item yang akan digunakan dalam proses pengolahan data nantinya.

Pada tahapan ini, diperlukannya analisis dataset yang bertujuan menyederhanakan dataset dan memfokuskan perhatian pada faktor-faktor yang paling mempengaruhi hasil analisis yang diinginkan. Melalui analisis kualitatif dan kuantitatif, atribut-atribut yang tidak relevan, redundan, atau kurang informatif dapat diidentifikasi dan dihapus, sementara atribut-atribut yang paling signifikan dan bermanfaat dipertahankan. Berikut atribut yang dipertahankan dan alasan mengapa atribut tersebut diperlukan :

1. KEL.UMUR (Kelompok Umur)

Atribut ini dipilih karena usia pasien dapat menjadi faktor yang signifikan dalam menentukan jenis perawatan yang diperlukan. Berdasarkan kelompok umur, perawatan dan diagnosis bisa berbeda, sehingga atribut ini dapat membantu dalam pengklasifikasian pasien ke dalam kelompok perawatan yang sesuai.

2. JNS.KELAMIN (Jenis Kelamin)

Faktor jenis kelamin juga dapat mempengaruhi jenis dan kebutuhan perawatan. Misalnya, beberapa kondisi medis lebih umum pada jenis kelamin tertentu. Oleh karena itu, atribut ini dipilih untuk membantu dalam pengklasifikasian pasien berdasarkan karakteristik ini.

3. CARA MASUK (Cara Masuk)

Cara pasien masuk ke rumah sakit dapat memberikan informasi tentang jenis kecelakaan atau keadaan darurat tertentu yang mungkin mempengaruhi jenis perawatan yang diperlukan. Atribut ini penting untuk memahami konteks kedatangan pasien ke rumah sakit.

4. OPNAME KASUS (Kasus)

Opname kasus ini mengacu pada kondisi medis pasien yang memerlukan perawatan di rumah sakit (rawat inap), dimana pasien harus tinggal di rumah sakit untuk pengawasan dan perawatan yang intensif. Informasi tentang opname kasus membantu dalam menentukan jenis perawatan yang tepat yang diperlukan oleh pasien.

5. JENIS PENYAKIT

Atribut ini merupakan atribut yang mendasar untuk menganalisis kategori diagnosis berdasarkan klasifikasi tingkat keparahan penyakit. Ini penting untuk memahami kondisi medis pasien yang diperlukan.

3.2.2. Data Cleaning

Setelah data dipilih dan diseleksi sesuai dengan item data yang digunakan, dilakukan proses pembersihan data agar tidak ada duplikasi data, nilai yang hilang, data outliers, dan lainnya. Pembersihan data adalah proses untuk mengisi nilai yang hilang atau yang kurang tepat, mengoreksi data yang tidak konsisten, dan mendeteksi redundansi data seperti duplikasi data.

Berikut beberapa atribut yang dihapus dalam pembersihan data beserta alasannya :

1. Sebagian atau seluruh data kosong.
 - GOL.DARAH
 - KET. ASAL PASIEN
 - COVID19_NO_SEP
 - KETERANGAN (CARA KELUAR)
 - ICD 10.EXTERNA+S1+AG1:AK+AG1:AJ1
 - EXTERNAL CAUSES
 - NO.DTD EXTERNAL CAUSES
 - DTD EXTERNAL CAUSES
 - ICD 9CM.01
2. Tidak Relevan
 - NO.REGISTRASI
 - NO.RM
 - STATUS
 - ASAL PASIEN
 - NO.KAMAR/TT
 - KELAS
 - NO.JAMINAN
 - TGL.MASUK
 - TGL.PULANG
 - LOS
 - CARA KELUAR
 - KETERANGAN (CARA KELUAR)
 - KEADAAN KELUAR
 - DX.MASUK
 - DPJP
 - NO.DTD
 - ICD 10.01
 - HAK KELAS
 - KODE INACBG
 - DESKRIPSI INACBG
 - BIAYA RS
 - TARIFF INA CBG (HAK)

- BAYAR (IUR)
 - SELISIH
3. Data tidak informatif
- CARA BAYAR
 - BAYAR (IUR)

3.3. Analisis Data Menggunakan Perhitungan Manual

Penerapan Algoritma Decision Tree Classification pada proses manual dalam pengelompokan data berdasarkan cara masuk pasien menggunakan aplikasi Microsoft Excel 2021. Pada pengujian data secara manual ini, menggunakan sampel data sebagian dari 4266 records data yang diperoleh yaitu sebanyak 15 sampels.

Berikut data 15 sampels yang diambil :

Tabel 3.1 Sampel dari Dataset

KEL.UMUR	JNS.KELAMIN	OPNAME	JENIS PENYAKIT	CARA MASUK
15 - 24 TAHUN	LAKI-LAKI	FALSE	SEDANG	RAWAT DARURAT
5 - 14 TAHUN	LAKI-LAKI	FALSE	SEDANG	RAWAT JALAN
45 - 64 TAHUN	PEREMPUAN	FALSE	BERAT	RAWAT DARURAT
> 65 TAHUN	LAKI-LAKI	TRUE	SEDANG	RAWAT INAP
> 65 TAHUN	PEREMPUAN	TRUE	BERAT	RAWAT INAP
25 - 44 TAHUN	LAKI-LAKI	TRUE	BERAT	RAWAT INAP
25 - 44 TAHUN	PEREMPUAN	FALSE	BERAT	RAWAT DARURAT
25 - 44 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT JALAN
1 - 4 TAHUN	PEREMPUAN	TRUE	RINGAN	RAWAT INAP
45 - 64 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT JALAN
45 - 64 TAHUN	LAKI-LAKI	TRUE	RINGAN	RAWAT INAP
> 65 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT DARURAT

15 - 24 TAHUN	PEREMPUAN	TRUE	BERAT	RAWAT INAP
25 - 44 TAHUN	PEREMPUAN	FALSE	SEDANG	RAWAT JALAN
25 - 44 TAHUN	PEREMPUAN	FALSE	SEDANG	RAWAT JALAN

Adapun tahapan yang diperlukan adalah sebagai berikut :

1. Menentukan target dari data untuk diambil keputusan

Target ini adalah variabel yang ingin diprediksi atau diklasifikasikan oleh model, dan penentuan target yang tepat akan mempengaruhi keberhasilan seluruh proses analisis dan pengambilan keputusan. Dalam dataset ini, fokusnya adalah pada kolom 'cara masuk', yang menjadi tujuan utama untuk diprediksi. Jadi, dalam analisis data ini, 'cara masuk' adalah apa yang ingin diketahui atau diprediksi dari data tersebut.

“Cara Masuk”

Rawat Inap = 6

Rawat Darurat = 4

Rawat Jalan = 5

- Hitung Entropi dari Atribut Target

Entropi dihitung berdasarkan penjumlahan dari setiap kategori kasus (dalam hal ini rawat jalan, rawat darurat, dan rawat inap), di mana setiap jumlah kasus dibagi dengan total jumlah data dan dikalikan dengan logaritma dari hasil pembagian tersebut, kemudian dikalikan dengan nilai negatif. Hasil dari semua perhitungan tersebut kemudian dijumlahkan. Fungsi dari menghitung entropi dari atribut target adalah untuk mengukur tingkat ketidakpastian atau kekacauan dalam distribusi nilai-nilai pada atribut target tersebut.

Rumus :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i$$

$$Info(D) = \left(-\frac{6}{15} \times \log_2\left(\frac{6}{15}\right)\right) + \left(-\frac{4}{15} \times \log_2\left(\frac{4}{15}\right)\right) + \left(-\frac{5}{15} \times \log_2\left(\frac{5}{15}\right)\right) = 1,566$$

- Hitung Entropi dari tiap kelas pada atribut predaktor

Menghitung entropi dari tiap kelas pada atribut prediktor membantu mengevaluasi tingkat ketidakpastian atau keragaman dalam distribusi nilai-nilai. Proses ini melibatkan perhitungan entropi untuk setiap nilai atau kelas pada atribut prediktor, yang kemudian dijumlahkan untuk mendapatkan entropi total dari atribut tersebut. Dengan mengetahui entropi tiap kelas, kita dapat memahami seberapa bervariasi distribusi nilai-nilai dalam atribut dan memilih strategi analisis yang sesuai untuk menginterpretasikan data.

Rumus :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

a. KEL.UMUR

- $Info(D>65) = \left(-\frac{2}{3} \times \log_2\left(\frac{2}{3}\right)\right) + \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{0}{3} \times \log_2\left(\frac{0}{3}\right)\right) = 0,918$
- $Info(D45-64) = \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) = 1,585$
- $Info(D25-44) = \left(-\frac{1}{5} \times \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{1}{5} \times \log_2\left(\frac{1}{5}\right)\right) + \left(-\frac{3}{5} \times \log_2\left(\frac{3}{5}\right)\right) = 1,371$
- $Info(D15-24) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{0}{2} \times \log_2\left(\frac{0}{2}\right)\right) = 1$
- $Info(D5-14) = \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0$
- $Info(D1-4) = \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = 0$

b. JNS.KELAMIN

$$\begin{aligned}
 - \text{Info(Dperempuan)} &= \left(-\frac{3}{7} \times \text{Log}2\left(\frac{3}{7}\right)\right) + \left(-\frac{2}{7} \times \text{Log}2\left(\frac{2}{7}\right)\right) + \\
 &\quad \left(-\frac{2}{7} \times \text{Log}2\left(\frac{2}{7}\right)\right) = 1,557 \\
 - \text{Info(Dlaki-laki)} &= \left(-\frac{3}{8} \times \text{Log}2\left(\frac{3}{8}\right)\right) + \left(-\frac{2}{8} \times \text{Log}2\left(\frac{2}{8}\right)\right) + \\
 &\quad \left(-\frac{3}{8} \times \text{Log}2\left(\frac{3}{8}\right)\right) = 1,561
 \end{aligned}$$

c. OPNAME

$$\begin{aligned}
 - \text{Info(Dtrue)} &= \left(-\frac{6}{6} \times \text{Log}2\left(\frac{6}{6}\right)\right) + \left(-\frac{0}{6} \times \text{Log}2\left(\frac{0}{6}\right)\right) + \\
 &\quad \left(-\frac{0}{6} \times \text{Log}2\left(\frac{0}{6}\right)\right) = 0 \\
 - \text{Info(Dfalse)} &= \left(-\frac{0}{9} \times \text{Log}2\left(\frac{0}{9}\right)\right) + \left(-\frac{4}{9} \times \text{Log}2\left(\frac{4}{9}\right)\right) + \\
 &\quad \left(-\frac{5}{9} \times \text{Log}2\left(\frac{5}{9}\right)\right) = 0,991
 \end{aligned}$$

d. JENIS PENYAKIT

$$\begin{aligned}
 - \text{Info(Dberat)} &= \left(-\frac{3}{5} \times \text{Log}2\left(\frac{3}{5}\right)\right) + \left(-\frac{2}{5} \times \text{Log}2\left(\frac{2}{5}\right)\right) + \\
 &\quad \left(-\frac{0}{5} \times \text{Log}2\left(\frac{0}{5}\right)\right) = 0,971 \\
 - \text{Info(Dsedang)} &= \left(-\frac{1}{5} \times \text{Log}2\left(\frac{1}{5}\right)\right) + \left(-\frac{1}{5} \times \text{Log}2\left(\frac{1}{5}\right)\right) + \\
 &\quad \left(-\frac{3}{5} \times \text{Log}2\left(\frac{3}{5}\right)\right) = 1,371 \\
 - \text{Info(Dringan)} &= \left(-\frac{2}{5} \times \text{Log}2\left(\frac{2}{5}\right)\right) + \left(-\frac{1}{5} \times \text{Log}2\left(\frac{1}{5}\right)\right) + \\
 &\quad \left(-\frac{2}{5} \times \text{Log}2\left(\frac{2}{5}\right)\right) = 1,522
 \end{aligned}$$

- Hitung Information Gain pada setiap atribut prediktor

Information Gain adalah ukuran untuk mengevaluasi seberapa baik sebuah atribut dapat memisahkan atau mengklasifikasikan data. Untuk menghitung Information Gain pada setiap atribut prediktor, langkahnya adalah menghitung perbedaan antara entropi atribut target awal dengan entropi setiap atribut prediktor. Atribut dengan Information Gain tertinggi dipilih sebagai atribut yang paling informatif untuk membangun model prediksi atau klasifikasi. Information Gain

membantu dalam memilih atribut yang paling berpengaruh dalam memahami pola-pola dalam data.

Rumus :

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * \log_2 p_i$$

- Gain (KEL.UMUR) = $\text{Info (D)} - \text{Info}_{\text{kel.umur}} (\text{D})$
 $= 1,566 - \left(\left(\frac{3}{15} \right) \times 0,918 \right) - \left(\left(\frac{3}{15} \right) \times 1,585 \right) - \left(\left(\frac{5}{15} \right) \times 1,371 \right) -$
 $\left(\left(\frac{2}{15} \right) \times 1 \right) - \left(\left(\frac{1}{15} \right) \times 0 \right) - \left(\left(\frac{1}{15} \right) \times 0 \right)$
 $= 0,477$
- Gain (JNS.KELAMIN) = $\text{Info (D)} - \text{Info}_{\text{jns.kelamin}} (\text{D})$
 $= 1,566 - \left(\left(\frac{7}{15} \right) \times 1,557 \right) - \left(\left(\frac{8}{15} \right) \times 1,561 \right)$
 $= 0,006$
- Gain (OPNAME) = $\text{Info (D)} - \text{Info}_{\text{opname}} (\text{D})$
 $= 1,566 - \left(\left(\frac{6}{15} \right) \times 0 \right) - \left(\left(\frac{9}{15} \right) \times 0,991 \right)$
 $= 0,971$
- Gain (JENIS PENYAKIT) = $\text{Info (D)} - \text{Info}_{\text{jenispenyakit}} (\text{D})$
 $= 1,566 - \left(\left(\frac{5}{15} \right) \times 0,971 \right) - \left(\left(\frac{5}{15} \right) \times 1,371 \right) - \left(\left(\frac{5}{15} \right) \times 1,522 \right)$
 $= 0,277$

Tabel 3.2 Perhitungan Entropi dan Gain 1

		Jumlah (S)	Rawat Inap	Rawat Darurat	Rawat Jalan	Entropy	Gain
Total		15	6	4	5	1,56559623	
Kel. Umur							0,4746276986
	>65	3	2	1	0	0,9182958341	
	45 - 64	3	1	1	1	1,584962501	
	25 - 44	5	1	1	3	1,370950594	
	15 - 24	2	1	1	0	1	
	5 -14	1	0	0	1	0	

	1 - 4	1	1	0	0	0	
Jenis Kelamin							0,0064747671 63
	Perempuan	7	3	2	2	1,556656707	
	Laki-laki	8	3	2	3	1,561278124	
Opname							0,9709505945
	TRUE	6	6	0	0	0	
	FALSE	9	0	4	5	0,9910760598	
Jenis Penyakit							0,2776531358
	Berat	5	3	2	0	0,9709505945	
	Sedang	5	1	1	3	1,370950594	
	Ringan	5	2	1	2	1,521928095	

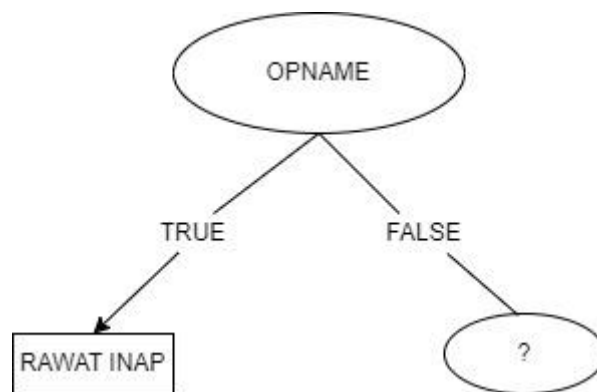
Dari hasil perhitungan gain, atribut 'Opname' memperoleh nilai gain tertinggi. Oleh karena itu, dalam mengambil keputusan berikutnya, prioritas diberikan pada atribut 'Opname' sebagai atribut yang paling informatif untuk langkah selanjutnya dalam memisahkan atau mengklasifikasikan data.

Tabel 3.3 Pengambilan keputusan dari Atribut Opname

KEL.UMUR	JNS.KELAMIN	OPNAME	JENIS PENYAKIT	CARA MASUK
15 - 24 TAHUN	LAKI-LAKI	FALSE	SEDANG	RAWAT DARURAT
5 - 14 TAHUN	LAKI-LAKI	FALSE	SEDANG	RAWAT JALAN
45 - 64 TAHUN	PEREMPUAN	FALSE	BERAT	RAWAT DARURAT
> 65 TAHUN	LAKI-LAKI	TRUE	SEDANG	RAWAT INAP
> 65 TAHUN	PEREMPUAN	TRUE	BERAT	RAWAT INAP
25 - 44 TAHUN	LAKI-LAKI	TRUE	BERAT	RAWAT INAP
25 - 44 TAHUN	PEREMPUAN	FALSE	BERAT	RAWAT DARURAT
25 - 44 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT JALAN

1 - 4 TAHUN	PEREMPUAN	TRUE	RINGAN	RAWAT INAP
45 - 64 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT JALAN
45 - 64 TAHUN	LAKI-LAKI	TRUE	RINGAN	RAWAT INAP
> 65 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT DARURAT
15 - 24 TAHUN	PEREMPUAN	TRUE	BERAT	RAWAT INAP
25 - 44 TAHUN	PEREMPUAN	FALSE	SEDANG	RAWAT JALAN
25 - 44 TAHUN	PEREMPUAN	FALSE	SEDANG	RAWAT JALAN

Dari tabel yang disajikan, jika 'Opname' bernilai True, maka pasien akan diklasifikasikan ke dalam kategori rawat inap. Namun, jika 'Opname' bernilai False, keputusan mengenai status pasien harus ditentukan berdasarkan atribut lain dalam dataset.



Gambar 3.2 Pohon Keputusan Berdasarkan Atribut Opname

Selanjutnya akan dilakukan perhitungan berdasarkan opname yang bernilai False untuk mendapatkan informasi gain dengan perhitungan yang sama seperti yang telah dilakukan diatas.

Tabel 3.4 Atribut Opname Bernilai False

KEL.UMUR	JNS.KELAMIN	OPNAME	JENIS PENYAKIT	CARA MASUK
----------	-------------	--------	----------------	------------

15 - 24 TAHUN	LAKI-LAKI	FALSE	SEDANG	RAWAT DARURAT
5 - 14 TAHUN	LAKI-LAKI	FALSE	SEDANG	RAWAT JALAN
45 - 64 TAHUN	PEREMPUAN	FALSE	BERAT	RAWAT DARURAT
25 - 44 TAHUN	PEREMPUAN	FALSE	BERAT	RAWAT DARURAT
25 - 44 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT JALAN
45 - 64 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT JALAN
> 65 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT DARURAT
25 - 44 TAHUN	PEREMPUAN	FALSE	SEDANG	RAWAT JALAN
25 - 44 TAHUN	PEREMPUAN	FALSE	SEDANG	RAWAT JALAN

2. Menentukan jumlah Atribut Target berdasarkan Atribut Opname

Hal ini melibatkan analisis untuk menentukan jumlah kategori Atribut Target yang relevan tergantung pada nilai Atribut Opname (True atau False). Dalam hal ini, jumlah atribut target berdasarkan False.

“Cara Masuk”

Rawat Darurat = 4

Rawat Jalan = 5

- Hitung Entropi dari Atribut Target

Menghitung entropi dengan menjumlahkan dari setiap kategori kasus pada Atribut Cara Masuk yang terdiri dari rawat jalan, rawat darurat, dan rawat inap untuk Atribut Opname.

Rumus :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i$$

$$Info(D) = \left(-\frac{4}{9} \times \log_2\left(\frac{4}{9}\right)\right) + \left(-\frac{5}{9} \times \log_2\left(\frac{5}{9}\right)\right) = 0,991$$

- Hitung Entropi dari tiap kelas pada atribut predaktor

Menghitung entropi untuk setiap nilai atau kelas pada atribut opname, yang kemudian dijumlahkan untuk mendapatkan entropi total dari atribut opname.

Rumus :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i \cdot \log_2 p_i$$

a. KEL.UMUR

- $Info(D>65) = \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = 0$
- $Info(D45-64) = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) + \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right)\right) = 1$
- $Info(D25-44) = \left(-\frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \times \log_2\left(\frac{3}{4}\right)\right) = 0,8113$
- $Info(D15-24) = \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = 0$
- $Info(D5-14) = \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0$

b. JNS.KELAMIN

- $Info(Dperempuan) = \left(-\frac{2}{4} \times \log_2\left(\frac{2}{4}\right)\right) + \left(-\frac{2}{4} \times \log_2\left(\frac{2}{4}\right)\right) = 1$
- $Info(Dlaki-laki) = \left(-\frac{2}{5} \times \log_2\left(\frac{2}{5}\right)\right) + \left(-\frac{3}{5} \times \log_2\left(\frac{3}{5}\right)\right) = 0,9709$

c. JENIS PENYAKIT

- $Info(Dberat) = \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right)\right) + \left(-\frac{0}{2} \times \log_2\left(\frac{0}{2}\right)\right) = 0$
- $Info(Dsedang) = \left(-\frac{1}{4} \times \log_2\left(\frac{1}{4}\right)\right) + \left(-\frac{3}{4} \times \log_2\left(\frac{3}{4}\right)\right) = 0,8113$
- $Info(Dringan) = \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} \times \log_2\left(\frac{2}{3}\right)\right) = 0,9183$

-

- Hitung Information Gain pada setiap atribut prediktor

Menghitung Information Gain pada setiap atribut prediktor adalah dengan menghitung perbedaan antara entropi atribut target awal dengan entropi setiap atribut prediktor.

Rumus :

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * \log_2 \text{pi}$$

- Gain (KEL.UMUR) = $\text{Info (D)} - \text{Infokel.umur (D)}$
 $= 0,991 - \left(\left(\frac{1}{9} \right) \times 0 \right) - \left(\left(\frac{2}{9} \right) \times 1 \right) - \left(\left(\frac{4}{9} \right) \times 0,8113 \right) - \left(\left(\frac{1}{9} \right) \times 0 \right) - \left(\left(\frac{1}{9} \right) \times 0 \right)$
 $= 0,4083$
- Gain (JNS.KELAMIN) = $\text{Info (D)} - \text{Infojns.kelamin (D)}$
 $= 0,991 - \left(\left(\frac{4}{9} \right) \times 1 \right) - \left(\left(\frac{5}{9} \right) \times 0,9709 \right)$
 $= 0,0072$
- Gain (JENIS PENYAKIT) = $\text{Info (D)} - \text{Infojenispenyakit (D)}$
 $= 0,991 - \left(\left(\frac{2}{9} \right) \times 0 \right) - \left(\left(\frac{4}{9} \right) \times 0,8113 \right) - \left(\left(\frac{3}{9} \right) \times 0,9183 \right)$
 $= 0,3244$

Tabel 3.5 Perhitungan Entropi dan Gain 2

		Jumlah (S)	Rawat Inap	Rawat Darurat	Rawat Jalan	Entropy	Gain
Opname False		9	0	4	5	0,9910760598	
Kel. Umur							0,4082857823
	>65	1	0	1	0	0	
	45 - 64	2	0	1	1	1	
	25 - 44	4	0	1	3	0,8112781245	
	15 - 24	1	0	1	0	0	
	5 - 14	1	0	0	1	0	
	1 - 4	0	0	0	0	0	

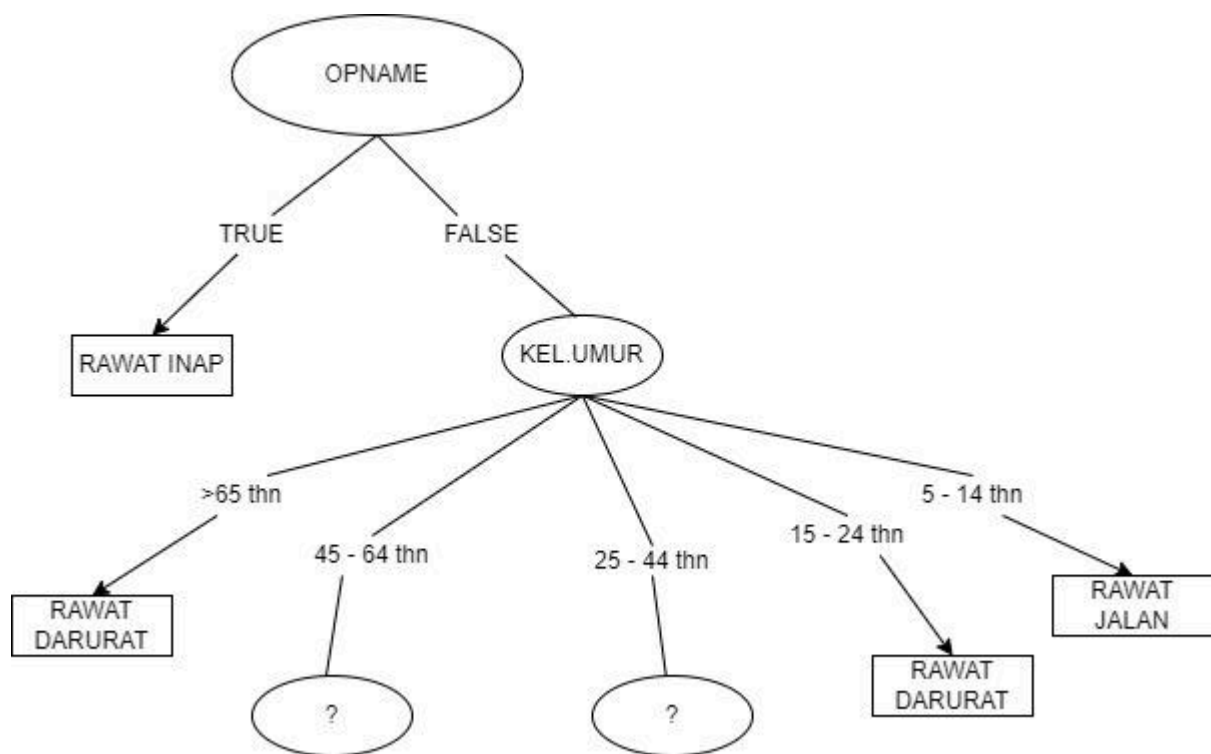
Jenis Kelamin							0,0072146184 75
	Perempuan	4	0	2	2	1	
	Laki-laki	5	0	2	3	0,9709505945	
Jenis Penyakit							0,3244093932
	Berat	2	0	2	0	0	
	Sedang	4	0	1	3	0,8112781245	
	Ringan	3	0	1	2	0,9182958341	

Dari hasil perhitungan gain, atribut 'Kel.Umur' memperoleh nilai gain tertinggi. Oleh karena itu, dalam mengambil keputusan berikutnya, prioritas diberikan pada atribut 'Kel.Umur' sebagai atribut yang paling informatif untuk langkah selanjutnya dalam memisahkan atau mengklasifikasikan data.

Tabel 3.6 Pengambilan Keputusan dari Atribut Kel.Umur

KEL.UMUR	JNS.KELAMIN	OPNAME	JENIS PENYAKIT	CARA MASUK
15 - 24 TAHUN	LAKI-LAKI	FALSE	SEDANG	RAWAT DARURAT
5 - 14 TAHUN	LAKI-LAKI	FALSE	SEDANG	RAWAT JALAN
45 - 64 TAHUN	PEREMPUAN	FALSE	BERAT	RAWAT DARURAT
25 - 44 TAHUN	PEREMPUAN	FALSE	BERAT	RAWAT DARURAT
25 - 44 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT JALAN
45 - 64 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT JALAN
> 65 TAHUN	LAKI-LAKI	FALSE	RINGAN	RAWAT DARURAT
25 - 44 TAHUN	PEREMPUAN	FALSE	SEDANG	RAWAT JALAN
25 - 44 TAHUN	PEREMPUAN	FALSE	SEDANG	RAWAT JALAN

Berdasarkan tabel yang dihasilkan, jika pasien dalam kelompok usia di atas 65 tahun, mereka akan diklasifikasikan ke dalam kategori rawat inap. Kelompok usia 15-24 tahun akan dikategorikan ke dalam rawat darurat, sementara kelompok usia 5-14 tahun akan masuk ke dalam kategori rawat jalan. Namun, untuk kelompok usia 25-44 dan 45-64 tahun, keputusan mengenai status pasien harus ditentukan berdasarkan atribut lain dalam dataset.



Gambar 3.3 Pohon Keputusan Berdasarkan Atribut Kel.Umur

Selanjutnya akan dilakukan perhitungan berdasarkan Kel.Umur 25-44 Tahun untuk mendapatkan informasi gain dengan perhitungan yang sama seperti yang telah dilakukan diatas.

Tabel 3.7 Atribut Kel.Umur Berumur 25-44 Tahun

KEL.UMUR	JNS.KELAMIN	CARA MASUK	OPNAME	JENIS PENYAKIT
25 - 44 TAHUN	PEREMPUAN	RAWAT DARURAT	FALSE	Berat
25 - 44 TAHUN	PEREMPUAN	RAWAT JALAN	FALSE	Sedang

25 - 44 TAHUN	LAKI-LAKI	RAWAT JALAN	FALSE	Ringan
25 - 44 TAHUN	PEREMPUAN	RAWAT JALAN	FALSE	Sedang

3. Menentukan jumlah Atribut Target berdasarkan Atribut Kel.Umur

Hal ini melibatkan analisis untuk menentukan jumlah kategori Atribut Target yang relevan tergantung pada nilai Atribut Kel.Umur (25-44 Tahun & 45-64 Tahun). Dalam hal ini, jumlah atribut target berdasarkan umur 25-44 Tahun..

“Cara Masuk” =

Rawat Darurat = 1

Rawat Jalan = 3

- Hitung Entropi dari Atribut Target

Menghitung entropi dengan menjumlahkan dari setiap kategori kasus pada Atribut Cara Masuk yang terdiri dari rawat jalan, rawat darurat, dan rawat inap untuk Atribut Kel.Umur 25-44 Tahun.

Rumus :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i$$

$$\text{Info (D)} = \left(-\frac{1}{4} \times \log_2 \left(\frac{1}{4} \right) \right) + \left(-\frac{3}{4} \times \log_2 \left(\frac{3}{4} \right) \right) = 0,811$$

- Hitung Entropi dari tiap kelas pada atribut predaktor

Menghitung entropi untuk setiap nilai atau kelas pada atribut kel.umur, yang kemudian dijumlahkan untuk mendapatkan entropi total dari atribut kel.umur..

Rumus :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i$$

a. JNS.KELAMIN

$$- \text{Info(Dperempuan)} = \left(-\frac{1}{3} \times \log_2\left(\frac{1}{3}\right)\right) + \left(-\frac{2}{3} \times \log_2\left(\frac{2}{3}\right)\right) = 0,918$$

$$- \text{Info(Dlaki-laki)} = \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0$$

b. JENIS PENYAKIT

$$- \text{Info(Dberat)} = \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) = 0$$

$$- \text{Info(Dsedang)} = \left(-\frac{0}{2} \times \log_2\left(\frac{0}{2}\right)\right) + \left(-\frac{2}{2} \times \log_2\left(\frac{2}{2}\right)\right) = 0$$

$$- \text{Info(Dringan)} = \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right)\right) + \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right)\right) = 0$$

- Hitung Information Gain pada setiap atribut prediktor

Menghitung Information Gain pada setiap atribut prediktor adalah dengan menghitung perbedaan antara entropi atribut target awal dengan entropi setiap atribut prediktor.

Rumus :

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * \log_2 \text{ pi}$$

a. $\text{Gain (JNS.KELAMIN)} = \text{Info (D)} - \text{Info}_{\text{jns.kelamin}} (\text{D})$

$$= 0,811 - \left(\left(\frac{3}{4}\right) \times 0,918\right) - \left(\left(\frac{1}{4}\right) \times 0\right)$$

$$= 0,12255$$

b. $\text{Gain (JENIS PENYAKIT)} = \text{Info (D)} - \text{Info}_{\text{jenispenyakit}} (\text{D})$

$$= 0,811 - \left(\left(\frac{1}{4}\right) \times 0\right) - \left(\left(\frac{2}{4}\right) \times 0\right) - \left(\left(\frac{1}{4}\right) \times 0\right)$$

$$= 0,81127$$

Tabel 3.8 Perhitungan Entropi dan Gain 3

		Jumlah (S)	Rawat Inap	Rawat Darurat	Rawat Jalan	Entropy	Gain
Kel.Umur 25 -		4	0	1	3	0,8112781245	

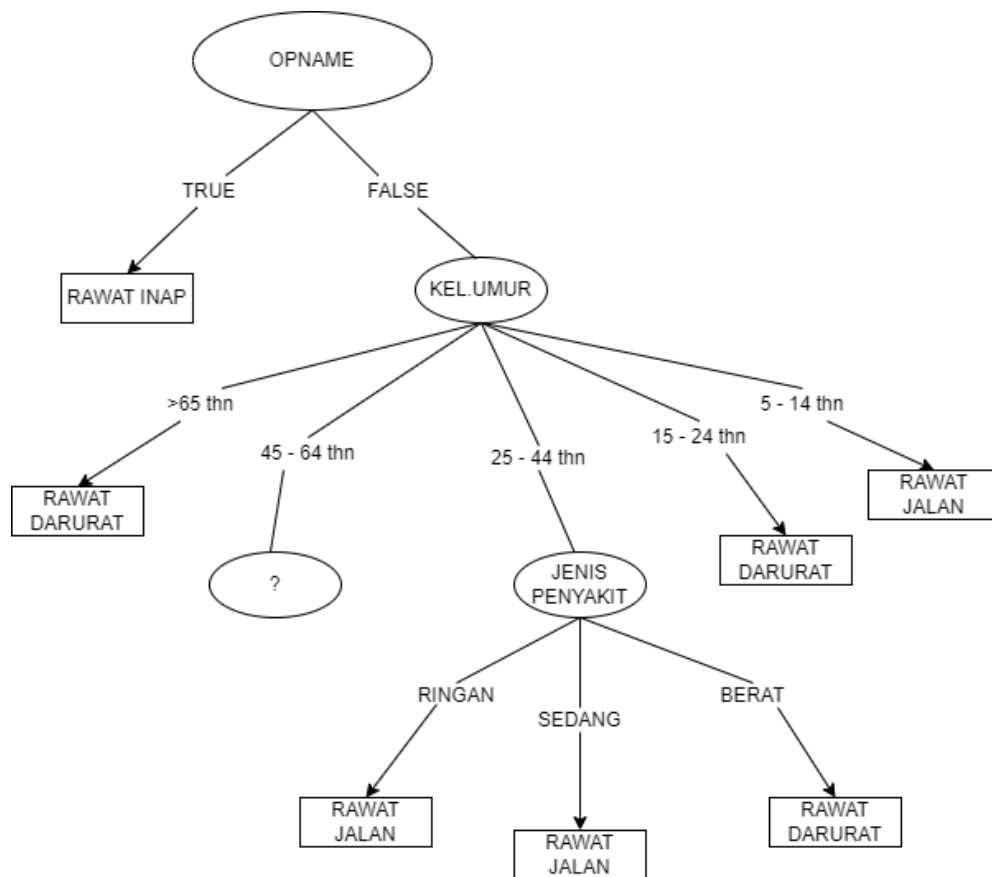
44 thn.							
Jenis Kelamin							0,1225562489
	Perempuan	3	0	1	2	0,9182958341	
	Laki-laki	1	0	0	1	0	
Jenis Penyakit							0,8112781245
	Berat	1	0	1	0	0	
	Sedang	2	0	0	2	0	
	Ringan	1	0	0	1	0	

Dari hasil perhitungan gain, atribut 'Jenis Penyakit' memperoleh nilai gain tertinggi. Oleh karena itu, dalam mengambil keputusan berikutnya, prioritas diberikan pada atribut 'Jenis Penyakit' sebagai atribut yang paling informatif untuk langkah selanjutnya dalam memisahkan atau mengklasifikasikan data.

Tabel 3.9 Pengambilan Keputusan dari Atribut Jenis Penyakit

KEL.UMUR	JNS.KELAMIN	CARA MASUK	OPNAME	JENIS PENYAKIT
25 - 44 TAHUN	PEREMPUAN	RAWAT DARURAT	FALSE	Berat
25 - 44 TAHUN	PEREMPUAN	RAWAT JALAN	FALSE	Sedang
25 - 44 TAHUN	LAKI-LAKI	RAWAT JALAN	FALSE	Ringan
25 - 44 TAHUN	PEREMPUAN	RAWAT JALAN	FALSE	Sedang

Berdasarkan tabel yang dihasilkan, jika pasien dengan Jenis Penyakit Ringan, mereka akan diklasifikasikan ke dalam kategori rawat jalan. Jenis penyakit sedang akan dikategorikan ke dalam rawat jalan, sementara jenis penyakit berat akan masuk ke dalam kategori rawat darurat.



Gambar 3.4 Pohon Keputusan Berdasarkan Atribut Jenis Penyakit Umur 25-44 Tahun

Selanjutnya akan dilakukan perhitungan berdasarkan Kel.Umur 45-64 Tahun untuk mendapatkan informasi gain dengan perhitungan yang sama seperti yang telah dilakukan diatas.

Tabel 3.10 Atribut Kel.Umur Berumur 45-64 Tahun

KEL.UMUR	JNS.KELAMIN	CARA MASUK	OPNAME	JENIS PENYAKIT
45 - 64 TAHUN	PEREMPUAN	RAWAT DARURAT	FALSE	Berat
45 - 64 TAHUN	LAKI-LAKI	RAWAT JALAN	FALSE	Ringan

3. Menentukan jumlah Atribut Target berdasarkan Atribut Kel.Umur

Hal ini melibatkan analisis untuk menentukan jumlah kategori Atribut Target yang relevan tergantung pada nilai Atribut Kel.Umur (25-44 Tahun & 45-64 Tahun). Dalam hal ini, jumlah atribut target berdasarkan umur 45-64 Tahun.

“Cara Masuk”

Rawat Darurat = 1

Rawat Jalan = 1

- Hitung Entropi dari Atribut Target

Menghitung entropi dengan menjumlahkan dari setiap kategori kasus pada Atribut Cara Masuk yang terdiri dari rawat jalan, rawat darurat, dan rawat inap untuk Atribut Kel.Umur 45-64 Tahun.

Rumus :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i$$

$$\text{Info (D)} = \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right) \right) + \left(-\frac{1}{2} \times \log_2\left(\frac{1}{2}\right) \right) = 1$$

- Hitung Entropi dari tiap kelas pada atribut predaktor

Menghitung entropi untuk setiap nilai atau kelas pada atribut kel.umur, yang kemudian dijumlahkan untuk mendapatkan entropi total dari atribut kel.umur.

Rumus :

$$\text{Entropy (S)} = \sum_{i=1}^n -p_i * \log_2 p_i$$

a. JNS.KELAMIN

$$- \text{Info(Dperempuan)} = \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right) \right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right) \right) = 0$$

$$- \text{Info(Dlaki-laki)} = \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right) \right) + \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right) \right) = 0$$

b. JENIS PENYAKIT

$$- \text{Info(Dberat)} = \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right) \right) + \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right) \right) = 0$$

$$- \text{Info(Dringan)} = \left(-\frac{0}{1} \times \log_2\left(\frac{0}{1}\right) \right) + \left(-\frac{1}{1} \times \log_2\left(\frac{1}{1}\right) \right) = 0$$

- Hitung Information Gain pada setiap atribut prediktor

Menghitung Information Gain pada setiap atribut prediktor adalah dengan menghitung perbedaan antara entropi atribut target awal dengan entropi setiap atribut prediktor.

Rumus :

$$\text{Gain (S,A)} = \text{Entropy(S)} - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * \log_2 p_i$$

- a. $\text{Gain (JNS.KELAMIN)} = \text{Info (D)} - \text{Info}_{\text{jns.kelamin}} (\text{D})$
 $= 1 - \left(\left(\frac{1}{2} \right) \times 0 \right) - \left(\left(\frac{1}{2} \right) \times 0 \right)$
 $= 1$
- b. $\text{Gain (JENIS PENYAKIT)} = \text{Info (D)} - \text{Info}_{\text{jenispenyakit}} (\text{D})$
 $= 1 - \left(\left(\frac{1}{2} \right) \times 0 \right) - \left(\left(\frac{1}{2} \right) \times 0 \right)$
 $= 1$

Tabel 3.11 Perhitungan Entropi dan Gain 4

		Jumlah (S)	Rawat Inap	Rawat Darurat	Rawat Jalan	Entrophy	Gain
Total		2	0	1	1	1	
Jenis Kelamin							1
	Perempua n	1	0	1	0	0	
	Laki-laki	1	0	0	1	0	
Jenis Penyakit							1
	Berat	1	0	1	0	0	
	Sedang	0	0	0	0	0	
	Ringan	1	0	0	1	0	

Dikarenakan terdapat kesamaan nilai gain pada beberapa atribut, maka atribut dibagi menjadi dua node yang berbeda untuk memperjelas proses pengklasifikasian data. Node pertama didasarkan pada atribut Jenis Kelamin, sementara node kedua didasarkan pada atribut Jenis

Penyakit. Dengan pemisahan ini, kita dapat melakukan pengklasifikasian lebih lanjut berdasarkan informasi yang terkandung dalam atribut Jenis Kelamin dan Jenis Penyakit.

Tabel 3.12 Pengambilan Atribut Jenis Kelamin sebagai Node berikutnya

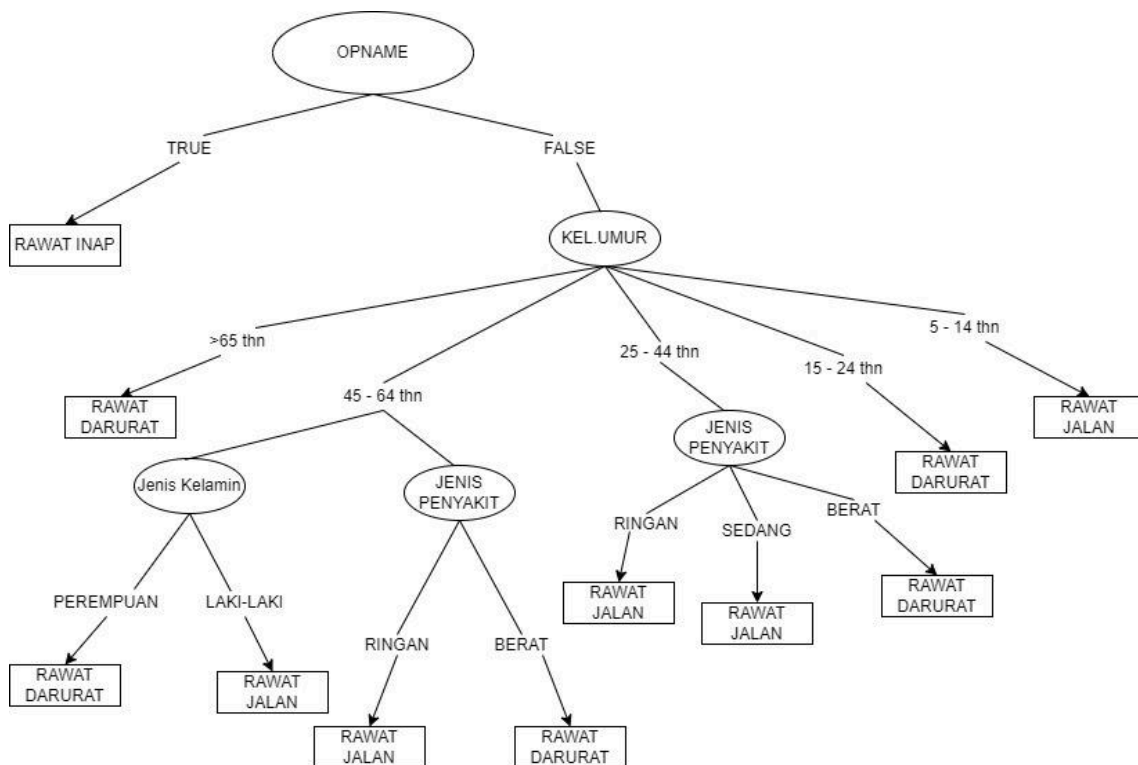
KELUMUR	JNS.KELAMIN	CARA MASUK	OPNAME	JENIS PENYAKIT
45 - 64 TAHUN	PEREMPUAN	RAWAT DARURAT	FALSE	Berat
45 - 64 TAHUN	LAKI-LAKI	RAWAT JALAN	FALSE	Ringan

Dari tabel terlihat bahwa berdasarkan Jenis Kelamin, perempuan berada pada layanan rawat darurat sementara laki-laki pada rawat jalan.

Tabel 3.13 Pengambilan Atribut Jenis Penyakit sebagai Node berikutnya

KELUMUR	JNS.KELAMIN	CARA MASUK	OPNAME	JENIS PENYAKIT
45 - 64 TAHUN	PEREMPUAN	RAWAT DARURAT	FALSE	Berat
45 - 64 TAHUN	LAKI-LAKI	RAWAT JALAN	FALSE	Ringan

Dari tabel terlihat bahwa berdasarkan Jenis Penyakit, kategori jenis penyakit berat berada pada layanan rawat darurat sementara kategori jenis penyakit ringan pada rawat jalan.



Gambar 3.5 Pohon Keputusan dari Hasil Perhitungan Manual

3.4. Implementasi Data Mining Menggunakan Python

Python adalah salah satu bahasa pemrograman yang paling populer di dunia saat ini, terutama dalam bidang data science dan machine learning. Dikenal karena sintaksnya yang sederhana dan mudah dipelajari, Python menawarkan berbagai library dan framework yang kuat, seperti NumPy, Pandas, Scikit-learn, dan TensorFlow, yang sangat membantu dalam analisis data, visualisasi, dan pengembangan model machine learning.

Salah satu alat yang paling sering digunakan oleh para ilmuwan data dan peneliti dalam ekosistem Python adalah Jupyter Notebook. Jupyter Notebook adalah aplikasi web open-source yang memungkinkan pengguna untuk membuat dan berbagi dokumen yang berisi kode langsung, visualisasi, dan narasi teks. Ini sangat berguna untuk eksplorasi data, dokumentasi, dan presentasi hasil analisis.

3.4.1. Pengolahan Data

1. Melakukan import library yang dibutuhkan

```
[1]: import pandas as pd
import numpy as np
import seaborn as sns
import matplotlib.pyplot as plt
%matplotlib inline
sns.set(color_codes=True)
```

Gambar 3.6 Kode Import Library

Terdapat empat library yang digunakan dalam mengklasifikasikan cara masuk pada rekam medis di Rumah Sakit Islam Ibnu Sina. Untuk analisis dan visualisasi data menggunakan Python, diperlukannya mengimpor Pandas, NumPy, Seaborn, dan Matplotlib, memungkinkan untuk melakukan manipulasi data, komputasi numerik, dan visualisasi data. Kemudian, pengaturan `%matplotlib inline` memastikan bahwa plot yang dihasilkan oleh Matplotlib akan ditampilkan secara langsung di dalam notebook, tanpa perlu memanggil `plt.show()`. Terakhir, menggunakan `sns.set(color_codes=True)` dengan mengatur pengaturan default Seaborn untuk menggunakan palet warna yang berbeda dari Matplotlib, memberikan estetika yang lebih menarik pada visualisasi yang akan dibuat.

2. Mengimport csv dan Membaca file RekamMedis.csv

```
[2]: data = pd.read_csv('RekamMedis.csv')
data.info()
data.head()
```

Gambar 3.7 Kode Import CSV

Selanjutnya membaca file CSV yang bernama 'RekamMedis.csv' ke dalam sebuah DataFrame menggunakan Pandas dengan perintah `pd.read_csv('RekamMedis.csv')`. DataFrame adalah struktur data utama dalam Pandas yang mirip dengan tabel database. Kemudian, `data.info()` digunakan untuk menampilkan informasi tentang DataFrame, seperti jumlah baris dan kolom, serta tipe data dari setiap kolom. Ini membantu kita memahami struktur data yang sedang kita kerjakan. Terakhir, `data.head()` digunakan untuk menampilkan lima baris pertama dari DataFrame, memungkinkan kita untuk melihat contoh data dan memeriksa apakah data telah terbaca dengan benar.

Output :

Hasil dari import data, menampilkan dataset sebagai berikut.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 4266 entries, 0 to 4265
Data columns (total 46 columns):
#   Column                                Non-Null Count  Dtype
---  -
0   NO.REGISTRASI                        4266 non-null   int64
1   NO.RM                                4266 non-null   int64
2   UMUR                                 4266 non-null   object
3   KEL.UMUR                             4266 non-null   object
4   JNS.KELAMIN                         4266 non-null   object
5   GOL.DARAH                           30 non-null     object
6   SUKU                                 4266 non-null   object
7   BAHASA                              4266 non-null   object
8   STATUS                              4266 non-null   object
9   ASAL PASIEN                         4266 non-null   object
10  KET. ASAL PASIEN                    9 non-null      object
11  CARA MASUK                          4266 non-null   object
12  POLIKLINIK/RUANG                    4266 non-null   object
13  OPNAME                              4266 non-null   bool
14  NO.KAMAR/TT                         284 non-null    object
15  KELAS                               284 non-null    object
16  CARA BAYAR                          4266 non-null   object
17  NO.JAMINAN                          4266 non-null   object
18  COVID19_NO_SEP                     0 non-null      float64
19  TGL.MASUK                          4266 non-null   object
20  TGL.PULANG                         4266 non-null   object
21  LOS                                 4266 non-null   int64
22  CARA KELUAR                         4266 non-null   object
23  KETERANGAN (CARA KELUAR)            186 non-null    object
24  KEADAAN KELUAR                     4266 non-null   object
25  DX.MASUK                           3280 non-null   object
26  DPJP                                4266 non-null   object
27  NO.DTD                              4266 non-null   float64
28  KASUS                              4265 non-null   object
29  DTD UTAMA                          4266 non-null   object
30  ICD 10.01                          4266 non-null   object
31  DIAGNOSIS 01                      4266 non-null   object
32  JENIS PENYAKIT                     4266 non-null   object
33  ICD 10.EXTERNA+S1+AG1:AK+AG1:AJ1  0 non-null      float64
34  EXTERNAL CAUSES                    0 non-null      float64
35  NO.DTD EXTERNAL CAUSES              0 non-null      float64
36  DTD EXTERNAL CAUSES                 0 non-null      float64
37  ICD 9CM.01                         774 non-null    float64
38  PROCEDURE 01                       774 non-null    object
39  HAK KELAS                          4266 non-null   int64
40  KODE INACBG                        4266 non-null   object
41  DESKRIPSI INACBG                   4266 non-null   object
42  BIAYA RS                           4266 non-null   object
43  TARIFF INA CBG (HAK)                4266 non-null   object
44  BAYAR (IUR)                        4266 non-null   object
45  SELISIH                            4266 non-null   object
dtypes: bool(1), float64(7), int64(4), object(34)
```

Gambar 3.8 Tipe Data Dataset Rekam Medis

memory usage: 1.5+ MB

[2]:

	NO.REGISTRASI	NO.RM	UMUR	KEL.UMUR	JNS.KELAMIN	GOL.DARAH	SUKU	BAHASA	STATUS	ASAL PASIEN	...	DTD EXTERNAL CAUSES	ICD 9CM.01	PROCEDURE 01
0	1414580001	141458	71 TAHUN	> 65 TAHUN	PEREMPUAN	NaN	MINANGKABAU	MINANGKABAU	BARU	DATANG SENDIRI	...	NaN	NaN	NaN
1	1414960001	141496	34 TAHUN	25 - 44 TAHUN	LAKI-LAKI	NaN	MINANGKABAU	MINANGKABAU	BARU	DATANG SENDIRI	...	NaN	99.04	Transfusion of packed cells
2	907050004	90705	40 TAHUN	25 - 44 TAHUN	PEREMPUAN	NaN	MINANGKABAU	MINANGKABAU	LAMA	DATANG SENDIRI	...	NaN	NaN	NaN
3	573030024	57303	51 TAHUN	45 - 64 TAHUN	PEREMPUAN	NaN	BELUM TERDEFINISI	BELUM TERDEFINISI	LAMA	DATANG SENDIRI	...	NaN	NaN	NaN
4	1415770001	141577	63 TAHUN	45 - 64 TAHUN	LAKI-LAKI	NaN	MINANGKABAU	MINANGKABAU	BARU	DATANG SENDIRI	...	NaN	NaN	NaN

5 rows x 46 columns

Gambar 3.9 Dataset Rekam Medis

3. Membaca tipe - tipe data dari file tersebut

[3]: `data.dtypes`

Gambar 3.10 Kode Membaca Tipe Data

Kemudian melakukan pemanggilan metode `dtypes` pada objek `DataFrame` yang disebut `data`. Metode ini digunakan untuk menampilkan tipe data dari setiap kolom dalam `DataFrame` tersebut. Ini memberikan informasi tentang jenis data yang disimpan dalam setiap kolom, apakah itu integer, float, string, atau tipe data lainnya.

Output :

Tipe data yang diperoleh antara lain sebagai berikut.

```
[3]: NO.REGISTRASI      int64
      NO.RM             int64
      UMUR              object
      KEL.UMUR          object
      JNS.KELAMIN       object
      GOL.DARAH         object
      SUKU              object
      BAHASA            object
      STATUS            object
      ASAL PASIEN       object
      KET. ASAL PASIEN  object
      CARA MASUK        object
      POLIKLINIK/RUANG  object
      OPNAME            bool
      NO.KAMAR/TT       object
      KELAS             object
      CARA BAYAR        object
      NO.JAMINAN        object
      COVID19_NO_SEP    float64
      TGL.MASUK         object
      TGL.PULANG        object
      LOS              int64
      CARA KELUAR       object
      KETERANGAN (CARA KELUAR) object
      KEADAAN KELUAR    object
      DX.MASUK          object
      DPJP             object
      NO.DTD            float64
      KASUS             object
      DTD UTAMA         object
      ICD 10.01         object
      DIAGNOSIS 01      object
      JENIS PENYAKIT    object
      ICD 10.EXTERNA+S1+AG1:AK+AG1:AJ1 float64
      EXTERNAL CAUSES   float64
      NO.DTD EXTERNAL CAUSES float64
      DTD EXTERNAL CAUSES float64
      ICD 9CM.01        float64
      PROCEDURE 01      object
      HAK KELAS         int64
      KODE INACBG       object
      DESKRIPSI INACBG  object
      BIAYA RS          object
      TARIFF INA CBG (HAK) object
      BAYAR (IUR)       object
      SELISIH           object
      dtype: object
```

Gambar 3.11 Tipe Data Dataset Rekam Medis

Dari atribut atribut diatas, Terdapat empat jenis tipe data yaitu integer, object, float serta boolean.

4. Menghapus kolom - kolom yang tidak diperlukan

```
[4]: data = data.drop(['NO.REGISTRASI', 'NO.RM', 'UMUR', 'GOL.DARAH', 'SUKU', 'BAHASA', 'STATUS', 'ASAL PASIEN', 'KET. ASAL PASIEN', 'POLIKLINIK/RUANG', 'NO.KAI', 'NO.KAMAR/TT', 'KELAS', 'CARA BAYAR', 'NO.JAMINAN', 'COVID19_NO_SEP', 'TGL.MASUK', 'TGL.PULANG', 'LOS', 'CARA KELUAR', 'KETERANGAN (CARA KELUAR)', 'KEADAAN KELUAR', 'DX.MASUK', 'DPJP', 'NO.DTD', 'KASUS', 'DTD UTAMA', 'ICD 10.01', 'DIAGNOSIS 01', 'JENIS PENYAKIT', 'ICD 10.EXTERNA+S1+AG1:AK+AG1:AJ1', 'EXTERNAL CAUSES', 'NO.DTD EXTERNAL CAUSES', 'DTD EXTERNAL CAUSES', 'ICD 9CM.01', 'PROCEDURE 01', 'HAK KELAS', 'KODE INACBG', 'DESKRIPSI INACBG', 'BIAYA RS', 'TARIFF INA CBG (HAK)', 'BAYAR (IUR)', 'SELISIH'])
      data.head(5)
```

Gambar 3.12 Kode Menghapus Kolom

Tahap selanjutnya kita menggunakan metode `drop()` pada DataFrame yang disebut data untuk menghapus kolom yang telah ditentukan. Parameter pertama dari metode `drop()` adalah daftar nama kolom yang ingin dihapus, dan parameter kedua, `axis=1`, menunjukkan bahwa kita ingin menghapus kolom tersebut (bukan baris). Setelah kolom-kolom tersebut dihapus, `data.head(5)` digunakan untuk menampilkan lima baris pertama dari DataFrame yang telah dimodifikasi.

Output :

Hasil yang didapat dari menghapus kolom-kolom yang tidak diperlukan dapat dilihat seperti gambar dibawah. Terdapat enam kolom tersisa yaitu KEL.UMUR, JNS.KELAMIN, CARA MASUK, OPNAME, KASUS, dan JENIS PENYAKIT.

[4]:

	KEL.UMUR	JNS.KELAMIN	CARA MASUK	OPNAME	KASUS	JENIS PENYAKIT
0	> 65 TAHUN	PEREMPUAN	RAWAT INAP	True	BARU	Berat
1	25 - 44 TAHUN	LAKI-LAKI	RAWAT INAP	True	BARU	Berat
2	25 - 44 TAHUN	PEREMPUAN	RAWAT INAP	True	LAMA	Berat
3	45 - 64 TAHUN	PEREMPUAN	RAWAT INAP	True	LAMA	Sedang
4	45 - 64 TAHUN	LAKI-LAKI	RAWAT INAP	False	BARU	Berat

Gambar 3.13 Tampilan Kolom yang diperlukan

5. Menampilkan jumlah baris dan kolom

```
[5]: data.shape
```

Gambar 3.14 Kode Data Shape

Setelah itu memanggil atribut `shape` dari objek DataFrame yang disebut data. Atribut `shape` memberikan tuple yang berisi jumlah baris dan kolom dalam DataFrame tersebut. Dengan menggunakan kode ini, kita dapat dengan cepat melihat dimensi dari DataFrame, yaitu berapa banyak baris dan kolom yang ada.

Output :

Dengan menggunakan kode diatas, maka didapatkan jumlah baris yaitu 4266 serta jumlah kolom adalah enam.

```
[5]: (4266, 6)
```

Gambar 3.15 Output Data Shape

6. Menghitung jumlah nilai dalam setiap kolom yang sudah dimodifikasi

```
[6]: data.count()
```

Gambar 3.16 Kode Menghitung Nilai Setiap Kolom

Untuk menghitung jumlah nilai dalam setiap kolom DataFrame dapat dilakukan dengan kode diatas. Ini memberikan informasi tentang seberapa lengkap data dalam setiap kolomnya. Hasilnya adalah sebuah Series yang berisi jumlah values untuk setiap kolom.

Output :

Hasil dari menghitung jumlah nilai dalam setiap kolom adalah terdapat 4266 baris pada kolom KEL.UMUR, JNS.KELAMIN, CARA MASUK, OPNAME, JENIS PENYAKIT. Namun pada kolom KASUS hanya terdapat 4265 baris data. Hal ini menandakan bahwa terdapat missing value pada salah satu nilai pada kolom KASUS.

```
[6]: KEL.UMUR      4266
      JNS.KELAMIN  4266
      CARA MASUK   4266
      OPNAME       4266
      KASUS        4265
      JENIS PENYAKIT 4266
      dtype: int64
```

Gambar 3.17 Output Menghitung Nilai Setiap Kolom

7. Menghapus baris yang terdapat duplikat data

```
[7]: data = data.drop_duplicates()
      data.head(5)
```

Gambar 3.18 Kode Menghapus Data Duplikat

Selanjutnya menghapus baris-baris yang merupakan duplikat satu sama lain dalam DataFrame. Setelah operasi ini, DataFrame akan berisi setiap baris hanya sekali, tanpa ada duplikat. Kemudian, `data.head(5)` digunakan untuk menampilkan lima baris pertama dari DataFrame yang telah dimodifikasi. Ini membantu untuk memeriksa apakah operasi penghapusan duplikat berhasil dilakukan dan memastikan bahwa DataFrame terlihat seperti yang diinginkan setelah operasi tersebut.

Output :

Berikut hasil menghapus data yang duplikat.

[7]:

	KEL.UMUR	JNS.KELAMIN	CARA MASUK	OPNAME	KASUS	JENIS PENYAKIT
0	> 65 TAHUN	PEREMPUAN	RAWAT INAP	True	BARU	Berat
1	25 - 44 TAHUN	LAKI-LAKI	RAWAT INAP	True	BARU	Berat
2	25 - 44 TAHUN	PEREMPUAN	RAWAT INAP	True	LAMA	Berat
3	45 - 64 TAHUN	PEREMPUAN	RAWAT INAP	True	LAMA	Sedang
4	45 - 64 TAHUN	LAKI-LAKI	RAWAT INAP	False	BARU	Berat

Gambar 3.19 Output Data yang Telah Dibersihkan

8. Menghitung jumlah nilai dalam setiap kolom yang sudah dimodifikasi

```
[8]: data.count()
```

Gambar 3.20 Kode Menghitung Jumlah Nilai Setiap Kolom

Setelah dilakukan menghapus baris yang terdapat duplikasi, kita menampilkan jumlah nilai dalam setiap kolom DataFrame. Ini memberikan informasi tentang seberapa lengkap data dalam setiap kolomnya. Hasilnya adalah sebuah Series yang berisi jumlah values untuk setiap kolom.

Output :

Setelah dilakukannya penghapusan data yang duplikat maka tersisa 176 baris dari kolom KEL.UMUR, JNS.KELAMIN, CARA MASUK, OPNAME, JENIS PENYAKIT. Serta 175 baris dari kolom KASUS.


```
[8]: KEL.UMUR      176
      JNS.KELAMIN   176
      CARA MASUK    176
      OPNAME        176
      KASUS          175
      JENIS PENYAKIT 176
      dtype: int64
```

Gambar 3.21 Output Menghitung Nilai Setiap Kolom

9. Menampilkan informasi keberadaan nilai null dalam dataset

```
[9]: print(data.isnull().sum())
```

Gambar 3.22 Kode Mencetak Data yang Bernilai Null

Kode di atas mencetak hasil dari metode `isnull()` yang diterapkan pada objek DataFrame `data`, diikuti oleh metode `sum()`. Metode `isnull()` digunakan untuk menghasilkan DataFrame yang memiliki nilai `True` di setiap sel yang bernilai null (NaN), dan `False` di setiap sel yang tidak null. Kemudian, metode `sum()` digunakan untuk menjumlahkan nilai `True` dalam setiap kolom DataFrame, yang menghasilkan jumlah nilai null (NaN) dalam setiap kolom. Hasilnya adalah sebuah Series yang menunjukkan jumlah nilai null dalam setiap kolom DataFrame. Ini memberikan informasi tentang keberadaan nilai null dalam dataset, yang penting untuk memahami kualitas dan keutuhan data sebelum melakukan analisis lebih lanjut.

Output :

Berikut hasil perhitungan kolom yang terdapat nilai null didalamnya. Dapat dilihat dari gambar dibawah, kolom `KASUS` memiliki satu nilai null didalamnya.

```
KEL.UMUR      0
JNS.KELAMIN    0
CARA MASUK     0
OPNAME         0
KASUS          1
JENIS PENYAKIT 0
dtype: int64
```

Gambar 3.23 Output Data yang Bernilai Null

10. Menghapus Data yang Kosong dengan `data.dropna()`

```
[10]: data = data.dropna()  
data.count()
```

Gambar 3.24 Kode Menghapus Data yang Bernilai Null

Langkah ini penting untuk membersihkan dataset dari baris atau kolom yang mengandung nilai kosong (NaN), sehingga menghindari gangguan atau kesalahan saat proses analisis atau pemodelan. Metode `dropna()` digunakan pada dataframe `data` dengan parameter `inplace=True` untuk menghapus nilai kosong secara langsung dari dataframe tersebut.

Output :

Hasil dari penghapusan data yang bernilai null dapat dilihat pada gambar. Pada gambar terlihat bahwa seluruh kolom bernilai sama yaitu 175. Sehingga tidak terdapat lagi data yang bernilai Null.

```
[10]: KEL.UMUR      175  
      JNS.KELAMIN   175  
      CARA.MASUK    175  
      OPNAME        175  
      KASUS         175  
      JENIS.PENYAKIT 175  
      dtype: int64
```

Gambar 3.25 Output Menghitung Nilai Setiap Kolom

11. Memeriksa Jumlah Data Kosong dengan `data.isnull().sum()`

```
[11]: print(data.isnull().sum())
```

Gambar 3.26 Kode Memeriksa Jumlah Data yang Bernilai Null

Setelah menghapus nilai kosong, digunakan metode `isnull().sum()` untuk memverifikasi apakah dataset sudah bersih dari nilai kosong. Metode ini menghitung jumlah nilai kosong (NaN) di setiap kolom dataframe `data`, sehingga memastikan tidak ada lagi nilai kosong yang tersisa sebelum melanjutkan proses analisis atau pemodelan.

Output :

Hasil dari perhitungan nilai Null menunjukkan bahwa tidak terdapat lagi nilai yang kosong (Null) pada data tersebut. sehingga dapat dipastikan semua data memiliki nilainya masing-masing.

```
KEL.UMUR      0
JNS.KELAMIN   0
CARA_MASUK    0
OPNAME        0
KASUS         0
JENIS_PENYAKIT 0
dtype: int64
```

Gambar 3.27 Output Jumlah Data yang Bernilai Null

12. Mengubah Data Kategorikal Menjadi Data Numerik

```
[12]: from sklearn.preprocessing import LabelEncoder

encoder = {}
for col in data.columns:
    encoder[col] = LabelEncoder()
    data[col] = encoder[col].fit_transform(data[col])
print(data)
```

Gambar 3.28 Kode Menggunakan Label Encoder

Kode tersebut menggunakan library `scikit-learn` untuk melakukan label encoding pada setiap kolom dalam dataset `data`. `LabelEncoder` dari `sklearn.preprocessing` digunakan untuk mengubah data kategorikal dalam setiap kolom menjadi nilai numerik. Pertama, sebuah dictionary `encoder` dibuat untuk menyimpan objek `LabelEncoder` untuk setiap kolom. Kemudian, kode ini melakukan iterasi melalui setiap kolom dalam `data`, membuat dan menyimpan objek `LabelEncoder` untuk kolom tersebut, serta melakukan transformasi data dengan metode `fit_transform`. Akhirnya, setiap kolom dalam dataset `data` diubah menjadi nilai numerik yang sesuai dan hasil akhirnya dicetak. Kode ini berguna untuk mempersiapkan data kategorikal agar dapat digunakan dalam algoritma machine learning yang hanya menerima input numerik.

Output :

Berikut hasil dari pengubahan nilai kategorikal menjadi nilai numerik pada data.

	KEL.UMUR	JNS.KELAMIN	CARA MASUK	OPNAME	KASUS	JENIS PENYAKIT
0	8	1	1	1	0	0
1	3	0	1	1	0	0
2	3	1	1	1	1	0
3	5	1	1	1	1	2
4	5	0	1	0	0	0
...
3790	6	1	1	1	0	2
3805	7	1	0	0	1	1
4004	3	0	2	0	0	2
4085	7	1	2	0	1	1
4129	6	0	2	0	1	0

[175 rows x 6 columns]

Gambar 3.29 Output Pengubahan Nilai Menggunakan Label Encoder

13. Menentukan Data Atribut dan Data Target

```
[13]: x = data [['KEL.UMUR', 'JNS.KELAMIN', 'OPNAME', 'KASUS', 'JENIS PENYAKIT']]

[14]: y = data ['CARA MASUK']

[15]: from sklearn.model_selection import train_test_split
      X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=123)
```

Gambar 3.30 Kode Menentukan Data Atribut Dan data Target

Dalam parameter X digunakan untuk menentukan kolom predictor yang digunakan untuk memprediksi hasil. Sedangkan parameter y digunakan untuk menentukan kolom yang akan dijadikan target. Parameter X berisikan kolom KEL.UMUR, JNS.KELAMIN, OPNAME, KASUS, JENIS PENYAKIT. Serta parameter y berisikan kolom CARA MASUK. Dataset dibagi menjadi data atribut dan data target menggunakan fungsi `train_test_split` dari `scikit-learn`. `X_train` dan `X_test` akan berisi fitur-fitur (atribut) dari dataset, sedangkan `y_train` dan `y_test` akan berisi label (target) yang sesuai. Proporsi data uji biasanya ditentukan dengan parameter `test_size`.

14. Membuat Decision Tree Classifier dan Menghitung Akurasi

```
[16]: from sklearn.tree import DecisionTreeClassifier

      tree_model = DecisionTreeClassifier()
      tree_model = tree_model.fit(X_train, y_train)
```

Gambar 3.31 Kode Pembuatan Model Decision Tree

```
[17]: from sklearn.metrics import accuracy_score

y_pred = tree_model.predict(X_test)
acc_score = round(accuracy_score(y_pred, y_test), 3)
print('Accuracy:', acc_score)
```

Gambar 3.32 Kode Akurasi Pemodelan

Decision tree classifier dibuat menggunakan DecisionTreeClassifier() dari scikit-learn. Setelah model dibuat, model dilatih menggunakan data latih (X_train dan y_train) dengan metode fit(). Selanjutnya, model digunakan untuk memprediksi data uji (X_test) dan akurasi model dievaluasi menggunakan accuracy_score() dari scikit-learn.

Output :

Hasil dari perhitungan akurasi yang dilakukan oleh kode diatas, didapatkan bahwa akurasi dari pemodelan ini adalah 0,611.

```
Accuracy: 0.611
```

Gambar 3.33 Output Akurasi Pemodelan

15. Visualisasi Decision Tree

```
[18]: import matplotlib.pyplot as plt
from sklearn import tree

plt.figure(figsize=(100,90))
tree.plot_tree(tree_model, feature_names=['KEL.UMUR', 'JNS.KELAMIN', 'OPNAME', 'KASUS', 'JENIS PENYAKIT'],
               class_names=['RAWAT INAP', 'RAWAT JALAN', 'RAWAT DARURAT'],
               filled=True, rounded=True,
               fontsize=10, label='all')

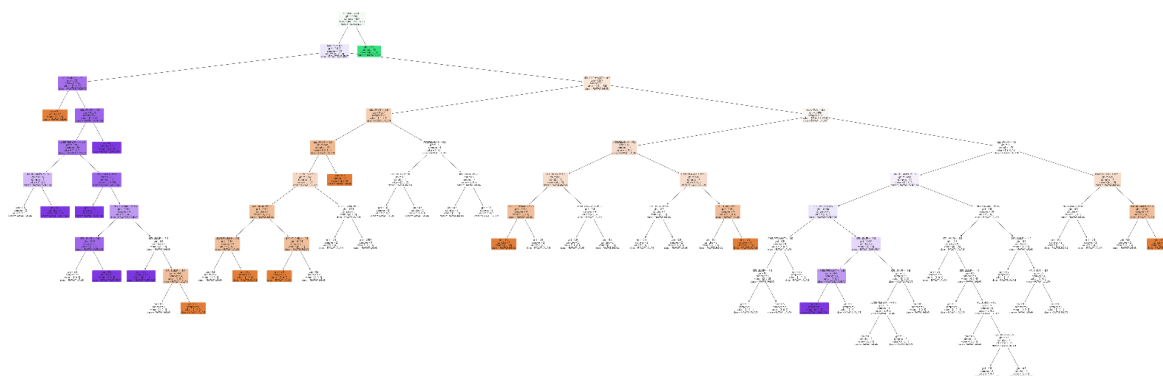
plt.show()
```

Gambar 3.34 Kode Visualisasi Decision Tree

Langkah terakhir adalah visualisasi decision tree yang telah dibuat untuk memahami bagaimana model melakukan pengambilan keputusan. Ini dilakukan dengan menggunakan modul tree dan matplotlib.pyplot dari scikit-learn untuk menggambar dan menampilkan struktur decision tree secara grafis. Visualisasi ini membantu dalam interpretasi dan pemahaman cara kerja model decision tree yang telah dibuat.

Output :

Dari data rekam medis di Rumah Sakit Islam Ibnu Sina, didapatkan pemodelan klasifikasi menggunakan algoritma *Decision Tree* seperti gambar dibawah.



Gambar 3.35 Visualisasi Decision Tree Rekam Medis

(Link Gambar: [Decision Tree Rekam Medis](#))

3.4.2. Hasil Pemodelan

Dari pemodelan decision tree yang dilakukan, dapat disimpulkan bahwa setiap variabel yang digunakan mempengaruhi hasil keputusan akhir. Kolom OPNAME berperan sebagai root node, yang menentukan arah keputusan awal berdasarkan nilai-nilai yang diambil. Setelah itu, setiap percabangan dalam pohon keputusan terjadi berdasarkan nilai dari variabel yang dipilih. Misalnya, jika nilai OPNAME tidak lebih kecil atau sama dengan 0.5, maka langkah berikutnya bergantung pada nilai KASUS. Jika nilai KASUS memenuhi suatu kondisi tertentu, maka variabel lain seperti KEL.UMUR atau JENIS PENYAKIT akan menjadi pertimbangan selanjutnya. Proses ini berlanjut sampai keputusan akhir tercapai, dimana berbagai variabel telah digunakan untuk menentukan tindakan atau klasifikasi yang tepat.

Dengan demikian, setiap variabel yang dimasukkan ke dalam model memiliki peran penting dalam menentukan hasil keputusan akhir. Variabel-variabel tersebut secara berurutan digunakan untuk membangun serangkaian pertanyaan atau kondisi yang membantu dalam mengklasifikasikan data dengan tepat sesuai dengan aturan yang ditetapkan oleh decision tree.

3.5. Perbandingan Hasil

Perbandingan antara hasil perhitungan manual dengan menggunakan 10 sampel data dan perhitungan dengan Python menggunakan 157 sampel data menunjukkan perbedaan yang signifikan, meskipun keduanya memiliki root node yang sama, yaitu OPNAME. Perhitungan manual yang hanya melibatkan 10 sampel data cenderung lebih terbatas dalam representasi variasi yang ada dalam dataset, sehingga keputusan yang diambil mungkin tidak sepenuhnya mencerminkan kompleksitas sebenarnya dari masalah. Di sisi lain, perhitungan dengan Python menggunakan 157 sampel data memberikan gambaran yang lebih luas tentang pola dan karakteristik dari dataset, karena melibatkan lebih banyak data yang mewakili berbagai kemungkinan skenario. Oleh karena itu, hasil yang dihasilkan dari perhitungan dengan Python cenderung lebih dapat diandalkan dan menggambarkan keadaan yang lebih akurat dari pada hasil perhitungan manual yang terbatas. Hal ini menekankan pentingnya menggunakan jumlah sampel data yang memadai dalam melakukan analisis dan pengambilan keputusan menggunakan teknik pemodelan seperti decision tree.

3.6. Lampiran Kode

```
import pandas as pd
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import accuracy_score
import matplotlib.pyplot as plt
from sklearn import tree

# Memuat data
data = pd.read_csv('RekamMedis.csv')
data.info()
data.head()

# Menghapus kolom yang tidak diperlukan
data = data.drop(['NO.REGISTRASI', 'NO.RM', 'UMUR', 'GOL.DARAH',
'SUKU', 'BAHASA', 'STATUS', 'ASAL PASIEN', 'KET. ASAL PASIEN',
'POLIKLINIK/RUANG', 'NO.KAMAR/TT', 'KELAS', 'CARA BAYAR',
'NO.JAMINAN', 'COVID19_NO_SEP', 'TGL.MASUK', 'TGL.PULANG', 'LOS',
'CARA KELUAR', 'KETERANGAN (CARA KELUAR)', 'KEADAAN KELUAR',
'DX.MASUK', 'DPJP', 'NO.DTD', 'DTD UTAMA', 'ICD 10.01', 'DIAGNOSIS
01', 'ICD 10.EXTERNA+S1+AG1:AK+AG1:AJ1', 'EXTERNAL CAUSES', 'NO.DTD
EXTERNAL CAUSES', 'DTD EXTERNAL CAUSES', 'ICD 9CM.01', 'PROCEDURE
01', 'HAK KELAS', 'KODE INACBG', 'DESKRIPSI INACBG', 'BIAYA RS',
'TARIFF INA CBG (HAK)', 'BAYAR (IUR)', 'SELISIH'], axis=1)
```

```

# Menghapus baris duplikat dan baris dengan nilai NaN
data = data.drop_duplicates().dropna()

# Mengubah nilai-nilai kategorikal menjadi numerik
encoder = {}
for col in data.columns:
    encoder[col] = LabelEncoder()
    data[col] = encoder[col].fit_transform(data[col])

# Memisahkan fitur dan target
X = data[['KEL.UMUR', 'JNS.KELAMIN', 'OPNAME', 'KASUS', 'JENIS
PENYAKIT']]
y = data['CARA MASUK']

# Membagi data menjadi data latih dan data uji
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.1, random_state=123)

# Membangun model decision tree
tree_model = DecisionTreeClassifier()
tree_model = tree_model.fit(X_train, y_train)

# Memprediksi target pada data uji
y_pred = tree_model.predict(X_test)

# Menghitung akurasi model
accuracy = round(accuracy_score(y_pred, y_test), 3)
print('Accuracy:', accuracy)

# Visualisasi pohon keputusan
plt.figure(figsize=(90, 30))
tree.plot_tree(tree_model, feature_names=['KEL.UMUR',
'JNS.KELAMIN', 'OPNAME', 'KASUS', 'JENIS PENYAKIT'],
class_names=['RAWAT JALAN', 'RAWAT INAP', 'RAWAT DARURAT'],
filled=True, rounded=True, fontsize=10)
plt.savefig("decision_tree.png")
plt.show()

```


BAB IV

PENUTUP

4.1. Kesimpulan

Penelitian ini berfokus pada klasifikasi rekam medis pasien di Rumah Sakit Islam Ibnu Sina menggunakan algoritma decision tree untuk memprediksi cara masuk pasien. Data rekam medis yang telah dikumpulkan mengalami proses pre-processing, termasuk data reduction dan data cleaning, untuk memastikan kualitas data yang digunakan dalam analisis.

Dari hasil pemodelan decision tree yang telah dilakukan, dapat disimpulkan bahwa setiap variabel yang digunakan dalam model berperan penting dalam menentukan keputusan akhir. Variabel OPNAME berfungsi sebagai root node, yang menjadi titik awal untuk percabangan keputusan berdasarkan nilai-nilai spesifik. Variabel-variabel lain seperti KASUS, KEL.UMUR, dan JENIS PENYAKIT turut mempengaruhi arah keputusan pada setiap percabangan hingga mencapai klasifikasi akhir.

Perbandingan antara perhitungan manual dengan menggunakan 10 sampel data dan perhitungan menggunakan Python dengan 157 sampel data menunjukkan adanya perbedaan signifikan. Perhitungan manual yang hanya menggunakan sedikit sampel data cenderung kurang representatif dan mungkin tidak mencerminkan kompleksitas dataset secara keseluruhan. Sebaliknya, perhitungan dengan menggunakan lebih banyak sampel data memberikan hasil yang lebih akurat dan dapat diandalkan karena mencakup lebih banyak variasi dan skenario.

Dengan demikian, dapat disimpulkan bahwa penggunaan dataset yang lebih besar dan beragam sangat penting dalam analisis data dan pemodelan untuk menghasilkan keputusan yang lebih tepat dan akurat. Teknik pemodelan seperti decision tree memerlukan jumlah data yang memadai untuk mengidentifikasi pola-pola yang relevan dan mendukung pengambilan keputusan yang lebih baik dalam konteks klinis dan manajemen rumah sakit.

4.2. Saran

Penelitian ini masih memiliki kekurangan yang perlu diperbaiki untuk mencapai hasil yang lebih optimal. Oleh karena itu, untuk pengembangan penelitian berikutnya, beberapa saran yang dapat dipertimbangkan adalah sebagai berikut:

1. Penggunaan Dataset yang Lebih Luas dan Bervariasi

Untuk meningkatkan keakuratan dan generalisasi model klasifikasi, disarankan untuk menggunakan dataset yang lebih besar dan bervariasi. Data dari berbagai periode waktu dan sumber yang berbeda dapat memberikan wawasan yang lebih komprehensif tentang pola perawatan pasien.

2. Peningkatan Kualitas Data

Proses preprocessing yang lebih mendetail dapat dilakukan untuk memastikan kualitas data yang lebih baik. Ini termasuk penanganan nilai yang hilang, deteksi dan penanganan outliers yang lebih cermat peningkatan akurasi data, serta verifikasi konsistensi data.

3. Penggunaan Algoritma Lain untuk Perbandingan

Selain algoritma decision tree, disarankan untuk menerapkan dan membandingkan hasil dari berbagai algoritma klasifikasi lainnya seperti Random Forest, Support Vector Machine (SVM), dan Neural Networks. Ini dapat membantu menentukan algoritma mana yang memberikan performa terbaik untuk kasus spesifik ini.

DAFTAR PUSTAKA

- Abduh, R. (2021). Kajian hukum rekam medis sebagai alat bukti malapraktik medis. *De Lega Lata: Jurnal Ilmu Hukum*, 6(1), 221-234.
- Anjaryani, W. D. (2009). *Kepuasan pasien rawat inap terhadap pelayanan perawat di RSUD Tugurejo Semarang* (Doctoral dissertation, Universitas Diponegoro).
- Erfavira, A., & Kirana, S. (2012). *Perbedaan Kelengkapan Pengisian Rekam Medis antara Instalasi Rawat Jalan dan Instalasi Rawat Darurat di Poli Bedah RSUP dr. Kariadi Semarang* (Doctoral dissertation, Fakultas Kedokteran).
- García, S., Luengo, J., & Herrera, F. (2015). Data preprocessing in data mining. *Springer*.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Jothi, N., Rashid, N. A., & Husain, W. (2015). Data mining in healthcare – A review. *Procedia Computer Science*, 72, 306-313.
- Kotsiantis, S. B., Zaharakis, I., & Pintelas, P. (2010). Supervised machine learning: A review of classification techniques. *Emerging Artificial Intelligence Applications in Computer Engineering*, 160, 3-24.
- Larose, D. T. (2005). An introduction to data mining. *Traduction et adaptation de Thierry Vallaud*.
- Lavrac, N. (1999). Selected techniques for data mining in medicine. *Artificial Intelligence in Medicine*, 16(1), 3-23.

- Maroco, J., Silva, D., Rodrigues, A., Guerreiro, M., Santana, I., de Mendonça, A., & Martins, I. (2011). Data mining methods in the prediction of Dementia: A real-data comparison of the accuracy, sensitivity and specificity of linear discriminant analysis, logistic regression, neural networks, support vector machines, classification trees and random forests. *BMC Research Notes*, 4(1), 299.
- Moertini, V. S. (2002). Data mining sebagai solusi bisnis. *Bandung*, 7(1), 44-56.
- Nguyen, L., Bellucci, E., & Nguyen, L. T. (2014). Electronic health records implementation: An evaluation of information system impact and contingency factors. *International Journal of Medical Informatics*, 83(11), 779-796.
- Patil, V. M., & Parmar, K. P. (2016). Decision tree classifier: A detailed survey. *International Journal of Information and Computational Technology*, 6(3), 164-170.
- Supartiningsih, S. (2017). Kualitas pelayanan kepuasan pasien rumah sakit: kasus pada pasien rawat jalan. *Jurnal medicoeticolegal dan manajemen rumah sakit*, 6(1), 9-15.
- Santosa, B. (2007). Data mining teknik pemanfaatan data untuk keperluan bisnis. *Yogyakarta: Graha Ilmu*, 978(979), 756.