

**REKOMENDASI RUMAH SAKIT RUJUKAN UNTUK PASIEN BERDASARKAN
DATA DIAGNOSA DAN TINGKAT KEPARAHAN MENGGUNAKAN
METODE RANDOM FOREST DI PUSKESMAS SIGAMBAL**

SKRIPSI

DISUSUN OLEH

ZAHRA AMANDA
NPM. 2209010246



UMSU

Unggul | Cerdas | Terpercaya

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

**REKOMENDASI RUMAH SAKIT RUJUKAN UNTUK PASIEN BERDASARKAN
DATA DIAGNOSA DAN TINGKAT KEPARAHAN MENGGUNAKAN
METODE RANDOM FOREST DI PUSKESMAS SIGAMBAL**

SKRIPSI

**Diajukan sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer
(S.Kom) dalam Program Studi Sistem Infromasi pada Fakultas Ilmu Komputer
dan Teknologi Informasi, Universitas Muhammadiyah Sumatera Utara**

**ZAHRA AMANDA
NPM. 2209010246**

**PROGRAM STUDI SISTEM INFORMASI
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA**

MEDAN

2026

LEMBAR PENGESAHAN

Judul Skripsi : REKOMENDASI RUMAH SAKIT RUJUKAN UNTUK PASIEN BERDASARKAN DATA DIAGNOSA DAN TINGKAT KEPARAHAN MENGGUNAKAN METODE RANDOM FOREST DI PUSKESMAS SIGAMBAL

Nama Mahasiswa : ZAHRA AMANDA

NPM : 2209010246

Program Studi : SISTEM INFORMASI

Menyetujui
Komisi Pembimbing



(Dr. Firahmi Rizky, S.Kom., M.Kom)
NIDN. 0116079201

Ketua Program Studi



(Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom)
NIDN. 0116079201

Dekan



(Dr. Al-Khowarizmi, S.Kom., M.Kom.)
NIDN. 0127099201

PERNYATAAN ORISINALITAS

**REKOMENDASI RUMAH SAKIT RUJUKAN UNTUK PASIEN
BERDASARKAN DATA DIAGNOSA DAN TINGKAT
KEPARAHAN MENGGUNAKAN METODE
RANDOM FOREST DI PUSKESMAS
SIGAMBAL**

SKRIPSI

Saya menyatakan bahwa karya tulis ini adalah hasil karya sendiri, kecuali beberapa kutipan dan ringkasan yang masing-masing disebutkan sumbernya.

Medan, Maret 2026
Yang membuat pernyataan



Zahra Amanda
NPM. 220901046

PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN
AKADEMIS

Sebagai sivitas akademika Universitas Muhammadiyah Sumatera Utara, saya bertanda tangan dibawah ini:

Nama : ZAHRA AMANDA
NPM : 2209010246
Program Studi : Sistem Informasi
Karya Ilmiah : Skripsi

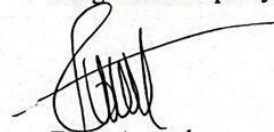
Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Universitas Muhammadiyah Sumatera Utara Hak Bebas Royalti Non-Eksekutif (*Non-Exclusive Royalty free Right*) atas penelitian skripsi saya yang berjudul:

**REKOMENDASI RUMAH SAKIT RUJUKAN UNTUK PASIEN BERDASARKAN
DATA DIAGNOSA DAN TINGKAT KEPARAHAN MENGGUNAKAN
METODE RANDOM FOREST DI PUSKESMAS SIGAMBAL**

Beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Non-Eksekutif ini, Universitas Muhammadiyah Sumatera Utara berhak menyimpan, mengalih media, memformat, mengelola dalam bentuk database, merawat dan mempublikasikan Skripsi saya ini tanpa meminta izin dari saya selama tetap mencantumkan nama saya sebagai penulis dan sebagai pemegang dan atau sebagai pemilik hak cipta.

Demikian pernyataan ini dibuat dengan sebenarnya.

Medan, Maret 2026
Yang membuat pernyataan



Zahra Amanda
NPM. 2209010246

RIWAYAT HIDUP

DATA PRIBADI

Nama Lengkap : Zahra Amanda
Tempat dan Tanggal Lahir : Sigambal, 14 Agustus 2004
Alamat Rumah : Jl. Sadikun Lestari
Telepon/Faks/HP : 082363780442
E-mail : zahraamnda254@gmail.com
Instansi Tempat Kerja : -
Alamat Kantor : -

DATA PENDIDIKAN

| | | |
|-----|-------------------------|-------------|
| SD | : SDN 112143 | TAMAT: 2016 |
| SMP | : MTsN 1 Labuhanbatu | TAMAT: 2019 |
| SMA | : SMAN 1 Rantau Selatan | TAMAT: 2022 |

KATA PENGANTAR



Alhamdulillah, puji Syukur penulis panjatkan kehadiran Allah SWT atas segala limpahan taufiq, rahmat hidayah, serta inayah-Nya, sehingga penulis dapat menyusun dan menyelesaikan laporan skripsi dengan judul “Rekomendasi Rumah Sakit Rujukan Untuk Pasien Berdasarkan Data Diagnosa dan Tingkat Keparahan Menggunakan Metode Random Forest di Puskesmas Sigambal”. Skripsi ini disusun untuk memenuhi syarat dalam menyelesaikan studi dan memperoleh gelar sarjana untuk program studi Sistem Informasi Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatra Utara.

Selama proses penelitian hingga penyusunan skripsi ini, saya telah memperoleh banyak pelajaran berharga saat menyusun laporan tugas akhir. Demikian pula, berbagai tantangan yang muncul ternyata memberikan pengalaman yang bermanfaat untuk kehidupan mendatang. Semua pencapaian ini tidak mungkin terwujud tanpa kontribusi dari orang-orang di sekitar saya, yang terus memberikan dorongan dan semangat dalam menyelesaikan laporan tugas akhir ini. Melalui momen ini, saya ingin menyampaikan rasa syukur yang mendalam serta penghormatan yang tulus kepada

1. Bapak Prof. Dr. Akrim, M.Pd., Selaku Rektor Universitas Muhammadiyah Sumatera Utara (UMSU).
2. Bapak Dr. Al-Khowarizmi, S.Kom., M.Kom. Selaku Dekan Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
3. Ibuk Dr. Firahmi Rizky, S.Kom., M.Kom., selaku Wakil Dekan I Fakultas

Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU, dan Juga Selaku Dosen Pembimbing yang telah meluangkan waktu, tenaga, dan pikiran dalam memberikan arahan, bimbingan, serta masukan yang sangat berarti selama proses penyusunan skripsi ini hingga selesai dengan baik.

4. Bapak Mhd. Basri, S.Si., M.Kom., selaku Wakil Dekan III Fakultas Ilmu Komputer dan Teknologi Informasi (FIKTI) UMSU.
5. Bapak Mahardika Abdi Prawira Tanjung, S.Kom., M.Kom., Selaku Ketua Program Studi Sistem Informasi yang selalu memberikan dukungan
6. Bapak Mulkan Azhari, S.Kom., M.Kom., selaku Sekretaris Program Studi Sistem Informasi yang telah membantu dan memberikan dukungan selama proses perkuliahan dan penyusunan skripsi.
7. Teruntuk Bapak dan Ibu Dosen (FIKTI) UMSU atas ilmu pengetahuan yang telah diberikan kepada saya selama masa perkuliahan dan seluruh Staff dan Keanggotaan Biro Kemahasiswaan yang mendukung dalam proses pengerjaan skripsi ini.
8. Kedua orang tua tercinta, Bapak Syahrial, S.Pd., M.Si. dan Ibu Suryawati, A.Md.Keb., yang selalu memberikan doa, kasih sayang, serta dukungan yang tiada henti kepada penulis. Setiap pencapaian yang penulis raih hingga saat ini, termasuk selesainya skripsi ini, tidak terlepas dari doa, pengorbanan, dan perjuangan beliau. Penulis menyampaikan terima kasih yang tulus atas segala cinta, motivasi, dan dukungan yang selalu menjadi sumber kekuatan bagi penulis dalam menyelesaikan pendidikan dan penyusunan skripsi ini.
9. Teruntuk saudara penulis, Richa Syahfitri, S.T. dan Randhi Azmi, yang

selalu menjadi penyemangat, tempat berbagi cerita, serta memberikan dukungan dan hiburan di setiap keadaan. Terima kasih atas perhatian, kebersamaan, dan dukungan yang selalu diberikan kepada penulis selama proses perkuliahan hingga penyusunan skripsi ini.

10. Teruntuk Teman-teman penulis dari perantauan, Annisa Alsadilla Zahra, Putri Salsabila, Siti Nasuha, dan Febriana Adelia, yang telah menjadi keluarga dan tempat pulang bagi penulis selama menjalani masa perantauan. Terima kasih atas kebersamaan, dukungan, serta kepedulian yang selalu diberikan dalam setiap keadaan, baik suka maupun duka selama proses penyusunan skripsi ini.

11. Teruntuk teman-teman penulis selama masa perkuliahan, Elsa Ramadhani, Nadifa Aprillia, Amanda Meisyah Putri, dan Devy Liani, yang telah memberikan kebersamaan, dukungan, serta semangat selama menjalani proses perkuliahan hingga penyusunan skripsi ini. Penulis mengucapkan terima kasih atas segala bantuan, motivasi, serta kenangan kebersamaan yang telah dilalui bersama selama masa studi.

12. Untuk diri sendiri, Zahra Amanda, terima kasih karena telah berusaha dan bertahan melalui setiap proses yang tidak selalu mudah selama perjalanan menyelesaikan pendidikan ini. Terima kasih karena tidak menyerah ketika menghadapi berbagai tantangan, rasa lelah, dan keraguan yang datang silih berganti. Setiap langkah, sekecil apa pun, telah membawa penulis sampai pada titik ini. Semoga segala usaha, kesabaran, dan perjuangan yang telah dilalui menjadi pelajaran berharga serta motivasi untuk terus berkembang, melangkah lebih jauh, dan meraih impian di masa yang akan datang.

**REKOMENDASI RUMAH SAKIT RUJUKAN UNTUK PASIEN
BERDASARKAN DATA DIAGNOSA DAN TINGKAT
KEPARAHAN MENGGUNAKAN METODE
RANDOM FOREST DI PUSKESMAS
SIGAMBAL**

ABSTRAK

Proses penentuan rumah sakit rujukan bagi pasien di fasilitas kesehatan tingkat pertama sering kali dilakukan berdasarkan pertimbangan subjektif tenaga kesehatan tanpa analisis sistematis terhadap data pasien. Padahal, data diagnosa dan tingkat keparahan pasien memiliki peran penting dalam menentukan kebutuhan rujukan yang tepat. Penelitian ini bertujuan untuk membangun sistem rekomendasi rumah sakit rujukan untuk pasien berdasarkan data diagnosa dan tingkat keparahan menggunakan metode Random Forest di Puskesmas Sigambal. Dataset yang digunakan dalam penelitian ini merupakan data riwayat rujukan pasien sebanyak 120 data dengan atribut umur, jenis kelamin, diagnosa, tingkat keparahan, dan rumah sakit rujukan sebagai variabel target. Data diagnosa dan tingkat keparahan menjadi variabel utama dalam proses klasifikasi. Tahapan penelitian meliputi preprocessing data, transformasi atribut kategorikal, pembagian data training dan testing dengan rasio 80:20, serta pembentukan model Random Forest. Hasil evaluasi model menggunakan confusion matrix menunjukkan nilai accuracy sebesar 91,4%, precision sebesar 89,8%, recall sebesar 90,5%, dan F1-score sebesar 90,1%. Hasil tersebut menunjukkan bahwa metode Random Forest mampu mengklasifikasikan rumah sakit rujukan dengan tingkat akurasi yang baik berdasarkan data diagnosa dan tingkat keparahan pasien. Dengan demikian, sistem yang dibangun dapat membantu tenaga kesehatan dalam memberikan rekomendasi rumah sakit rujukan secara lebih objektif, cepat, dan berbasis data di Puskesmas Sigambal.

Kata Kunci: Random Forest, Klasifikasi, Diagnosa, Tingkat Keparahan, Rumah Sakit Rujukan.

**REFERRAL HOSPITAL RECOMMENDATION SYSTEM
FOR PATIENTS BASED ON DIAGNOSIS DATA AND
SEVERITY LEVEL USING THE RANDOM F
OREST METHOD AT PUSKESMAS
SIGAMBAL**

ABSTRACT

The process of determining referral hospitals for patients in primary healthcare facilities is often conducted based on subjective considerations without systematic analysis of patient data. In fact, diagnosis data and severity levels play a crucial role in determining appropriate referral decisions. This study aims to develop a referral hospital recommendation system for patients based on diagnosis data and severity level using the Random Forest method at Puskesmas Sigambal. The dataset used in this study consists of 120 patient referral records, including attributes such as age, gender, diagnosis, severity level, and referral hospital as the target variable. Diagnosis and severity level serve as the main variables in the classification process. The research stages include data preprocessing, categorical data transformation, splitting the dataset into training and testing sets with an 80:20 ratio, and building the Random Forest model. The evaluation results using a confusion matrix show an accuracy of 91.4%, precision of 89.8%, recall of 90.5%, and F1-score of 90.1%. These results indicate that the Random Forest method is capable of classifying referral hospitals with good performance based on patients' diagnosis data and severity levels. Therefore, the developed system can assist healthcare personnel in providing referral hospital recommendations in a more objective, fast, and data-driven manner at Puskesmas Sigambal.

Keywords: Random Forest, Classification, Diagnosis, Severity Level, Referral Hospital.

DAFTAR ISI

| | |
|--|-------------|
| LEMBAR PENGESAHAN | i |
| PERNYATAAN ORISINALITAS | ii |
| PERNYATAAN PERSETUJUAN PUBLIKASI | iii |
| RIWAYAT HIDUP | iv |
| KATA PENGANTAR | v |
| ABSTRAK | viii |
| ABSTRACT | ix |
| DAFTAR ISI | x |
| DAFTAR TABEL | xiii |
| DAFTAR GAMBAR | xiv |
| DAFTAR RUMUS | xv |
| BAB I PENDAHULUAN | 1 |
| 1.1. Latar Belakang Masalah..... | 1 |
| 1.2. Rumusan Masalah | 3 |
| 1.3. Batasan Masalah..... | 4 |
| 1.4. Tujuan Penelitian..... | 4 |
| 1.5. Manfaat Penelitian..... | 5 |
| BAB II LANDASAN TEORI | 6 |
| 2.1. Data Mining | 6 |
| 2.2. Klasifikasi..... | 8 |
| 2.3. Machine Learning | 10 |
| 2.4. Algoritma Random Forest..... | 13 |
| 2.4.1. Konsep Dasar Decision Tree dan Struktur Simpul (Node) | 16 |
| 2.4.2. Pengukuran Ketidakmurnian Data dengan Gini Index | 16 |
| 2.4.3. Mekanisme Ensemble: Bootstrap Sampling dan Majority Voting | 17 |
| 2.4.4. Pra-pemrosesan Data dan Kategorisasi Diagnosa | 18 |
| 2.4.5. Evaluasi Model Menggunakan Confusion Matrix | 18 |
| 2.5. Diagnosa..... | 19 |
| 2.6. Tingkat Keparahan | 21 |
| 2.7. Sistem Rujukan Kesehatan..... | 23 |

| | |
|---|-----------|
| 2.8. Puskesmas Sigambal | 25 |
| 2.9. Unified Modeling Language (UML) | 27 |
| 2.10. Penelitian Terdahulu | 30 |
| 2.11. Kerangka Berpikir | 33 |
| BAB III ANALISA DAN PERANCANGAN SISTEM | 35 |
| 3.1. Analisis Permasalahan | 35 |
| 3.2. Algoritma Random Forest | 37 |
| 3.3. Flowchart Sistem | 49 |
| 3.4. Perancangan Sistem | 50 |
| 3.5. Perancangan Antarmuka | 56 |
| BAB IV IMPLEMENTASI DAN PENGUJIAN SISTEM | 59 |
| 4.1. Kebutuhan Sistem | 59 |
| 4.1.1. Kebutuhan Perangkat Keras (Hardware) | 59 |
| 4.1.2. Kebutuhan Perangkat Lunak (Software) | 60 |
| 4.2. Implementasi Sistem | 61 |
| 4.2.1. Implementasi Halaman Login | 61 |
| 4.2.2. Implementasi Halaman Dashboard | 62 |
| 4.2.3. Implementasi Halaman Input Data | 63 |
| 4.2.4. Implementasi Halaman Hasil Rekomendasi | 64 |
| 4.2.5. Implementasi Halaman Evaluasi Model | 65 |
| 4.3. Pengujian Sistem | 66 |
| 4.3.1. Pengujian Fungsional Sistem | 66 |
| 4.3.2. Pengujian Input Data Pasien | 67 |
| 4.3.3. Pengujian Proses Rekomendasi | 68 |
| 4.3.4. Pengujian Penyimpanan Data | 69 |
| 4.3.5. Kesimpulan Pengujian Sistem | 70 |
| 4.4. Evaluasi dan Analisis Model Random Forest | 70 |
| 4.4.1. Confusion Matrix | 71 |
| 4.4.2. Perhitungan Accuracy | 71 |
| 4.4.3. Perhitungan Precision | 72 |
| 4.4.4. Recall | 72 |
| 4.4.5. F1-Score | 72 |

| | |
|-----------------------------|-----------|
| BAB V PENUTUP | 76 |
| 5.1. Kesimpulan..... | 76 |
| 5.2. Saran..... | 77 |
| DAFTAR PUSTAKA | 79 |

DAFTAR TABEL

| | |
|--|----|
| Tabel 2.1 Simbol <i>Use case Diagram</i> | 29 |
| Tabel 2.2 Simbol <i>Activity Diagram</i> | 29 |
| Tabel 2.3 Simbol <i>Class Diagram</i> | 30 |
| Tabel 2.4 Penelitian Terdahulu..... | 31 |
| Tabel 3.1 Data Mentah Rujukan Pasien | 38 |
| Tabel 3.2 Pengelompokan Diagnosa | 40 |
| Tabel 3.3 Dataset Penelitian | 41 |
| Tabel 3.4 Data Sampel Perhitungan Manual | 44 |
| Tabel 4.1 Pengujian Fitur Login..... | 68 |
| Tabel 4.2 Pengujian Input Data Pasien..... | 69 |
| Tabel 4.3 Pengujian Proses Rekomendasi | 70 |
| Tabel 4.4 Pengujian Penyimpanan Data..... | 70 |

DAFTAR GAMBAR

| | |
|---|----|
| Gambar 2.1 Kerangka Berpikir | 35 |
| Gambar 3.1 Pohon keputusan Algoritma Random Forest..... | 48 |
| Gambar 3.2 Flowchart Sistem Rekomendasi Rumah Sakit Rujukan | 50 |
| Gambar 3.3 Use Case Diagram | 53 |
| Gambar 3.4 Activity Diagram | 54 |
| Gambar 3.5 Class Diagram..... | 56 |
| Gambar 3.6 Desain Halaman Login | 57 |
| Gambar 3.7 Desain Dashboard..... | 57 |
| Gambar 3.8 Desain input Data | 58 |
| Gambar 3.9 Hasil Rekomendasi | 58 |
| Gambar 3.10 Halaman Evaluasi Model..... | 59 |
| Gambar 4.1 Implementasi Halaman Login..... | 63 |
| Gambar 4.2 Implementasi Halaman Dashboard..... | 64 |
| Gambar 4.3 Implementasi Halaman Input Data | 65 |
| Gambar 4.4 Implementasi Halaman Hasil Rekomendasi | 66 |
| Gambar 4.5 Implementasi Halaman Evaluasi Model | 67 |

DAFTAR RUMUS

| | |
|--|----|
| Rumus matematis <i>Gini Index</i> | 17 |
| Rumus Menghitung Gini setelah Split..... | 47 |

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Perkembangan teknologi informasi telah membawa perubahan besar dalam cara organisasi mengelola dan memanfaatkan data. Data yang sebelumnya hanya berfungsi sebagai arsip kini dipandang sebagai sumber pengetahuan strategis. Melalui pendekatan *data mining*, data dapat diolah untuk menemukan pola, hubungan, dan informasi tersembunyi yang tidak terlihat melalui analisis konvensional. *Data mining* memungkinkan ekstraksi pengetahuan dilakukan secara sistematis dengan memanfaatkan teknik statistik dan komputasional (Kusrini & Luthfi, 2009). Seiring berkembangnya *machine learning*, berbagai metode analitik digunakan untuk membangun model prediktif, salah satunya melalui teknik klasifikasi (Prasetyo, 2012).

Dalam konteks pengelolaan informasi, klasifikasi merupakan teknik untuk mengelompokkan data ke dalam kategori tertentu berdasarkan atribut yang dimiliki. Salah satu algoritma klasifikasi yang dikenal memiliki performa unggul adalah *Random Forest*. Algoritma ini bekerja dengan mengombinasikan sejumlah pohon keputusan (*decision trees*) melalui mekanisme *ensemble learning* untuk meningkatkan akurasi prediksi dan mengurangi risiko *overfitting*. *Random Forest* juga dinilai sangat handal dalam menangani dataset medis yang memiliki banyak variabel kategori, toleran terhadap *noise*, serta menghasilkan model yang stabil (Prasetyo, 2012).

Pada institusi layanan kesehatan tingkat pertama seperti Puskesmas Sigambal, data pasien dikumpulkan secara rutin melalui proses administrasi. Data tersebut mencakup atribut krusial seperti diagnosa penyakit dan tingkat keparahan kondisi pasien. Namun, berdasarkan observasi, pemanfaatan data di Puskesmas Sigambal saat ini masih terbatas pada fungsi dokumentasi dan pelaporan administratif semata. Potensi prediktif dari data historis pasien belum dimanfaatkan secara optimal untuk mendukung proses pengambilan keputusan klinis maupun operasional.

Masalah utama muncul pada aktivitas penentuan rekomendasi rumah sakit rujukan. Proses ini memiliki peran vital dalam memastikan pasien memperoleh layanan lanjutan yang tepat dan cepat. Namun, dalam praktiknya di Puskesmas Sigambal, penentuan rujukan sering kali didasarkan pada pertimbangan subjektif, intuisi petugas, atau kebiasaan operasional tanpa adanya panduan berbasis data historis. Pendekatan subjektif ini berpotensi menimbulkan ketidakonsistenan keputusan rujukan, yang pada akhirnya dapat mengakibatkan penumpukan pasien pada rumah sakit tertentu (mismatch kapasitas) atau ketidaksesuaian antara fasilitas rumah sakit tujuan dengan kebutuhan spesifik pasien.

Padahal, variabel seperti diagnosa dan tingkat keparahan merupakan prediktor kuat dalam model klasifikasi. Dengan memanfaatkan teknik *machine learning*, data historis pasien dapat dianalisis untuk menemukan pola hubungan antara karakteristik medis pasien dengan keputusan rujukan yang paling efektif. Penggunaan *Random Forest* dalam hal ini menjadi solusi tepat karena algoritma ini mampu memproses variabel diagnosa yang beragam dan memberikan hasil

keputusan berdasarkan *majority voting* dari banyak pohon keputusan, sehingga rekomendasinya lebih objektif dan akurat.

Meskipun penelitian mengenai *Random Forest* telah banyak dilakukan, penerapannya secara spesifik untuk sistem rekomendasi rujukan di tingkat puskesmas dengan mengintegrasikan variabel tingkat keparahan masih relatif terbatas. Kondisi ini menunjukkan adanya kesenjangan (*gap*) antara ketersediaan data dengan teknologi pengolahannya di fasilitas kesehatan primer.

Berdasarkan uraian tersebut, diperlukan suatu pendekatan analitik yang mampu mengoptimalkan pemanfaatan data pasien menjadi informasi rujukan yang presisi. Oleh karena itu, penelitian ini mengusulkan judul: “Rekomendasi Rumah Sakit Rujukan Untuk Pasien Berdasarkan Data Diagnosa Dan Tingkat Keparahan Menggunakan Metode Random Forest Di Puskesmas Sigambal”. Penelitian ini diharapkan dapat memberikan kontribusi nyata dalam bidang Sistem Informasi, khususnya dalam pengembangan model prediktif yang mendukung efektivitas layanan kesehatan.

1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah diuraikan, maka rumusan masalah dalam penelitian ini adalah:

1. Bagaimana mengolah data historis pasien di Puskesmas Sigambal agar dapat digunakan sebagai dasar pengambilan keputusan rujukan yang objektif berdasarkan atribut diagnosa dan tingkat keparahan?
2. Bagaimana mengimplementasikan algoritma *Random Forest* dalam membangun model klasifikasi untuk menghasilkan rekomendasi rumah sakit rujukan?

3. Bagaimana tingkat akurasi dan kinerja model *Random Forest* dalam memberikan rekomendasi rumah sakit rujukan yang tepat bagi pasien di Puskesmas Sigambal?

1.3. Batasan Masalah

Agar penelitian lebih terarah dan sesuai dengan tujuan yang telah ditetapkan, maka batasan masalah dalam penelitian ini adalah sebagai berikut:

1. Data yang digunakan terbatas pada data pasien yang telah dianonimkan dari Puskesmas Sigambal, dengan atribut utama berupa diagnosa dan tingkat keparahan, sesuai periode dan ketersediaan data selama penelitian.
2. Proses analisis data dilakukan menggunakan pendekatan data mining dengan teknik klasifikasi, dengan algoritma yang digunakan hanya Random Forest, tanpa perbandingan dengan metode klasifikasi lainnya.
3. Penelitian difokuskan pada pembangunan model klasifikasi untuk menghasilkan rekomendasi rumah sakit rujukan, tanpa mempertimbangkan faktor eksternal seperti jarak geografis, kapasitas rumah sakit secara real-time, maupun analisis klinis medis secara mendalam.

1.4. Tujuan Penelitian

Berdasarkan rumusan masalah yang telah ditetapkan, maka tujuan dari penelitian ini adalah:

1. Melakukan analisis terhadap data pasien di Puskesmas Sigambal, khususnya data diagnosa dan tingkat keparahan, serta mengidentifikasi

pemanfaatannya sebagai atribut yang relevan dalam proses klasifikasi berbasis data mining.

2. Membangun model klasifikasi menggunakan algoritma Random Forest dengan memanfaatkan data diagnosa dan tingkat keparahan pasien, sehingga model yang dihasilkan mampu memberikan rekomendasi rumah sakit rujukan secara sistematis dan berbasis data.
3. Menguji serta mengevaluasi kinerja model Random Forest menggunakan metrik evaluasi klasifikasi untuk mengetahui tingkat akurasi dan keandalan model dalam menghasilkan rekomendasi rumah sakit rujukan.

1.5. Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini adalah sebagai berikut:

1. Penelitian ini diharapkan dapat memberikan kontribusi akademik dalam pengembangan keilmuan Sistem Informasi, khususnya pada penerapan teknik data mining dan algoritma Random Forest dalam pembangunan model klasifikasi.
2. Penelitian ini diharapkan dapat menjadi sarana bagi penulis untuk mengimplementasikan serta memperdalam pemahaman terkait data mining, machine learning, dan proses analisis data dalam penyelesaian permasalahan nyata.
3. Penelitian ini diharapkan dapat memberikan alternatif solusi berbasis teknologi dalam menghasilkan rekomendasi rumah sakit rujukan yang lebih sistematis, objektif, dan berbasis data di Puskesmas Sigambal.

BAB II

LANDASAN TEORI

2.1. Data Mining

Data mining merupakan salah satu cabang ilmu dalam bidang teknologi informasi yang berfokus pada proses penggalian pengetahuan dari sekumpulan data. Meningkatnya volume data yang dihasilkan dari berbagai aktivitas organisasi menuntut adanya metode yang mampu mengolah data secara efektif sehingga dapat menghasilkan informasi yang bermakna. Data yang tersimpan dalam basis data tidak selalu memberikan manfaat secara langsung apabila tidak dianalisis dengan teknik yang tepat. Oleh karena itu, data mining berkembang sebagai solusi untuk mengekstraksi pola, hubungan, dan kecenderungan tertentu dari data berukuran besar.

Data mining merupakan proses pencarian pola atau informasi yang menarik dari data terpilih dengan menggunakan teknik atau metode tertentu. Konsep ini menegaskan bahwa data mining tidak sekadar berfokus pada pengolahan data, tetapi juga berorientasi pada penemuan pengetahuan baru yang memiliki nilai guna. Selain itu, data mining merupakan bagian dari proses Knowledge Discovery in Databases (KDD), yang mencakup tahapan seleksi data, pembersihan data, transformasi data, proses penambangan data, serta interpretasi hasil. Dengan demikian, data mining dipahami sebagai proses yang sistematis dan terstruktur dalam menghasilkan informasi yang bermakna dari data (Kusrini & Luthfi, 2009; Prasetyo, 2012).

Data mining kombinasi berbagai disiplin ilmu, seperti statistika, machine learning, kecerdasan buatan, dan manajemen basis data (Larose, 2005). Integrasi berbagai disiplin ini memungkinkan data mining digunakan untuk menyelesaikan permasalahan analitik yang kompleks. Dalam praktiknya, data mining digunakan

untuk mengidentifikasi pola tersembunyi, membangun model prediktif, serta menghasilkan pengetahuan yang dapat mendukung kebutuhan organisasi (Gorunescu, 2011).

Tahapan dalam data mining umumnya mengikuti model proses tertentu agar analisis berjalan secara sistematis. Salah satu model yang banyak digunakan adalah CRISP-DM (Cross Industry Standard Process for Data Mining). Model ini terdiri dari enam tahapan utama, yaitu pemahaman bisnis (business understanding), pemahaman data (data understanding), persiapan data (data preparation), pemodelan (modeling), evaluasi (evaluation), dan implementasi (deployment) (Prasetyo, 2012). Pendekatan ini membantu memastikan bahwa setiap proses analisis dilakukan secara terarah sesuai tujuan penelitian.

Fungsi utama dalam data mining meliputi beberapa teknik analitik, di antaranya klasifikasi, klustering, asosiasi, dan prediksi. Klasifikasi digunakan untuk memetakan data ke dalam kategori tertentu, klustering untuk mengelompokkan data berdasarkan kemiripan karakteristik, asosiasi untuk menemukan hubungan antar variabel, dan prediksi untuk memperkirakan nilai di masa mendatang (Kusrini & Luthfi, 2009). Pemilihan teknik yang digunakan sangat bergantung pada tujuan analisis dan karakteristik data.

Dalam konteks penelitian ini, data mining digunakan sebagai pendekatan untuk mengolah data pasien yang tersedia di Puskesmas Sigambal. Data berupa diagnosa dan tingkat keparahan pasien memiliki potensi untuk dianalisis menggunakan teknik klasifikasi. Melalui proses ini, data historis dapat dimanfaatkan untuk menemukan pola hubungan antara atribut pasien dan

rekomendasi rumah sakit rujukan. Pendekatan data mining memungkinkan analisis dilakukan secara objektif, sistematis, dan berbasis data.

Penerapan data mining dalam lingkungan organisasi memberikan berbagai manfaat, antara lain meningkatkan kualitas analisis data, membantu identifikasi pola yang tidak terlihat secara langsung, serta mendukung pembangunan model prediktif (Gorunescu, 2011). Oleh karena itu, data mining menjadi salah satu pendekatan yang relevan dalam pengembangan solusi berbasis data pada bidang Sistem Informasi.

Berdasarkan uraian tersebut, dapat disimpulkan bahwa data mining merupakan proses analitik yang berperan penting dalam mengubah data mentah menjadi pengetahuan yang bernilai. Dalam penelitian ini, data mining digunakan untuk mendukung pembangunan model klasifikasi menggunakan algoritma Random Forest guna menghasilkan rekomendasi rumah sakit rujukan berdasarkan data diagnosa dan tingkat keparahan pasien.

2.2. Klasifikasi

Klasifikasi merupakan salah satu teknik utama dalam data mining yang digunakan untuk mengelompokkan data ke dalam kelas atau kategori tertentu berdasarkan atribut yang dimiliki. Teknik ini bertujuan membangun sebuah model yang mampu mengenali pola dari data historis sehingga dapat digunakan untuk memprediksi kelas pada data baru. Dalam praktiknya, klasifikasi banyak digunakan karena mampu menghasilkan analisis yang sistematis dan berbasis data, terutama pada permasalahan yang membutuhkan keputusan prediktif (Kusrini & Luthfi, 2009).

Dalam konteks pembelajaran mesin, klasifikasi termasuk ke dalam pendekatan *supervised learning*, yaitu metode pembelajaran yang menggunakan

data latih yang telah memiliki label kelas. Proses klasifikasi dilakukan dengan mempelajari hubungan antara variabel input dan variabel target sehingga model yang dihasilkan mampu melakukan prediksi secara otomatis. Pendekatan ini dinilai efektif karena model belajar langsung dari contoh data yang tersedia (Prasetyo, 2012).

Proses klasifikasi umumnya terdiri dari dua tahap utama, yaitu tahap pembentukan model (*training*) dan tahap pengujian model (*testing*). Pada tahap training, algoritma mempelajari pola dari data historis, sedangkan pada tahap testing, model diuji menggunakan data yang belum pernah dianalisis sebelumnya. Tahap pengujian ini bertujuan untuk mengetahui sejauh mana model mampu melakukan generalisasi terhadap data baru dan tidak hanya bekerja baik pada data latih (Suyanto, 2018).

Klasifikasi memiliki peranan penting dalam analisis prediktif karena mampu mengidentifikasi kecenderungan tertentu berdasarkan pola data. Teknik ini banyak diterapkan pada berbagai bidang, seperti prediksi perilaku pengguna, analisis risiko, penentuan prioritas, hingga sistem rekomendasi. Keunggulan klasifikasi terletak pada kemampuannya dalam menangani data berukuran besar serta menghasilkan model yang dapat digunakan secara berulang (Santoso, 2017).

Berbagai algoritma dapat digunakan dalam proses klasifikasi, antara lain Decision Tree, Naïve Bayes, K-Nearest Neighbor, Support Vector Machine, serta Random Forest. Setiap algoritma memiliki karakteristik yang berbeda dalam hal kompleksitas, akurasi, serta kebutuhan komputasi. Oleh karena itu, pemilihan algoritma klasifikasi perlu disesuaikan dengan karakteristik data dan tujuan

penelitian yang ingin dicapai (Witten, Frank, & Hall, 2011; diadaptasi dalam literatur Indonesia).

Evaluasi model klasifikasi dilakukan untuk mengetahui tingkat kinerja model yang dibangun. Salah satu metode evaluasi yang paling umum digunakan adalah confusion matrix, yang menghasilkan beberapa metrik pengukuran, seperti akurasi, precision, recall, dan F1-score. Akurasi menunjukkan tingkat ketepatan prediksi secara keseluruhan, precision menggambarkan ketepatan prediksi pada kelas tertentu, recall menunjukkan kemampuan model dalam mengenali kelas target, sedangkan F1-score merupakan nilai keseimbangan antara precision dan recall (Kusrini & Luthfi, 2009; Prasetyo, 2012).

Dalam penelitian ini, teknik klasifikasi digunakan untuk membangun model prediktif dalam menghasilkan rekomendasi rumah sakit rujukan. Data pasien berupa diagnosa dan tingkat keparahan digunakan sebagai atribut input dalam model klasifikasi. Pendekatan ini memungkinkan pemanfaatan data historis dilakukan secara lebih objektif dan terstruktur, sehingga rekomendasi yang dihasilkan tidak hanya berdasarkan pertimbangan subjektif, tetapi juga didukung oleh analisis data (Suyanto, 2018).

Berdasarkan uraian tersebut, klasifikasi dapat dipahami sebagai teknik data mining yang berfungsi membangun model prediktif berbasis data historis. Teknik ini relevan digunakan dalam penelitian yang berorientasi pada rekomendasi karena mampu menghasilkan keputusan yang konsisten, terukur, dan berbasis pola data.

2.3. Machine Learning

Machine learning merupakan salah satu cabang dari kecerdasan buatan (*artificial intelligence*) yang berfokus pada pengembangan sistem yang mampu mempelajari pola dari data. Pendekatan ini memungkinkan komputer melakukan proses pembelajaran secara

otomatis tanpa harus diprogram secara eksplisit untuk setiap aturan yang digunakan. Machine learning bekerja dengan memanfaatkan data historis sebagai dasar pembentukan model yang dapat digunakan untuk prediksi maupun klasifikasi (Suyanto, 2018).

Konsep utama dalam machine learning adalah kemampuan sistem untuk belajar dari pengalaman, yaitu melalui data yang diberikan selama proses pelatihan (*training*). Data tersebut digunakan untuk menemukan pola atau hubungan tertentu yang selanjutnya direpresentasikan dalam bentuk model. Model inilah yang kemudian digunakan untuk menganalisis data baru. Pendekatan ini menjadikan machine learning relevan dalam pengolahan data modern yang memiliki volume besar dan kompleksitas tinggi (Prasetyo, 2014).

Machine learning secara umum dibagi menjadi beberapa kategori, yaitu supervised learning, unsupervised learning, dan reinforcement learning. Supervised learning merupakan pendekatan pembelajaran yang menggunakan data latih berlabel, sehingga model belajar mengenali hubungan antara variabel input dan output. Unsupervised learning digunakan pada data tanpa label untuk menemukan pola tersembunyi seperti pengelompokan (*clustering*). Reinforcement learning menitikberatkan pada proses pembelajaran melalui interaksi sistem dengan lingkungan (Suyanto, 2018).

Dalam konteks data mining, supervised learning sering digunakan karena mampu menghasilkan model prediktif yang dapat diterapkan pada berbagai kasus analitik. Teknik klasifikasi termasuk dalam supervised learning, di mana model dibangun untuk memetakan data ke dalam kelas tertentu. Keunggulan pendekatan ini terletak pada kemampuannya menghasilkan prediksi berbasis pola data historis (Kusrini & Luthfi, 2009).

Penerapan machine learning memberikan berbagai manfaat dalam pengolahan data, antara lain meningkatkan akurasi analisis, mempercepat proses identifikasi pola, serta mengurangi ketergantungan pada proses manual. Machine learning juga memungkinkan pengembangan sistem yang adaptif terhadap perubahan pola data, sehingga model dapat diperbarui sesuai dinamika data yang tersedia (Santoso, 2017).

Berbagai algoritma machine learning dikembangkan untuk menyelesaikan permasalahan klasifikasi, seperti Decision Tree, Naïve Bayes, Support Vector Machine, K-Nearest Neighbor, serta Random Forest. Setiap algoritma memiliki pendekatan berbeda dalam membangun model, baik dari sisi kompleksitas, kebutuhan komputasi, maupun tingkat akurasi yang dihasilkan. Pemilihan algoritma perlu mempertimbangkan karakteristik data dan tujuan analisis yang ingin dicapai (Suyanto, 2018).

Dalam penelitian ini, machine learning digunakan sebagai pendekatan utama dalam membangun model klasifikasi. Data pasien berupa diagnosa dan tingkat keparahan diperlakukan sebagai atribut input untuk membentuk model prediktif menggunakan algoritma Random Forest. Pendekatan ini memungkinkan analisis dilakukan secara sistematis, objektif, serta berbasis pola data historis.

Berdasarkan uraian tersebut, machine learning dapat dipahami sebagai pendekatan pembelajaran berbasis data yang memungkinkan sistem menghasilkan model prediktif secara otomatis. Pendekatan ini relevan digunakan dalam penelitian berbasis data mining karena mampu mengolah data menjadi informasi yang bernilai.

2.4. Algoritma Random Forest

Random Forest merupakan salah satu algoritma dalam machine learning yang dikembangkan untuk meningkatkan performa model klasifikasi maupun regresi. Algoritma ini diperkenalkan oleh Leo Breiman sebagai metode berbasis ensemble learning, yaitu pendekatan yang menggabungkan sejumlah model pembelajaran untuk memperoleh hasil prediksi yang lebih akurat dan stabil. Random Forest dibangun atas dasar konsep bahwa kombinasi beberapa model yang beragam mampu menghasilkan performa yang lebih baik dibandingkan model tunggal (Breiman, 2001).

Secara konseptual, Random Forest merupakan pengembangan dari algoritma Decision Tree. Decision Tree dikenal sebagai metode yang mudah dipahami dan mampu merepresentasikan aturan keputusan secara jelas, namun metode ini memiliki kelemahan berupa kecenderungan mengalami *overfitting*. Overfitting terjadi ketika model terlalu menyesuaikan diri dengan data latih sehingga kinerjanya menurun saat diterapkan pada data baru. Random Forest hadir untuk mengatasi kelemahan tersebut dengan membangun banyak pohon keputusan yang bekerja secara independen, kemudian menggabungkan hasil prediksi masing-masing pohon (Han et al., 2012).

Proses pembentukan Random Forest melibatkan dua unsur utama, yaitu *bootstrap sampling* dan *random feature selection*. *Bootstrap sampling* merupakan teknik pengambilan sampel data secara acak dengan pengembalian dari dataset pelatihan, sehingga setiap pohon keputusan dilatih menggunakan subset data yang berbeda. Sementara itu, *random feature selection* dilakukan dengan memilih atribut secara acak pada setiap node pohon untuk menentukan pemisahan terbaik.

Kombinasi kedua teknik ini bertujuan menciptakan variasi model sehingga korelasi antar pohon dapat dikurangi (Breiman, 2001).

Random Forest bekerja melalui beberapa tahapan, dimulai dari pembentukan sejumlah pohon keputusan berdasarkan data latih. Setiap pohon melakukan proses klasifikasi terhadap data uji dan menghasilkan prediksi kelas. Selanjutnya, seluruh hasil prediksi digabungkan menggunakan mekanisme *majority voting*, di mana kelas dengan jumlah suara terbanyak ditetapkan sebagai hasil akhir. Pendekatan agregasi ini menjadikan Random Forest lebih stabil karena keputusan tidak bergantung pada satu pohon saja (Suyanto, 2018).

Keunggulan Random Forest terletak pada kemampuannya menangani data dengan kompleksitas tinggi. Algoritma ini mampu bekerja baik pada data dengan jumlah atribut yang besar, toleran terhadap *noise*, serta relatif tidak sensitif terhadap perubahan kecil pada dataset. Selain itu, Random Forest memiliki kemampuan dalam mengukur tingkat kepentingan atribut (*feature importance*), sehingga dapat membantu dalam memahami kontribusi variabel terhadap model (Prasetyo, 2014).

Dalam perspektif Sistem Informasi, pemilihan algoritma tidak hanya mempertimbangkan tingkat akurasi, tetapi juga stabilitas model dan kemampuannya untuk diintegrasikan ke dalam sistem aplikasi yang berjalan. Penerapan metode data mining dalam sistem pendukung keputusan harus mampu menghasilkan rekomendasi yang konsisten dan dapat diimplementasikan secara praktis dalam lingkungan organisasi (Syahra, 2023). Selain itu, pengembangan sistem berbasis analitik perlu memperhatikan aspek integrasi antara model dan aplikasi agar sistem dapat dimanfaatkan secara optimal oleh pengguna (Ginting et al., 2024). Oleh karena itu, Random Forest dinilai sesuai untuk diterapkan dalam

penelitian ini karena mampu memberikan keseimbangan antara performa klasifikasi dan stabilitas implementasi dalam sistem berbasis web.

Kinerja Random Forest dipengaruhi oleh beberapa parameter penting, di antaranya jumlah pohon (*number of trees*), jumlah atribut yang dipilih secara acak (*number of features*), serta kedalaman pohon (*maximum depth*). Jumlah pohon yang memadai dapat meningkatkan stabilitas model, namun jumlah yang terlalu besar dapat meningkatkan beban komputasi. Oleh karena itu, pemilihan parameter perlu dilakukan secara proporsional untuk memperoleh keseimbangan antara akurasi dan efisiensi model (Suyanto, 2018).

Evaluasi model Random Forest dilakukan untuk mengetahui tingkat performa model yang dihasilkan. Metode evaluasi yang umum digunakan adalah *confusion matrix*, yang menghasilkan metrik seperti akurasi, precision, recall, dan F1-score. Evaluasi ini penting untuk memastikan bahwa model memiliki tingkat keandalan yang memadai sebelum diimplementasikan dalam sistem (Ginting et al., 2024; Han et al., 2012).

Dalam konteks penelitian ini, Random Forest digunakan sebagai algoritma utama dalam membangun model klasifikasi untuk menghasilkan rekomendasi rumah sakit rujukan. Data pasien berupa diagnosa dan tingkat keparahan digunakan sebagai atribut input dalam model. Dengan memanfaatkan data historis pasien, Random Forest diharapkan mampu mengidentifikasi pola hubungan antara karakteristik pasien dan keputusan rujukan yang sesuai secara lebih objektif dan sistematis.

Berdasarkan uraian tersebut, Random Forest dapat dipahami sebagai algoritma klasifikasi berbasis *ensemble learning* yang efektif dalam meningkatkan

akurasi prediksi serta mengurangi risiko overfitting. Karakteristik tersebut menjadikan Random Forest relevan digunakan dalam penelitian berbasis data mining dan sistem pendukung keputusan, khususnya dalam pembangunan model rekomendasi rumah sakit rujukan berdasarkan data diagnosa dan tingkat keparahan pasien.

2.4.1. Konsep Dasar Decision Tree dan Struktur Simpul (Node)

Sebelum memahami mekanisme *Random Forest*, perlu dipahami konsep *Decision Tree* sebagai unit penyusun utamanya. *Decision Tree* adalah model klasifikasi yang merepresentasikan aturan keputusan dalam struktur hirarki pohon. Sesuai dengan kebutuhan analisis pada penelitian ini, setiap pohon dalam *Random Forest* memiliki anatomi simpul (*node*) sebagai berikut:

1. **Simpul Akar (*Root Node*):** Simpul paling atas yang mewakili seluruh sampel data dan merupakan titik awal pemilihan atribut terbaik untuk membagi data.
2. **Simpul Internal (*Internal Node*):** Simpul perantara yang merepresentasikan hasil pengujian terhadap atribut (seperti Diagnosa Kategori atau Tingkat Keparahan). Setiap simpul ini memiliki cabang yang mengarahkan data ke simpul berikutnya.
3. **Simpul Daun (*Leaf Node*):** Simpul akhir yang memuat label kelas atau keputusan final (dalam hal ini, nama Rumah Sakit Rujukan). Simpul ini tidak memiliki cabang lagi (Han et al., 2012).

2.4.2. Pengukuran Ketidakmurnian Data dengan Gini Index

Dalam menentukan atribut mana yang paling tepat untuk dijadikan simpul (*node*), algoritma menggunakan metrik pengukuran ketidakmurnian (*impurity*).

Salah satu metrik yang paling sering digunakan dalam *Random Forest* adalah *Gini Index*. *Gini Index* mengukur probabilitas variabel yang dipilih secara acak akan diklasifikasikan secara salah. Rumus matematis *Gini Index* adalah:

$$Gini(S) = 1 - \sum_{i=1}^n (p_i)^2. \quad \dots\dots\dots (2.1)$$

Keterangan:

- *Gini(S)*: Nilai ketidakmurnian dataset pada simpul tersebut.
- *p_i*: Proporsi atau probabilitas kemunculan kelas ke- *i* dalam dataset.

Pemisahan data (*splitting*) yang optimal terjadi ketika nilai Gini mencapai angka terkecil (mendekati 0), yang menandakan bahwa data pada simpul tersebut sudah murni atau homogen (Breiman, 2001).

2.4.3. Mekanisme Ensemble: Bootstrap Sampling dan Majority Voting

Kekuatan utama *Random Forest* terletak pada dua mekanisme yang menjaga variansi model:

1. Bootstrap Sampling (Bagging): Setiap pohon dalam "hutan" dilatih menggunakan sampel acak yang diambil dari dataset asli dengan pengembalian. Hal ini memastikan bahwa setiap pohon memiliki karakteristik yang berbeda satu sama lain.
2. Majority Voting: Setelah seluruh pohon memberikan hasil klasifikasi masing-masing, *Random Forest* akan melakukan agregasi. Keputusan akhir rujukan rumah sakit ditentukan berdasarkan suara terbanyak (*majority vote*) dari seluruh pohon yang terbentuk (Suyanto, 2018).

2.4.4. Pra-pemrosesan Data dan Kategorisasi Diagnosa

Untuk menghasilkan model yang akurat, data mentah harus melalui tahap pra-pemrosesan. Dalam konteks data medis Puskesmas, hal ini mencakup:

1. **Data Cleaning:** Menangani redundansi atau ketidakkonsistenan data diagnosa asli.
2. **Mapping Diagnosa:** Mengubah diagnosa asli yang sangat spesifik (berdasarkan kode ICD-10) menjadi Diagnosa Kategori yang lebih umum. Hal ini dilakukan untuk mengurangi kompleksitas variabel (*dimensionality reduction*) sehingga simpul pada pohon keputusan dapat mengenali pola rujukan dengan lebih stabil (Prasetyo, 2014).
3. **Label Encoding:** Transformasi data kategori (seperti jenis diagnosa) menjadi nilai numerik agar dapat dihitung melalui rumus *Gini Index*.

2.4.5. Evaluasi Model Menggunakan Confusion Matrix

Kinerja model yang dibangun perlu diuji untuk memastikan keandalannya. Metode evaluasi yang digunakan adalah *Confusion Matrix*, yang menghasilkan beberapa metrik penting seperti:

- **Akurasi (*Accuracy*):** Mengukur sejauh mana model benar dalam mengklasifikasikan rumah sakit rujukan secara keseluruhan.
- **Presisi (*Precision*):** Mengukur ketepatan antara hasil prediksi rujukan dengan data aktual.
- **Recall:** Mengukur kemampuan model dalam menemukan kembali informasi rujukan yang benar dari keseluruhan data aktual (Ginting et al., 2024).

2.5. Diagnosa

Diagnosa merupakan hasil identifikasi kondisi pasien yang diperoleh melalui proses pemeriksaan medis. Dalam praktik pelayanan kesehatan, diagnosa menjadi dasar utama dalam menentukan tindakan, terapi, maupun rujukan lanjutan yang diperlukan oleh pasien. Diagnosa menggambarkan jenis gangguan atau penyakit yang dialami pasien berdasarkan gejala, hasil pemeriksaan fisik, serta pemeriksaan penunjang yang dilakukan oleh tenaga medis (Kementerian Kesehatan Republik Indonesia, 2014b).

Secara konseptual, diagnosa tidak hanya berfungsi sebagai catatan medis, tetapi juga sebagai informasi penting yang merepresentasikan kondisi pasien secara terstruktur. Data diagnosa biasanya dicatat dalam sistem informasi kesehatan menggunakan standar tertentu agar dapat dipahami secara konsisten. Keberadaan data diagnosa memungkinkan institusi layanan kesehatan melakukan dokumentasi riwayat penyakit pasien, analisis kasus, serta evaluasi kualitas layanan yang diberikan (Kementerian Kesehatan Republik Indonesia, 2014b).

Dalam konteks pengelolaan data, diagnosa termasuk ke dalam atribut yang memiliki nilai analitik tinggi. Informasi diagnosa mampu mencerminkan karakteristik kasus pasien, pola penyakit yang sering muncul, serta kebutuhan layanan lanjutan. Oleh karena itu, data diagnosa sering dimanfaatkan dalam penelitian berbasis data mining karena mampu menggambarkan kondisi utama yang berkaitan dengan keputusan administratif layanan (Kusrini & Luthfi, 2009).

Pemanfaatan data diagnosa dalam analisis berbasis machine learning memungkinkan identifikasi pola hubungan antara kondisi pasien dan keputusan layanan. Diagnosa dapat diperlakukan sebagai variabel input dalam model

klasifikasi untuk memprediksi kategori tertentu, seperti tingkat risiko, jenis tindakan, maupun rekomendasi rujukan. Pendekatan ini memberikan peluang untuk mengoptimalkan pemanfaatan data historis pasien secara sistematis dan berbasis data (Suyanto, 2018).

Namun demikian, dalam praktiknya, data diagnosa sering kali hanya digunakan sebagai arsip atau dokumentasi administratif. Potensi diagnosa sebagai sumber analitik belum sepenuhnya dimanfaatkan, terutama dalam pembangunan model prediktif. Padahal, data diagnosa memiliki peran penting dalam merepresentasikan kondisi pasien yang menjadi dasar berbagai keputusan layanan (Jogiyanto, 2014).

Dalam penelitian ini, diagnosa digunakan sebagai salah satu variabel utama dalam proses klasifikasi. Data diagnosa merepresentasikan kondisi pasien yang menjadi dasar analisis model Random Forest. Dengan memanfaatkan data historis pasien, model klasifikasi diharapkan mampu mempelajari pola hubungan antara jenis diagnosa dan rekomendasi rumah sakit rujukan yang sesuai.

Penggunaan variabel diagnosa dalam penelitian berbasis data mining dinilai relevan karena mampu menggambarkan karakteristik utama pasien secara terstruktur. Variabel ini berperan sebagai atribut input yang berkontribusi dalam pembentukan model prediktif (Kusrini & Luthfi, 2009).

Berdasarkan uraian tersebut, diagnosa dapat dipahami sebagai informasi penting yang merepresentasikan kondisi pasien dan memiliki peran signifikan dalam berbagai keputusan layanan kesehatan. Dalam konteks penelitian ini, data diagnosa dimanfaatkan sebagai atribut dalam model klasifikasi untuk menghasilkan rekomendasi rumah sakit rujukan secara sistematis, objektif, dan berbasis data.

2.6. Tingkat Keparahan

Tingkat keparahan merupakan indikator yang digunakan untuk menggambarkan kondisi pasien berdasarkan tingkat urgensi dan kompleksitas penanganan yang dibutuhkan. Dalam praktik pelayanan kesehatan, tingkat keparahan berperan sebagai acuan dalam menentukan prioritas layanan, jenis tindakan yang diberikan, serta kebutuhan penanganan lanjutan. Variabel ini mencerminkan sejauh mana kondisi pasien memerlukan perhatian khusus, baik dalam bentuk perawatan segera maupun observasi lanjutan (Kementerian Kesehatan Republik Indonesia, 2014b).

Secara konseptual, tingkat keparahan umumnya diklasifikasikan ke dalam beberapa kategori, seperti ringan, sedang, dan berat. Kategori tersebut digunakan untuk mempermudah proses pengambilan keputusan layanan, khususnya dalam menentukan tingkat urgensi tindakan. Tingkat keparahan tidak hanya memiliki makna klinis, tetapi juga memiliki fungsi administratif karena sering digunakan dalam sistem pencatatan layanan pasien (Kementerian Kesehatan Republik Indonesia, 2014b).

Dalam konteks pengelolaan data, tingkat keparahan merupakan atribut yang memiliki nilai informasi tinggi. Variabel ini mampu merepresentasikan kondisi pasien secara terstruktur, sehingga dapat dimanfaatkan dalam analisis berbasis data mining. Tingkat keparahan menggambarkan intensitas kondisi pasien yang berkaitan langsung dengan kebutuhan layanan lanjutan, termasuk keputusan rujukan. Oleh karena itu, variabel ini relevan digunakan sebagai atribut input dalam model klasifikasi (Kusrini & Luthfi, 2009).

Pemanfaatan tingkat keparahan dalam pendekatan machine learning memungkinkan identifikasi pola hubungan antara kondisi pasien dan keputusan layanan. Variabel ini dapat digunakan bersama atribut lain, seperti diagnosa, untuk membangun model prediktif yang mampu menghasilkan rekomendasi atau klasifikasi tertentu. Dalam model klasifikasi, tingkat keparahan membantu algoritma memahami variasi tingkat urgensi kondisi pasien (Suyanto, 2018).

Namun demikian, dalam praktik operasional, data tingkat keparahan sering kali hanya digunakan sebagai informasi pendukung dalam dokumentasi layanan. Potensi variabel ini sebagai sumber analitik belum dimanfaatkan secara optimal. Padahal, tingkat keparahan memiliki peran penting dalam merepresentasikan kondisi pasien yang dapat memengaruhi berbagai keputusan administratif layanan (Jogiyanto, 2014).

Dalam penelitian ini, tingkat keparahan digunakan sebagai salah satu variabel utama dalam proses klasifikasi menggunakan algoritma Random Forest. Variabel ini dipadukan dengan data diagnosa untuk membentuk atribut input model. Dengan memanfaatkan data historis pasien, model diharapkan mampu mempelajari pola hubungan antara tingkat keparahan pasien dan rekomendasi rumah sakit rujukan yang sesuai.

Penggunaan tingkat keparahan dalam penelitian berbasis data mining dinilai relevan karena mampu menggambarkan kondisi pasien secara sistematis dan terstruktur. Variabel ini berperan sebagai atribut penting yang berkontribusi dalam pembentukan model prediktif. Dengan demikian, tingkat keparahan tidak hanya berfungsi sebagai indikator kondisi pasien, tetapi juga sebagai sumber data yang bernilai dalam analisis prediktif (Kusrini & Luthfi, 2009).

Berdasarkan uraian tersebut, tingkat keparahan dapat dipahami sebagai variabel yang menggambarkan intensitas kondisi pasien berdasarkan tingkat urgensi penanganan. Dalam konteks penelitian ini, data tingkat keparahan dimanfaatkan sebagai atribut dalam model klasifikasi Random Forest untuk menghasilkan rekomendasi rumah sakit rujukan secara sistematis, objektif, dan berbasis data.

2.7. Sistem Rujukan Kesehatan

Sistem rujukan kesehatan merupakan mekanisme pelayanan yang mengatur proses pelimpahan tanggung jawab penanganan pasien dari satu fasilitas kesehatan ke fasilitas kesehatan lainnya yang memiliki kemampuan layanan lebih tinggi. Sistem ini dirancang untuk memastikan bahwa pasien memperoleh layanan medis yang sesuai dengan kebutuhan dan tingkat kompleksitas kondisinya. Rujukan dilakukan ketika fasilitas kesehatan tingkat pertama memiliki keterbatasan dalam menyediakan layanan lanjutan yang dibutuhkan pasien (Kementerian Kesehatan Republik Indonesia, 2014b).

Secara konseptual, sistem rujukan tidak hanya berkaitan dengan aspek medis, tetapi juga mencakup proses administratif dan pengelolaan informasi. Proses rujukan melibatkan pencatatan data pasien, alasan rujukan, tujuan rujukan, serta dokumentasi layanan yang telah diberikan. Dengan demikian, sistem rujukan memiliki keterkaitan erat dengan sistem informasi kesehatan karena membutuhkan pengelolaan data yang akurat, terstruktur, dan mudah diakses (Kementerian Kesehatan Republik Indonesia, 2014b).

Tujuan utama sistem rujukan adalah meningkatkan efektivitas dan efisiensi pelayanan kesehatan. Melalui sistem rujukan, pasien dapat diarahkan ke fasilitas yang memiliki sumber daya, tenaga medis, dan peralatan yang memadai. Sistem ini juga membantu mengoptimalkan peran fasilitas kesehatan tingkat pertama agar tetap fokus pada layanan dasar. Ketepatan rujukan menjadi faktor penting karena berpengaruh terhadap kualitas layanan serta keselamatan pasien (Kementerian Kesehatan Republik Indonesia, 2012)

Dalam praktiknya, proses rujukan sering menghadapi berbagai tantangan, seperti keterlambatan pengambilan keputusan, ketidaksesuaian tujuan rujukan, maupun keterbatasan informasi pendukung. Keputusan rujukan yang kurang tepat dapat berdampak pada penurunan kualitas layanan, peningkatan beban fasilitas kesehatan, serta ketidakpuasan pasien. Oleh karena itu, diperlukan pendekatan yang mampu mendukung proses rujukan secara lebih sistematis dan objektif.

Dalam konteks Sistem Informasi, sistem rujukan dapat dipandang sebagai proses yang melibatkan aliran data, informasi, dan keputusan. Data pasien yang tersedia, seperti diagnosa dan tingkat keparahan, memiliki potensi untuk dianalisis guna mendukung ketepatan rekomendasi rujukan. Pemanfaatan data historis memungkinkan identifikasi pola yang dapat digunakan sebagai dasar rekomendasi berbasis analitik (Jogiyanto, 2014).

Pendekatan berbasis data mining memberikan peluang untuk meningkatkan kualitas keputusan rujukan melalui pemodelan prediktif. Dengan memanfaatkan algoritma machine learning, data pasien dapat diolah untuk menghasilkan rekomendasi tujuan rujukan yang lebih objektif. Pendekatan ini membantu

mengurangi ketergantungan pada pertimbangan subjektif semata serta meningkatkan konsistensi keputusan.

Dalam penelitian ini, sistem rujukan menjadi konteks utama penerapan model klasifikasi yang dibangun. Data pasien berupa diagnosa dan tingkat keparahan digunakan sebagai atribut input dalam model Random Forest. Model yang dihasilkan diharapkan mampu memberikan rekomendasi rumah sakit rujukan secara sistematis berdasarkan pola data historis pasien.

Pemanfaatan pendekatan data mining dalam sistem rujukan sejalan dengan kebutuhan pengelolaan data modern, di mana data tidak hanya berfungsi sebagai arsip, tetapi juga sebagai sumber pengetahuan. Dengan model klasifikasi yang dibangun, sistem diharapkan mampu mendukung proses rujukan secara lebih efektif, konsisten, dan berbasis data.

Berdasarkan uraian tersebut, sistem rujukan kesehatan dapat dipahami sebagai mekanisme pelayanan yang memerlukan pengelolaan informasi yang baik. Dalam konteks penelitian ini, sistem rujukan menjadi domain aplikasi bagi model klasifikasi Random Forest yang memanfaatkan data diagnosa dan tingkat keparahan pasien.

2.8. Puskesmas Sigambal

Puskesmas (Pusat Kesehatan Masyarakat) merupakan fasilitas pelayanan kesehatan tingkat pertama yang memiliki peran strategis dalam penyelenggaraan upaya kesehatan masyarakat dan pelayanan kesehatan perorangan. Puskesmas menjadi ujung tombak sistem pelayanan kesehatan karena berfungsi sebagai titik awal masyarakat dalam memperoleh layanan kesehatan dasar. Selain memberikan pelayanan promotif, preventif, kuratif, dan rehabilitatif, puskesmas juga memiliki

tanggung jawab dalam pencatatan serta pengelolaan data kesehatan masyarakat (Kementerian Kesehatan Republik Indonesia, 2014).

Sebagai fasilitas kesehatan tingkat pertama, puskesmas tidak hanya menjalankan fungsi pelayanan medis, tetapi juga menjalankan fungsi administratif dan manajerial. Berbagai aktivitas pelayanan menghasilkan data yang beragam, seperti data kunjungan pasien, hasil pemeriksaan, diagnosa, tindakan medis, hingga proses rujukan. Data tersebut dicatat sebagai bagian dari dokumentasi layanan yang berkelanjutan (Kementerian Kesehatan Republik Indonesia, 2014a)

Puskesmas Sigambal merupakan salah satu puskesmas yang menjadi lokasi penelitian ini. Dalam pelaksanaan layanan, Puskesmas Sigambal memberikan pelayanan kesehatan dasar kepada masyarakat serta melakukan rujukan pasien ke fasilitas kesehatan lanjutan apabila kondisi pasien memerlukan penanganan yang lebih kompleks. Proses rujukan dilakukan berdasarkan pertimbangan kondisi pasien, termasuk hasil pemeriksaan, diagnosa, serta tingkat keparahan yang dialami pasien.

Dalam perspektif Sistem Informasi, data yang dihasilkan dari aktivitas pelayanan puskesmas merupakan aset penting yang memiliki potensi analitik. Data pasien yang tersimpan dalam sistem informasi kesehatan tidak hanya berfungsi sebagai arsip administratif, tetapi juga dapat dimanfaatkan untuk mendukung analisis, evaluasi, dan pengembangan model prediktif. Pemanfaatan data historis memungkinkan identifikasi pola layanan yang dapat memberikan nilai tambah bagi institusi (Jogiyanto, 2014).

Namun demikian, dalam praktik operasional, pemanfaatan data pasien sering kali masih terbatas pada kebutuhan pelaporan rutin. Potensi data sebagai

sumber pengetahuan belum sepenuhnya dioptimalkan. Padahal, pendekatan data mining memungkinkan proses penggalian informasi dari data historis sehingga dapat menghasilkan model yang mendukung kebutuhan analitik organisasi (Kusrini & Luthfi, 2009).

Dalam penelitian ini, data pasien yang diperoleh dari Puskesmas Sigambal, khususnya data diagnosa dan tingkat keparahan, dimanfaatkan sebagai dataset dalam pembangunan model klasifikasi menggunakan algoritma Random Forest.

Pemanfaatan data tersebut bertujuan untuk membangun model yang mampu menghasilkan rekomendasi rumah sakit rujukan secara sistematis dan berbasis data.

Pemilihan Puskesmas Sigambal sebagai lokasi penelitian didasarkan pada relevansi data yang tersedia dengan kebutuhan penelitian. Selain itu, keberadaan proses rujukan pasien di puskesmas ini memberikan konteks nyata bagi penerapan model klasifikasi. Dengan pendekatan analitik berbasis data mining, penelitian ini diharapkan mampu menunjukkan bagaimana data operasional puskesmas dapat dimanfaatkan secara lebih optimal.

Berdasarkan uraian tersebut, Puskesmas Sigambal tidak hanya dipandang sebagai lokasi penelitian, tetapi juga sebagai sumber data yang memiliki nilai strategis. Pemanfaatan data pasien dalam penelitian ini menunjukkan pentingnya peran Sistem Informasi dalam mengubah data mentah menjadi informasi yang bernilai guna.

2.9. Unified Modeling Language (UML)

Unified Modeling Language (UML) merupakan bahasa pemodelan standar yang digunakan dalam analisis dan perancangan sistem perangkat lunak. UML

berfungsi untuk memvisualisasikan, menspesifikasikan, membangun, serta mendokumentasikan komponen sistem sehingga rancangan sistem dapat dipahami secara terstruktur (Rosa & Shalahuddin, 2018).





Penggunaan UML dalam pengembangan sistem informasi memberikan kemudahan dalam menggambarkan kebutuhan sistem, alur proses, serta hubungan antar komponen sistem sebelum tahap implementasi dilakukan. UML juga berperan sebagai alat komunikasi antara analis sistem dan pengembang, sehingga dapat meminimalkan kesalahan dalam proses perancangan (Pressman, 2015).

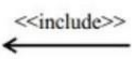
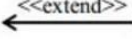
Beberapa diagram UML yang umum digunakan dalam perancangan sistem antara lain:

1. Use Case Diagram

Use Case Diagram digunakan untuk menggambarkan interaksi antara aktor dan sistem. Diagram ini menunjukkan fungsi utama sistem serta pihak yang terlibat dalam penggunaan sistem (Rosa & Shalahuddin, 2018).

Tabel 2.1 Simbol Use case Diagram


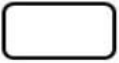



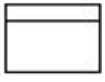
| Simbol | Nama | Keterangan |
|---|---------------------|---|
|  | <i>Actor</i> | Simbol ini digunakan untuk merepresentasikan peran pengguna, sistem lain, atau perangkat eksternal yang berinteraksi langsung dengan sistem melalui <i>use case</i> . |
|  | <i>Use case</i> | Simbol ini digunakan untuk menggambarkan fungsi atau layanan yang disediakan oleh sistem sebagai hasil interaksi antara sistem dengan aktor. |
|  | <i>Association</i> | Simbol ini digunakan untuk menunjukkan hubungan atau komunikasi antara aktor dengan <i>use case</i> dalam sistem. |
|  | <i>Generalisasi</i> | Simbol ini digunakan untuk menunjukkan menunjukkan spesialisasi aktor untuk dapat berpartisipasi dengan <i>use case</i> . |

| | | |
|---|-------------|---|
|  | <<include>> | Simbol ini digunakan untuk menunjukkan bahwa suatu <i>use case</i> selalu melibatkan <i>use case</i> lain sebagai bagian dari fungsionalitas utamanya. |
|  | <<extend>> | Simbol ini digunakan untuk menunjukkan bahwa suatu <i>use case</i> merupakan tambahan fungsional dari <i>use case</i> lainnya jika suatu kondisi terpenuhi. |

2. Activity Diagram

Activity Diagram digunakan untuk menggambarkan alur aktivitas atau proses yang terjadi dalam sistem. Diagram ini membantu dalam memahami urutan proses yang berjalan (Sukanto & Shalahuddin, 2018).

Tabel 2.2 Simbol Activity Diagram

| | | |
|---|-------------------------------|--|
|  | Status awal | Sebuah diagram aktivitas memiliki sebuah status awal. |
|  | Aktivitas | Aktivitas yang dilakukan sistem, aktivitas biasanya diawali dengan kata kerja. |
|  | Percabangan / <i>Decision</i> | Percabangan dimana ada pilihan aktivitas yang lebih dari satu. |
|  | Penggabungan / <i>Join</i> | Penggabungan dimana yang mana lebih dari satu aktivitas lalu digabungkan jadi satu. |
|  | Status Akhir | Status akhir yang dilakukan sistem, sebuah diagram aktivitas memiliki sebuah status akhir. |
|  | Swimlane | <i>Swimlane</i> memisahkan organisasi bisnis yang bertanggung jawab terhadap aktivitas yang terjadi. |

3. Class Diagram

Class Diagram digunakan untuk menggambarkan struktur sistem, termasuk kelas, atribut, metode, serta hubungan antar kelas (Rosa & Shalahuddin, 2018).

Tabel 2.3 Simbol *Class Diagram*

| Simbol | Nama | Keterangan |
|---------------|----------------------|---|
| | Class | Kelas pada struktur sistem. |
| | Interface | Sama dengan konsep interface dalam pemrograman berorientasi objek. |
| | Association | Relasi antarkelas dengan arti umum, asosiasi biasanya juga disertai dengan <i>Multiplicity</i> . |
| | Directed Association | Relasi antarkelas dengan makna yang jelas digunakan oleh kelas yang lain, asosiasi biasanya juga disertai dengan <i>Multiplicity</i> . |
| | Generalisasi | Relasi antarkelas dengan makna generalisasi-spesialisasi (umum-khusus). |
| | Dependency | Relasi antarkelas dengan makna ketergantungan antarkelas. |
| | Aggregation | Relasi antarkelas dengan makna <i>whole-part</i> yang bersifat lemah, di mana bagian masih dapat berdiri sendiri tanpa bergantung pada keseluruhan. |

Dalam penelitian ini, UML digunakan sebagai alat bantu dalam merancang sistem rekomendasi rumah sakit rujukan. UML membantu menggambarkan interaksi pengguna dengan sistem, alur proses klasifikasi menggunakan algoritma Random Forest, serta struktur data yang digunakan, khususnya data diagnosa dan tingkat keparahan pasien. Pemodelan ini bertujuan agar rancangan sistem terdokumentasi secara sistematis dan selaras dengan kebutuhan penelitian.

2.10. Penelitian Terdahulu

Penelitian terdahulu diperlukan untuk mengetahui perkembangan studi yang relevan dengan topik penelitian serta mengidentifikasi posisi penelitian yang dilakukan. Dalam sepuluh tahun terakhir, berbagai penelitian di Indonesia telah menerapkan algoritma Random Forest dan metode machine learning dalam bidang

kesehatan, khususnya untuk klasifikasi diagnosa penyakit dan prediksi risiko kondisi pasien.

Sebagian besar penelitian tersebut berfokus pada peningkatan akurasi model klasifikasi atau perbandingan performa antar algoritma. Namun, penelitian yang secara khusus mengembangkan model klasifikasi sebagai dasar rekomendasi rumah sakit rujukan berdasarkan data diagnosa dan tingkat keparahan pasien masih terbatas. Oleh karena itu, penelitian ini dilakukan untuk mengisi celah tersebut dengan membangun model rekomendasi berbasis Random Forest di lingkungan puskesmas.

Tabel 2.4 Penelitian Terdahulu

| NO | Judul Penelitian dan Penulis (Tahun) | Metode yang Digunakan | Fokus Penelitian | Keterbatasan Penelitian |
|-----------|---|------------------------------|---|--|
| 1 | <i>Penerapan Random Forest untuk Klasifikasi Diagnosa Penyakit - (Arisa, 2025)</i> | Random Forest | Klasifikasi diagnosa penyakit berdasarkan data pasien | Tidak membahas rekomendasi rujukan |
| 2 | <i>Prediksi Jumlah Pasien Masuk Rumah Sakit Menggunakan Random Forest - (Asaury et al., 2025)</i> | Random Forest | Prediksi jumlah pasien masuk RS (manajemen sumber daya) | Fokus pada jumlah pasien, bukan rekomendasi rujukan per pasien |
| 3 | | Random Forest | | Spesifik untuk diabetes tidak |

| | | | | |
|---|---|--------------------------------|---|--|
| | <i>Penerapan Algoritma Random Forest untuk Prediksi Risiko Diabetes - (Siswoyo & Nurhafidz, 2025)</i> | | Prediksi risiko diabetes berbasis data kesehatan klinis | menggabungkan variabel keparahan/rujukan |
| 4 | <i>Implementasi Algoritma Random Forest dalam Klasifikasi Diagnosis Penyakit Stroke – (I. et al., 2025)</i> | Random Forest | Klasifikasi diagnosis stroke menggunakan data rekam medis | Fokus penyakit spesifik (stroke); tidak membahas aplikasi rujukan di puskesmas |
| 5 | <i>Implementasi Random Forest untuk Klasifikasi Diagnosis Stroke (Siregar, 2023)</i> | Random Forest | Klasifikasi diagnosis stroke (perbandingan implementasi) | Dataset/studi kasus terbatas; tidak untuk rekomendasi rujukan |
| 6 | <i>Implementasi Random Forest pada Klasifikasi Penyakit (studi kasus jantung) - (S. et al., 2023)</i> | Random Forest vs Decision Tree | Klasifikasi penyakit jantung; perbandingan performa | Fokus pada perbandingan algoritma, bukan sistem rujukan |
| 7 | <i>Klasifikasi Penyakit Jantung Menggunakan Random Forest - (Hidayat, 2023)</i> | Random Forest | Klasifikasi penyakit jantung berdasarkan fitur pasien | Objek penelitian spesifik; keterbatasan generalisasi ke puskesmas |
| 8 | <i>Klasifikasi Diabetes /</i> | Random Forest / ML | Klasifikasi diabetes / DM | Fokus pada akurasi model; tidak |

| | | | | |
|----|---|-----------------------------|---|--|
| | <i>Penyakit Metabolik dengan Machine Learning</i> - (P. et al., 2024) | | pada dataset rumah sakit/puskesmas | menghubungkan ke keputusan rujukan |
| 9 | <i>Random Forest dan Aplikasi Pengukuran Kesehatan (studi sensor / tekanan darah)</i> - (Sunarya, 2024) | Random Forest | Aplikasi RF untuk analisis sinyal/sensor kesehatan (mis. tekanan darah) | Fokus teknis sensor; bukan alur layanan atau rujukan pasien |
| 10 | <i>Analisis Perbandingan Naïve Bayes & Random Forest dalam Klasifikasi Penyakit di Puskesmas</i> — (Virgiawan & Erizal, 2025) | Random Forest & Naïve Bayes | Perbandingan algoritma untuk klasifikasi penyakit di puskesmas | Tidak memproduksi modul rekomendasi rujukan, hanya perbandingan performa |

2.11. Kerangka Berpikir

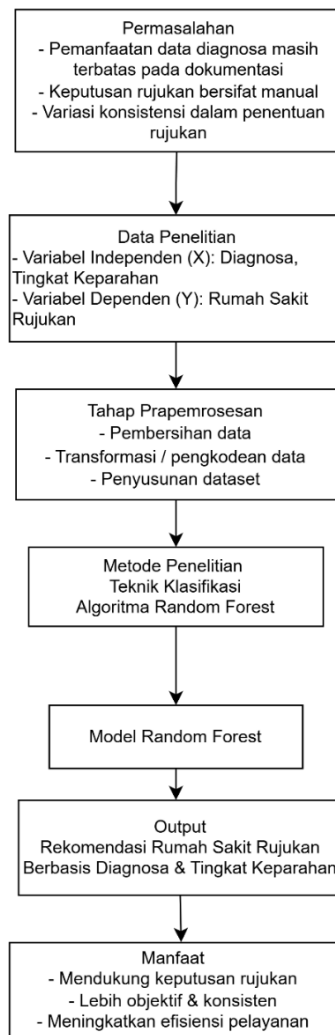
Kerangka berpikir merupakan gambaran konseptual yang menjelaskan alur logis penelitian berdasarkan permasalahan yang telah diidentifikasi. Kerangka ini disusun untuk memberikan arah yang sistematis dalam proses penelitian, mulai dari identifikasi masalah hingga menghasilkan solusi berupa model yang dapat diimplementasikan.

Permasalahan yang terjadi di Puskesmas Sigambal adalah pemanfaatan data diagnosa yang masih terbatas pada dokumentasi administratif, sementara proses penentuan rumah sakit rujukan masih dilakukan secara manual. Kondisi tersebut berpotensi menimbulkan variasi dalam pengambilan keputusan rujukan, sehingga diperlukan suatu pendekatan berbasis data untuk membantu menghasilkan rekomendasi yang lebih konsisten dan terukur.

Penelitian ini memanfaatkan data diagnosa dan tingkat keparahan pasien sebagai variabel independen, sedangkan rumah sakit rujukan sebagai variabel dependen. Data yang diperoleh selanjutnya melalui tahap prapemrosesan yang meliputi pembersihan data, transformasi atau pengkodean data, serta penyusunan dataset agar siap digunakan dalam proses pemodelan.

Metode yang digunakan dalam penelitian ini adalah teknik klasifikasi dengan algoritma Random Forest. Algoritma ini dipilih karena mampu menangani data dengan baik serta memiliki tingkat akurasi yang relatif tinggi dalam proses klasifikasi. Model yang dihasilkan kemudian digunakan untuk memberikan rekomendasi rumah sakit rujukan berdasarkan kombinasi diagnosa dan tingkat keparahan pasien.

Secara keseluruhan, kerangka berpikir penelitian ini menggambarkan alur mulai dari identifikasi permasalahan, pengolahan data, proses klasifikasi, hingga menghasilkan output berupa rekomendasi rumah sakit rujukan berbasis data.



Gambar 2.1 Kerangka Berpikir

BAB III

ANALISA DAN PERANCANGAN SISTEM

3.1. Analisis Permasalahan

Analisis permasalahan merupakan tahap awal dalam penelitian yang bertujuan untuk mengidentifikasi kondisi nyata di lapangan serta menentukan solusi yang relevan terhadap permasalahan tersebut. Pada tahap ini dilakukan pengkajian terhadap proses penentuan rumah sakit rujukan yang berjalan di Puskesmas Sigambal.

Berdasarkan hasil pengamatan dan studi awal, proses penentuan rumah sakit rujukan masih dilakukan secara manual oleh tenaga kesehatan. Keputusan rujukan umumnya didasarkan pada pertimbangan pengalaman, kebiasaan, serta penilaian subjektif terhadap kondisi pasien. Meskipun pendekatan tersebut tetap mempertimbangkan aspek medis, namun belum didukung oleh sistem berbasis data yang mampu membantu memberikan rekomendasi secara konsisten.

Di sisi lain, data diagnosa pasien sebenarnya telah terdokumentasi dengan baik. Namun, data tersebut lebih banyak digunakan sebagai arsip administratif dan belum dimanfaatkan sebagai sumber informasi untuk mendukung proses pengambilan keputusan. Padahal, menurut konsep data mining, data historis yang tersimpan dapat dieksplorasi untuk menemukan pola dan hubungan tersembunyi yang berguna dalam proses prediksi atau klasifikasi.

Permasalahan lain yang muncul adalah potensi ketidakkonsistenan dalam pemberian rujukan. Pasien dengan diagnosa yang serupa dan tingkat keparahan yang relatif sama dapat saja dirujuk ke rumah sakit yang berbeda, tergantung pada pertimbangan individu tenaga kesehatan. Kondisi ini menunjukkan bahwa proses pengambilan keputusan masih belum terstandarisasi secara sistematis.

Untuk mengatasi permasalahan tersebut, diperlukan suatu pendekatan berbasis data yang mampu mengolah informasi diagnosa dan tingkat keparahan pasien menjadi sebuah rekomendasi yang lebih objektif. Salah satu teknik yang dapat digunakan adalah klasifikasi dalam data mining. Klasifikasi merupakan metode pembelajaran terawasi (*supervised learning*) yang bertujuan untuk memetakan data ke dalam kelas tertentu berdasarkan pola yang dipelajari dari data sebelumnya.

Dalam penelitian ini, algoritma yang digunakan adalah Random Forest. Algoritma ini merupakan pengembangan dari metode decision tree yang bekerja dengan membangun banyak pohon keputusan dan menggabungkan hasilnya untuk meningkatkan akurasi prediksi. Random Forest dikenal memiliki performa yang baik dalam menangani data dengan variasi tinggi serta mampu mengurangi risiko overfitting dibandingkan satu pohon keputusan tunggal.

Dengan demikian, penelitian ini berupaya memanfaatkan data diagnosa dan tingkat keparahan pasien sebagai variabel input untuk membangun model klasifikasi yang dapat menghasilkan rekomendasi rumah sakit rujukan secara lebih sistematis, konsisten, dan berbasis data.

3.2. Algoritma Random Forest

Random Forest merupakan salah satu algoritma klasifikasi dalam machine learning yang termasuk dalam kategori ensemble learning. Algoritma ini bekerja dengan membangun sejumlah pohon keputusan (*decision tree*) dari sampel data yang dipilih secara acak menggunakan teknik bootstrap sampling. Setiap pohon keputusan akan menghasilkan prediksi kelas, kemudian hasil prediksi tersebut digabungkan menggunakan metode majority voting untuk memperoleh keputusan akhir yang lebih stabil dan akurat.

Dalam penelitian ini, algoritma Random Forest digunakan untuk menghasilkan rekomendasi rumah sakit rujukan berdasarkan atribut pasien yang terdiri dari umur, jenis kelamin, kategori diagnosa, dan tingkat keparahan, dengan rumah sakit rujukan sebagai variabel target.

3.2.1 Data Mentah

Data mentah yang digunakan dalam penelitian ini diperoleh dari laporan rujukan pasien di Puskesmas Sigambal. Data tersebut diambil dari sistem pencatatan rujukan pasien yang berisi berbagai informasi administratif dan medis pasien.

Data mentah terdiri dari berbagai atribut seperti nama pasien, nomor rekam medis, nomor identitas, alamat, diagnosa penyakit, dokter perujuk, rumah sakit tujuan rujukan, serta nomor rujukan BPJS. Data ini masih berupa data administratif yang belum siap digunakan dalam proses analisis data mining.

Berikut merupakan contoh sebagian data mentah yang diperoleh dari laporan rujukan pasien

Tabel 3.1 Data Mentah Rujukan Pasien

| No | Tanggal | Nama Pasien | Umur | Jenis Kelamin | Diagnosa | RS Tujuan Rujukan |
|-----------|----------------|---------------------|-------------|----------------------|---|------------------------------|
| 1 | 06-12-2025 | Agustina | 49 | Perempuan | Leiomyoma of uterus | RSU Karya Bakti |
| 2 | 06-12-2025 | Warni | 47 | Perempuan | Hypertension secondary to endocrine disorders | Klinik Utama Spesialis Bunda |
| 3 | 06-12-2025 | Sutan Muda Dongoran | 23 | Laki-laki | Benign neoplasm intrathoracic organs | RSUD Rantauprapat |
| 4 | 06-12-2025 | Gazali Sinambela | 59 | Laki-laki | Congestive heart failure | RSUD Rantauprapat |
| 5 | 06-12-2025 | Halomoan Rambe | 44 | Laki-laki | Low back pain | RSUD Rantauprapat |

Data mentah tersebut masih memiliki banyak atribut yang tidak digunakan dalam penelitian seperti NIK, nomor rekam medis, alamat, nomor asuransi, dokter perujuk, poli rujukan, serta nomor rujukan BPJS. Oleh karena itu diperlukan proses

data cleaning untuk menyeleksi atribut yang relevan dengan proses klasifikasi menggunakan algoritma Random Forest.

3.2.2 Data Cleaning

Tahap data cleaning dilakukan untuk mempersiapkan dataset agar dapat digunakan dalam proses klasifikasi. Pada tahap ini dilakukan seleksi atribut dan penyederhanaan diagnosa penyakit.

Beberapa atribut pada data mentah tidak digunakan karena tidak memiliki pengaruh terhadap proses penentuan rumah sakit rujukan. Atribut yang dihapus pada proses data cleaning antara lain:

- nomor rekam medis
- nomor identitas pasien
- alamat pasien
- nomor asuransi
- dokter perujuk
- poli rujukan
- nomor rujukan BPJS

Atribut yang dipertahankan dalam dataset penelitian adalah atribut yang memiliki keterkaitan dengan proses penentuan rujukan pasien, yaitu:

- umur pasien
- jenis kelamin pasien
- diagnosa penyakit
- kategori diagnosa
- tingkat keparahan
- rumah sakit rujukan

Selain itu dilakukan proses pengelompokan diagnosa penyakit. Pada data mentah terdapat berbagai jenis diagnosa yang sangat spesifik. Apabila seluruh diagnosa tersebut digunakan secara langsung, maka jumlah kategori diagnosa menjadi sangat banyak sehingga menyulitkan proses klasifikasi.

Untuk mengatasi hal tersebut, diagnosa penyakit kemudian dikelompokkan ke dalam beberapa kategori diagnosa yang memiliki karakteristik medis yang serupa.

Tabel 3.2 Pengelompokan Diagnosa

| Diagnosa Asli | Kategori Diagnosa |
|---|--------------------------|
| Leiomyoma of uterus | Gangguan Reproduksi |
| Hypertension secondary to endocrine disorders | Penyakit Kardiovaskular |
| Benign neoplasm intrathoracic organs | Tumor / Neoplasma |
| Congestive heart failure | Penyakit Jantung |
| Low back pain | Gangguan Tulang dan Otot |

Proses pengelompokan diagnosa ini bertujuan untuk mengurangi kompleksitas data serta membantu algoritma Random Forest dalam menemukan pola hubungan antara kondisi pasien dan rumah sakit rujukan yang sesuai.

Selain itu atribut jenis kelamin dikonversi menjadi bentuk numerik untuk mempermudah proses komputasi dalam algoritma Random Forest, dengan ketentuan:

- 1 = laki-laki
- 0 = perempuan

3.2.3. Dataset Setelah Cleaning

Setelah melalui proses data cleaning, dataset penelitian terdiri dari beberapa atribut utama yaitu umur, jenis kelamin, diagnosa asli, kategori diagnosa, tingkat keparahan, dan rumah sakit rujukan.

Dataset yang telah dibersihkan kemudian digunakan sebagai input dalam proses klasifikasi menggunakan algoritma Random Forest.

Tabel 3.3 Dataset Penelitian

| No | Umur | Jenis Kelamin | Diagnosa Asli | Kategori Diagnosa | Rumah Sakit | Keparahan |
|----|------|---------------|--|--------------------------|------------------------|-----------|
| 1 | 49 | 0 | Leiomyoma of uterus | Gangguan Reproduksi | RSU Karya Bakti | Berat |
| 2 | 47 | 0 | Hypertension secondary endocrine disorders | Penyakit Kardiovaskular | Klinik Spesialis Bunda | Sedang |
| 3 | 23 | 1 | Benign neoplasm intrathoracic organs | Tumor / Neoplasma | RSUD Rantauprapt | Berat |
| 4 | 59 | 1 | Congestive heart failure | Penyakit Jantung | RSUD Rantauprapt | Berat |
| 5 | 44 | 1 | Low back pain | Gangguan Tulang dan Otot | RSUD Rantauprapt | Ringan |

Dataset yang telah melalui proses pembersihan data inilah yang kemudian digunakan sebagai data input dalam proses perhitungan algoritma Random Forest untuk menghasilkan rekomendasi rumah sakit rujukan.

3.2.4. Penentuan Tingkat Keparahan Penyakit

Pada data mentah yang diperoleh dari laporan rujukan pasien, tidak terdapat atribut yang secara langsung menunjukkan tingkat keparahan penyakit pasien. Oleh karena itu, pada penelitian ini dilakukan proses penentuan tingkat keparahan berdasarkan jenis diagnosa yang tercatat pada data rujukan pasien.

Penentuan tingkat keparahan dilakukan dengan mengelompokkan diagnosa penyakit ke dalam tiga kategori tingkat keparahan yaitu ringan, sedang, dan berat. Pengelompokan ini dilakukan dengan mempertimbangkan karakteristik penyakit serta kebutuhan penanganan medis lanjutan yang biasanya memerlukan rujukan ke fasilitas kesehatan tingkat lanjut.

Secara umum klasifikasi tingkat keparahan dalam penelitian ini dijelaskan sebagai berikut:

- Ringan

Merupakan kondisi penyakit yang relatif tidak mengancam jiwa namun memerlukan pemeriksaan atau tindakan lanjutan oleh dokter spesialis.

- Sedang

Merupakan kondisi penyakit yang memerlukan pemeriksaan lanjutan atau pengobatan yang lebih intensif dibandingkan kondisi ringan.

- Berat

Merupakan kondisi penyakit yang memiliki tingkat risiko tinggi terhadap kesehatan pasien dan memerlukan penanganan medis lebih lanjut di rumah sakit rujukan.

3.2.5. Tahapan Algoritma Random Forest

Setelah dataset penelitian diperoleh melalui proses data cleaning dan penentuan tingkat keparahan, tahap selanjutnya adalah proses klasifikasi

menggunakan algoritma Random Forest. Algoritma ini bekerja dengan membangun beberapa pohon keputusan (*decision tree*) dari data training yang dipilih secara acak.

Secara umum tahapan algoritma Random Forest dalam penelitian ini adalah sebagai berikut:

1. Bootstrap Sampling

Mengambil sejumlah data secara acak dari dataset penelitian untuk membangun setiap pohon keputusan.

2. Pembentukan Decision Tree

Setiap sampel data digunakan untuk membangun satu pohon keputusan.

3. Pemilihan Atribut pada Node

Pada setiap node dilakukan pemilihan atribut terbaik berdasarkan nilai Gini Index.

4. Proses Klasifikasi

Setiap pohon keputusan menghasilkan prediksi rumah sakit rujukan.

5. Majority Voting

Hasil prediksi dari seluruh pohon keputusan digabungkan menggunakan metode majority voting untuk menentukan hasil akhir.

3.2.6. Perhitungan Manual

Untuk memberikan gambaran mengenai cara kerja algoritma Random Forest, dilakukan simulasi perhitungan manual menggunakan sebagian data sampel dari dataset penelitian. Data sampel ini digunakan untuk menunjukkan proses

pembentukan pohon keputusan dalam menentukan rekomendasi rumah sakit rujukan.

Tabel 3.4 Data Sampel Perhitungan Manual

| No | Umur | Jenis Kelamin | Diagnosa Kategori | Tingkat Keparahan | Rumah Sakit Rujukan |
|-----------|-------------|----------------------|----------------------------------|--------------------------|------------------------------|
| 1 | 3 | 1 | Penyakit Paru | Ringan | RSU ELPI Al-Azis |
| 2 | 41 | 1 | Tuberkulosis | Berat | RSUD Rantauprapat |
| 3 | 55 | 0 | Masalah Kesehatan Gigi dan Mulut | Ringan | RSU Karya Bakti Ujung Bandar |
| 4 | 55 | 0 | Komplikasi Pasca Operasi | Berat | RSU Karya Bakti Ujung Bandar |
| 5 | 53 | 0 | Gangguan Penglihatan | Ringan | RSUD Rantauprapat |
| 6 | 30 | 0 | Perawatan Lanjutan | Ringan | RSU Karya Bakti Ujung Bandar |
| 7 | 44 | 1 | Gangguan Tulang dan Sendi | Sedang | RSUD Rantauprapat |
| 8 | 22 | 0 | Kehamilan | Sedang | RSU ELPI Al-Azis |

Jumlah data = 8

Distribusi kelas:

- RSUD Rantauprapat = 3
- RSUD Karya Bakti Ujung Bandar = 3
- RSUD ELPI Al-Azis = 2

3.2.7. Perhitungan Gini Index Awal

Gini Index digunakan untuk mengukur tingkat ketidakteraturan data pada suatu node. Nilai Gini yang lebih kecil menunjukkan bahwa data pada node tersebut lebih homogen.

Rumus Gini Index:

$$Gini = 1 - \sum (p_i)^2$$

Probabilitas masing-masing kelas:

$$P(RSUD) = 3/8 = 0,375$$

$$P(RSU Karya Bakti) = 3/8 = 0,375$$

$$P(RSU ELPI) = 2/8 = 0,25$$

Perhitungan

$$Gini = 1 - (0,375^2 + 0,375^2 + 0,25^2)$$

$$Gini = 1 - (0,1406 + 0,1406 + 0,0625)$$

$$Gini = 1 - 0,3437$$

$$Gini = 0,6563$$

Nilai **0,6563** menunjukkan bahwa data masih bersifat heterogen sehingga perlu dilakukan pemisahan atribut.

1. Pemisahan Berdasarkan Atribut Tingkat Keparahan

A. Keparahan = ringan (4 data)

$$RSU ELPI = 1$$

$$RSUD = 1$$

RSU Karya Bakti = 2

$$P1 = 1/4 = 0,25$$

$$P2 = 1/4 = 0,25$$

$$P3 = 2/4 = 0,5$$

$$Gini_{ringan} = 1 - (0,25^2 + 0,25^2 + 0,5^2)$$

$$Gini_{ringan} = 1 - (0,0625 + 0,0625 + 0,25)$$

$$Gini_{ringan} = 1 - 0,375$$

$$Gini_{ringan} = 0,625$$

B. Keparahan = berat (2 data)

RSUD = 1

RSU Karya Bakti = 1

$$P = 1/2 = 0,5$$

$$Gini_{berat} = 1 - (0,5^2 + 0,5^2)$$

$$Gini_{berat} = 1 - (0,25 + 0,25)$$

$$Gini_{berat} = 1 - 0,5$$

$$Gini_{berat} = 0,5$$

C. Keparahan = sedang (2 data)

RSUD = 1

RSU ELPI = 1

$$Gini_{sedang} = 1 - (0,5^2 + 0,5^2)$$

$$Gini_{sedang} = 0,5$$

2. Menghitung Gini setelah Split

Rumus

$$Gini_{split} = \sum \frac{n_i}{n} \times Gini_i \quad \dots \dots \dots (3.1)$$

$$Gini_{split} = \sum \frac{n_i}{n} \times Gini_i$$

$$Gini_{split} = \frac{4}{8}(0,625) + \frac{2}{8}(0,5) + \frac{2}{8}(0,5)$$

$$Gini_{split} = 0,3125 + 0,125 + 0,125$$

$$Gini_{split} = 0,5625$$

Karena nilai Gini setelah split lebih kecil dari Gini awal, maka atribut tingkat keparahan dipilih sebagai root node.

3.2.8. Tahap Membangun Pohon Keputusan

Setelah atribut root node ditentukan, langkah selanjutnya adalah membangun struktur pohon keputusan berdasarkan atribut tersebut.

Struktur node pada pohon keputusan terdiri dari:

Root Node

Tingkat Keparahahan

Child Node

Ringan, Sedang, Berat

Leaf Node

- RSUD Rantauprapat
- RSUD Karya Bakti Ujung Bandar
- RSUD ELPI Al-Azis



Gambar 3.1 Pohon keputusan Algoritma Random Forest

3.2.9. Majority Voting

Dalam algoritma Random Forest, setiap pohon keputusan menghasilkan prediksi terhadap data pasien baru. Hasil akhir ditentukan menggunakan metode majority voting, yaitu kelas yang memiliki jumlah suara terbanyak dari seluruh pohon keputusan.

Misalkan pasien baru:

Diagnosa: Penyakit Paru

Tingkat Keparahan: Ringan

Hasil prediksi dari 3 pohon:

Tree 1 → RSU Karya Bakti

Tree 2 → RSU ELPI Al-Azis

Tree 3 → RSU Karya Bakti

Voting:

RSU Karya Bakti = 2 suara

RSU ELPI Al-Azis = 1 suara

Maka hasil akhir rekomendasi:

RSU Karya Bakti Ujung Bandar

3.2.10. Analisis Hasil Perhitungan

Berdasarkan perhitungan manual yang telah dilakukan, diperoleh nilai Gini awal sebesar 0,6563 yang menunjukkan bahwa data masih bersifat heterogen. Setelah dilakukan pemisahan berdasarkan atribut tingkat keparahan, nilai Gini menurun menjadi 0,5625. Penurunan nilai tersebut menunjukkan bahwa atribut tingkat keparahan mampu mengurangi tingkat impurity dan layak digunakan sebagai node akar dalam pembentukan pohon keputusan.

Simulasi majority voting juga menunjukkan bahwa Random Forest menentukan hasil akhir berdasarkan jumlah suara terbanyak dari beberapa pohon keputusan. Mekanisme ini menjadikan proses klasifikasi lebih stabil dibandingkan penggunaan satu pohon keputusan tunggal.

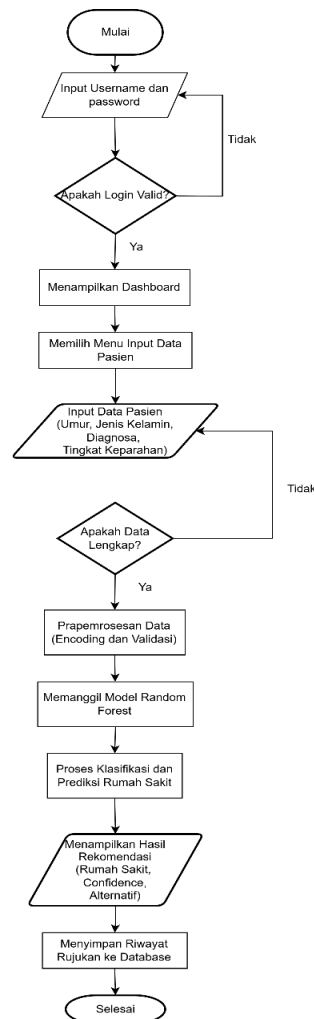
3.3. Flowchart Sistem

Flowchart sistem digunakan untuk menggambarkan alur proses kerja sistem rekomendasi rumah sakit rujukan yang telah dibangun. Flowchart ini menunjukkan tahapan mulai dari proses autentikasi pengguna hingga sistem menghasilkan rekomendasi rumah sakit berdasarkan model Random Forest.

Secara umum, sistem bekerja melalui beberapa tahapan utama, yaitu proses login, pengelolaan dashboard, input data pasien, proses klasifikasi menggunakan model Random Forest, serta penampilan hasil rekomendasi.

Flowchart Sistem Rekomendasi Rumah Sakit Rujukan menggambarkan alur kerja sistem mulai dari proses autentikasi pengguna hingga sistem menghasilkan rekomendasi rumah sakit berdasarkan model Random Forest. Proses dimulai dari login, input data pasien, validasi data, pemrosesan menggunakan

model, hingga menampilkan hasil rekomendasi dan menyimpan riwayat rujukan ke dalam database.



Gambar 3.2 Flowchart Sistem Rekomendasi Rumah Sakit Rujukan

3.4. Perancangan Sistem

Perancangan sistem merupakan tahap yang dilakukan setelah proses analisis kebutuhan sistem selesai dilakukan. Tahap ini bertujuan untuk merancang struktur dan komponen sistem yang akan dibangun sehingga sistem dapat bekerja sesuai dengan kebutuhan pengguna. Dalam penelitian ini, perancangan sistem dilakukan untuk membangun sistem rekomendasi rumah sakit rujukan yang memanfaatkan algoritma Random Forest dalam proses pengambilan keputusan.

Perancangan sistem dilakukan dengan menggambarkan model sistem secara terstruktur agar mempermudah proses pengembangan dan implementasi sistem. Melalui tahap perancangan ini, seluruh komponen sistem seperti alur proses, struktur data, serta antarmuka sistem dapat dirancang secara sistematis sehingga sistem yang dibangun dapat berjalan dengan baik.

Pada penelitian ini, proses perancangan sistem dilakukan menggunakan beberapa model pemodelan sistem yang meliputi diagram aktivitas (activity diagram), class diagram, serta perancangan antarmuka sistem. Diagram aktivitas digunakan untuk menggambarkan alur proses kerja sistem dari awal hingga menghasilkan rekomendasi rumah sakit rujukan. Class diagram digunakan untuk menggambarkan struktur kelas yang terdapat pada sistem serta hubungan antar kelas yang digunakan dalam pengolahan data. Sedangkan perancangan antarmuka sistem digunakan untuk memberikan gambaran tampilan sistem yang akan digunakan oleh pengguna.

Melalui tahap perancangan sistem ini diharapkan sistem rekomendasi rumah sakit rujukan yang dibangun dapat membantu tenaga medis dalam menentukan rumah sakit rujukan yang sesuai dengan kondisi pasien secara lebih cepat dan akurat.

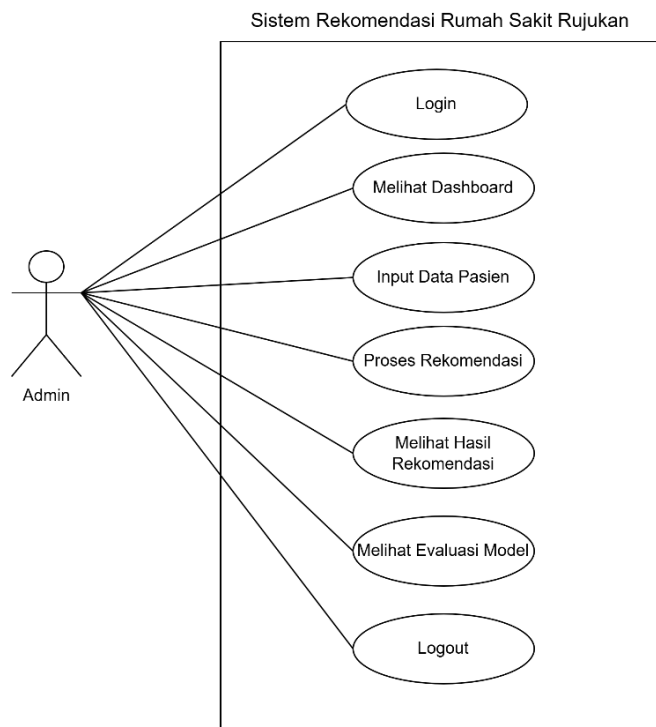
3.4.1. Use Case Diagram

Use Case Diagram merupakan salah satu bagian dari Unified Modeling Language (UML) yang digunakan untuk menggambarkan interaksi antara aktor dengan sistem. Use Case Diagram berfungsi untuk menunjukkan fungsi-fungsi utama yang tersedia di dalam sistem serta siapa saja yang dapat mengakses fungsi tersebut (Rosa & Shalahuddin, 2018).

Pada Sistem Rekomendasi Rumah Sakit Rujukan yang dibangun, terdapat satu aktor utama yaitu Admin atau Petugas Puskesmas. Admin merupakan pengguna yang memiliki hak akses untuk mengoperasikan seluruh fitur sistem, mulai dari proses autentikasi hingga memperoleh hasil rekomendasi rumah sakit.

Secara umum, Admin dapat melakukan beberapa aktivitas utama di dalam sistem, yaitu login, melihat dashboard, menginput data pasien, memproses rekomendasi, melihat hasil rekomendasi, melihat evaluasi model, serta logout. Seluruh aktivitas tersebut berada dalam batas sistem (system boundary) yang disebut sebagai *Sistem Rekomendasi Rumah Sakit Rujukan*.

Use Case Diagram ini menunjukkan bahwa seluruh proses sistem dikendalikan oleh Admin sebagai pengguna utama, dan sistem bekerja sebagai alat bantu dalam memberikan rekomendasi rumah sakit rujukan berdasarkan data pasien yang dimasukkan.



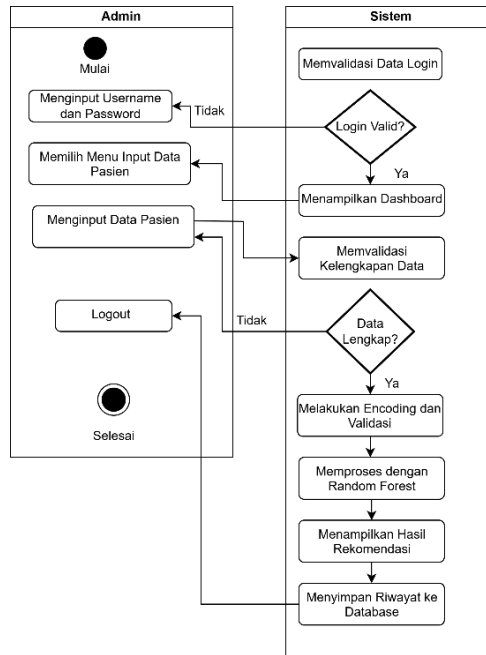
Gambar 3.3 Use Case Diagram

3.4.2. Activity Diagram

Activity Diagram pada Sistem Rekomendasi Rumah Sakit Rujukan menggambarkan alur aktivitas antara Admin dan Sistem mulai dari proses login hingga logout. Proses diawali ketika Admin memasukkan username dan password, kemudian sistem melakukan validasi data login. Jika login tidak valid, maka Admin diminta untuk menginput ulang data login. Jika login valid, sistem menampilkan dashboard sebagai halaman utama untuk mengakses fitur sistem.

Selanjutnya Admin memilih menu input data pasien dan mengisikan data yang diperlukan sesuai dengan atribut yang digunakan dalam proses klasifikasi. Sistem melakukan validasi kelengkapan data untuk memastikan seluruh informasi telah terisi dengan benar. Apabila data belum lengkap, maka Admin diminta melengkapi kembali data pasien. Jika data telah lengkap, sistem melakukan proses encoding dan validasi data sebelum diproses menggunakan algoritma Random Forest untuk menghasilkan rekomendasi rumah sakit rujukan berdasarkan kondisi pasien.

Hasil rekomendasi kemudian ditampilkan kepada Admin dan secara otomatis disimpan ke dalam database sebagai riwayat rujukan. Penyimpanan ini bertujuan untuk dokumentasi serta memudahkan proses monitoring dan evaluasi sistem di kemudian hari. Setelah seluruh proses selesai, Admin melakukan logout dan aktivitas sistem berakhir.



Gambar 3.4 Activity Diagram

3.4.3. Class Diagram

Class diagram digunakan untuk menggambarkan struktur kelas yang terdapat pada sistem serta hubungan antar kelas dalam sistem yang dibangun. Diagram ini menunjukkan atribut yang dimiliki oleh setiap kelas serta bagaimana data disimpan dan dikelola oleh sistem.

Pada sistem rekomendasi rumah sakit rujukan yang dibangun, terdapat dua kelas utama yaitu class users dan class riwayat. Kedua kelas tersebut digunakan untuk mengelola data pengguna sistem serta menyimpan data hasil rekomendasi rumah sakit rujukan.

Class users digunakan untuk menyimpan data pengguna yang memiliki akses ke dalam sistem. Class ini memiliki beberapa atribut yaitu id, username, dan password. Atribut id berfungsi sebagai primary key yang digunakan untuk mengidentifikasi setiap pengguna secara unik. Atribut username digunakan sebagai identitas pengguna saat melakukan login ke dalam sistem, sedangkan atribut

password digunakan sebagai kata sandi yang digunakan dalam proses autentikasi pengguna.

Class riwayat digunakan untuk menyimpan data riwayat rekomendasi rumah sakit rujukan yang dihasilkan oleh sistem. Class ini memiliki beberapa atribut yaitu id, umur, jk (jenis kelamin), diagnosa, kategori, keparahan, rs (rumah sakit rujukan), confidence, dan tanggal. Atribut id berfungsi sebagai primary key untuk membedakan setiap data riwayat yang tersimpan dalam database. Atribut umur, jk, diagnosa, kategori, dan keparahan merupakan data pasien yang digunakan sebagai input dalam proses klasifikasi menggunakan algoritma Random Forest. Atribut rs menyimpan hasil rekomendasi rumah sakit rujukan yang dihasilkan oleh sistem, sedangkan atribut confidence menyimpan nilai tingkat kepercayaan dari hasil prediksi model. Atribut tanggal digunakan untuk mencatat waktu ketika data rekomendasi tersebut disimpan dalam sistem.

Melalui class diagram ini dapat dilihat bahwa sistem menyimpan data pengguna serta riwayat hasil rekomendasi secara terstruktur di dalam database. Data tersebut kemudian dapat digunakan untuk kebutuhan monitoring, evaluasi sistem, serta pelaporan data rujukan pasien.



Gambar 3.5 Class Diagram

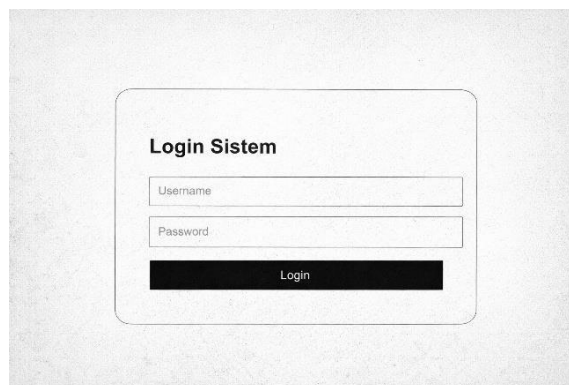
3.5. Perancangan Antarmuka

Perancangan antarmuka dilakukan untuk menggambarkan tampilan sistem sebelum tahap implementasi. Antarmuka dirancang berbasis web dengan tampilan sederhana, terstruktur, serta mudah digunakan oleh petugas puskesmas. Sistem dikembangkan menggunakan PHP dan MySQL dengan desain modern berbasis card layout dan visualisasi grafik untuk mempermudah pembacaan informasi.

Adapun perancangan antarmuka dalam sistem ini terdiri dari beberapa halaman utama sebagai berikut.

3.5.1. Halaman Login

Halaman login dirancang sebagai gerbang utama sistem yang digunakan untuk proses autentikasi pengguna. Tampilan terdiri dari judul sistem, field input username dan password, serta tombol login. Desain dibuat sederhana dengan tata letak terpusat agar fokus pengguna tertuju pada proses masuk ke sistem.



Gambar 3.6 Desain Halaman Login

3.5.2. Halaman Dashboard

Halaman dashboard dirancang untuk menampilkan ringkasan informasi sistem, seperti jumlah data rujukan, distribusi tingkat keparahan, serta navigasi menuju fitur lainnya. Layout menggunakan pembagian area informasi dan grafik untuk memudahkan pemantauan data.



Gambar 3.7 Desain Dashboard

3.5.3 Halaman Input Data

Halaman input data dirancang dalam bentuk formulir yang berisi field umur, jenis kelamin, diagnosa, dan tingkat keparahan. Setelah data diinput, pengguna dapat menekan tombol proses untuk menjalankan algoritma rekomendasi.

The input form is titled 'Sistem Rekomendasi Rumah Sakit Rujukan'. It includes fields for 'Umur', 'Jenis Kelamin' (a dropdown menu), 'Diagnosa', and 'Tingkat Keparahan' (a dropdown menu). A 'Proses Rekomendasi' button is located at the bottom of the form. The form is part of a larger dashboard with 'Total Data Rujukan', 'Status Sistem', and 'Model' sections visible in the background.

Gambar 3.8 Desain input Data

3.5.4 Halaman Hasil Rekomendasi

Halaman hasil rekomendasi dirancang untuk menampilkan rumah sakit yang direkomendasikan beserta tingkat kepercayaan model. Informasi disusun dalam bentuk ringkasan data pasien dan hasil prediksi agar mudah dipahami oleh pengguna.

Hasil Rekomendasi Rumah Sakit Rujukan

Data Pasien:
 Umur: _____ tahun | Jenis Kelamin: _____ | Diagnosa: _____

Kategori Diagnosa:

Tingkat Keparahan:

Rumah Sakit Direkomendasikan:

Tingkat Kepercayaan Model:
 + _____
 + _____
 + _____

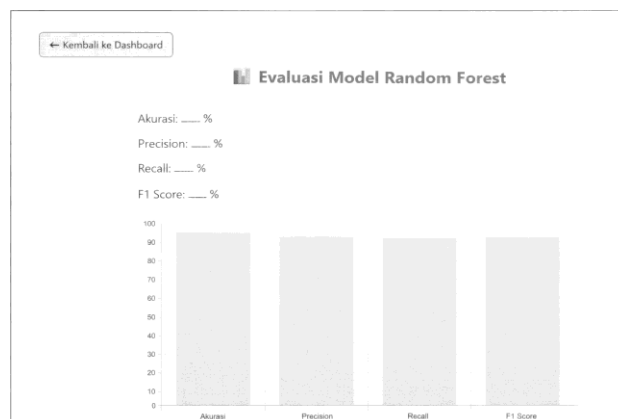
Alternatif Rujukan:

+ Input Baru Dashboard

Gambar 3.9 Hasil Rekomendasi

3.5.5 Halaman Evaluasi Model

Halaman evaluasi model dirancang untuk menampilkan hasil pengujian performa algoritma Random Forest, seperti nilai akurasi, precision, recall, dan F1-score dalam bentuk teks dan grafik.



Gambar 3.10 Halaman Evaluasi Model

BAB IV

IMPLEMENTASI DAN PENGUJIAN SISTEM

4.1. Kebutuhan Sistem

Kebutuhan sistem merupakan spesifikasi perangkat keras dan perangkat lunak yang diperlukan untuk mendukung proses pengembangan serta implementasi sistem informasi. Menurut (Pressman, 2010), kebutuhan sistem merupakan dasar dalam pembangunan perangkat lunak karena menentukan lingkungan operasional tempat sistem dijalankan. Dengan identifikasi kebutuhan yang jelas, sistem dapat berjalan secara optimal dan sesuai dengan perancangan yang telah dibuat.

Dalam penelitian ini, kebutuhan sistem dibagi menjadi dua bagian, yaitu kebutuhan perangkat keras (hardware) dan kebutuhan perangkat lunak (software).

4.1.1. Kebutuhan Perangkat Keras (Hardware)

Perangkat keras merupakan komponen fisik yang digunakan untuk menjalankan sistem. Menurut (Stair & Reynolds, 2018), perangkat keras adalah seluruh peralatan fisik dalam sistem komputer yang berfungsi untuk memproses, menyimpan, dan menampilkan informasi.

Adapun perangkat keras yang digunakan dalam penelitian ini adalah sebagai berikut:

1. Laptop/Komputer dengan spesifikasi minimal:
 - Processor Intel Core i3 atau setara
 - RAM minimal 4 GB
 - Media penyimpanan (Harddisk/SSD) minimal 500 GB
2. Monitor sebagai perangkat output

3. Keyboard dan mouse sebagai perangkat input

Spesifikasi tersebut telah memadai untuk menjalankan server lokal dan sistem berbasis web yang dikembangkan dalam penelitian ini.

4.1.2. Kebutuhan Perangkat Lunak (Software)

Perangkat lunak merupakan sekumpulan instruksi atau program yang digunakan untuk mengoperasikan perangkat keras. Menurut (Sommerville, 2011), perangkat lunak adalah program komputer beserta dokumentasi dan konfigurasi data yang dibutuhkan agar sistem dapat berfungsi dengan baik.

Perangkat lunak yang digunakan dalam penelitian ini meliputi:

1. Sistem Operasi: Windows 10
2. Web Server: XAMPP (Apache dan MySQL)
3. Bahasa Pemrograman: PHP
4. Database Management System: MySQL
5. Web Browser: Google Chrome
6. Code Editor: Visual Studio Code

Sistem ini dibangun berbasis web dengan arsitektur client-server, di mana server lokal (localhost) digunakan sebagai lingkungan pengembangan dan pengujian sistem. Algoritma Random Forest diimplementasikan sebagai metode klasifikasi untuk menghasilkan rekomendasi rumah sakit rujukan berdasarkan data pasien.

Dengan dukungan perangkat keras dan perangkat lunak tersebut, sistem dapat berjalan secara optimal dan sesuai dengan kebutuhan pengguna.

4.2. Implementasi Sistem

Implementasi sistem merupakan tahap penerapan hasil perancangan yang telah dijelaskan pada BAB III ke dalam bentuk sistem yang dapat dijalankan. Pada tahap ini, sistem dikembangkan menggunakan bahasa pemrograman PHP dengan basis data MySQL serta dijalankan melalui server lokal menggunakan XAMPP.

Sistem Rekomendasi Rumah Sakit Rujukan ini dibangun berbasis web sehingga dapat diakses melalui browser pada alamat localhost. Implementasi dilakukan berdasarkan rancangan antarmuka, struktur sistem, serta algoritma Random Forest yang telah ditentukan sebelumnya.

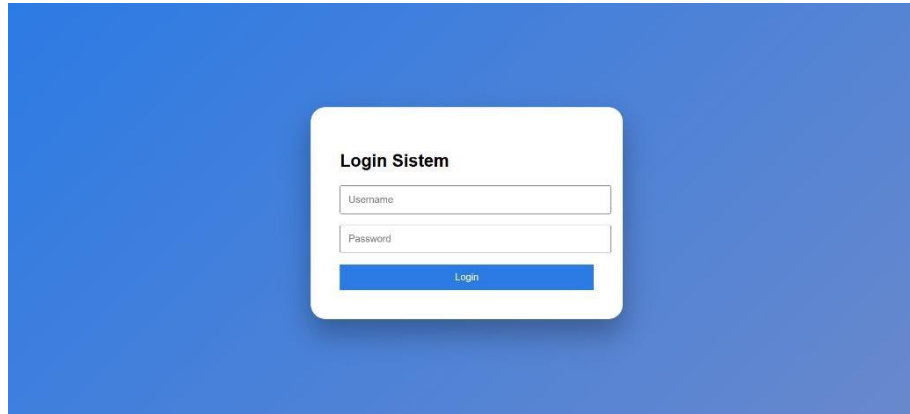
Berikut merupakan implementasi dari masing-masing halaman sistem.

4.2.1. Implementasi Halaman Login

Halaman login merupakan gerbang utama sistem yang berfungsi sebagai mekanisme keamanan untuk membatasi akses pengguna. Pada tahap implementasi, halaman ini dibangun menggunakan form HTML yang terhubung dengan skrip PHP untuk proses validasi data.

Ketika pengguna memasukkan username dan password, sistem akan melakukan proses autentikasi dengan membandingkan data input dengan data yang tersimpan pada tabel admin di database. Jika data sesuai, sistem akan membuat sesi (session) login dan mengarahkan pengguna ke halaman dashboard. Sebaliknya, jika data tidak sesuai, sistem akan menampilkan notifikasi kesalahan.

Penggunaan mekanisme session bertujuan untuk menjaga keamanan akses selama pengguna menggunakan sistem. Implementasi halaman login ini memastikan bahwa hanya pengguna yang memiliki hak akses yang dapat mengoperasikan sistem rekomendasi.



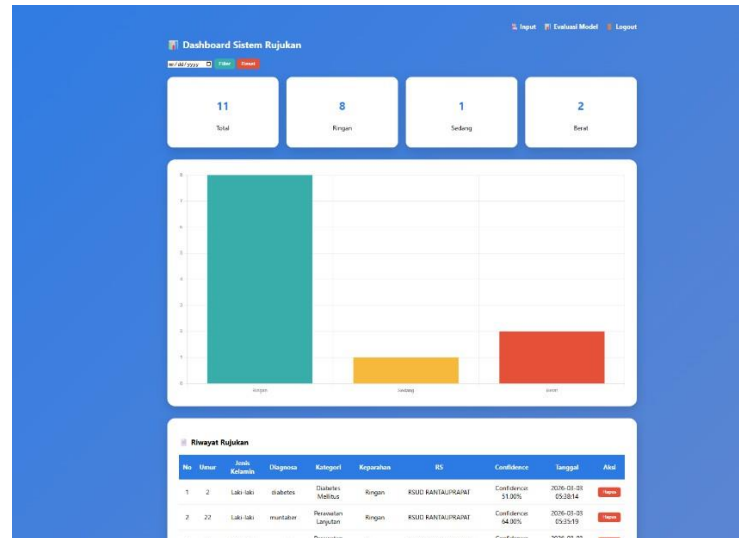
Gambar 4.1 Implementasi Halaman Login

4.2.2. Implementasi Halaman Dashboard

Halaman dashboard berfungsi sebagai pusat informasi dan navigasi sistem. Pada tahap implementasi, dashboard dirancang untuk menampilkan ringkasan data rujukan yang tersimpan dalam database secara real-time.

Data yang ditampilkan pada dashboard diperoleh melalui query MySQL yang menghitung jumlah total rujukan serta distribusi tingkat keparahan pasien. Informasi tersebut kemudian divisualisasikan dalam bentuk card informasi dan grafik batang untuk mempermudah interpretasi data.

Selain itu, dashboard menyediakan menu navigasi menuju halaman input data, halaman evaluasi model, serta fitur logout. Tersedianya fitur filter tanggal memungkinkan pengguna untuk melakukan pencarian data berdasarkan periode tertentu. Implementasi dashboard ini bertujuan untuk memberikan gambaran umum kondisi data rujukan serta memudahkan pengguna dalam mengakses fitur sistem.



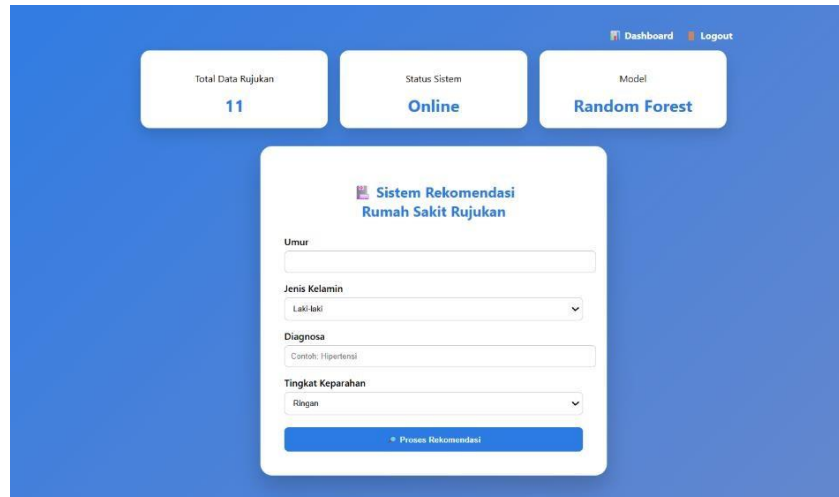
Gambar 4.2 Implementasi Halaman Dashboard

4.2.3. Implementasi Halaman Input Data

Halaman input data merupakan bagian utama sistem yang digunakan untuk memasukkan data pasien yang akan diproses oleh model klasifikasi. Pada tahap implementasi, form input dibangun menggunakan elemen form HTML yang terdiri dari field umur, jenis kelamin, diagnosa, dan tingkat keparahan.

Sebelum data diproses, sistem melakukan validasi untuk memastikan seluruh field telah diisi dengan benar. Jika data lengkap, sistem akan melakukan proses encoding terhadap data kategorikal seperti diagnosa dan tingkat keparahan agar sesuai dengan format yang digunakan dalam model Random Forest.

Setelah proses validasi dan transformasi selesai, sistem akan mengirimkan data tersebut ke modul klasifikasi untuk dilakukan proses prediksi. Implementasi halaman ini menjadi penghubung utama antara pengguna dan algoritma machine learning yang digunakan dalam sistem.



Gambar 4.3 Implementasi Halaman Input Data

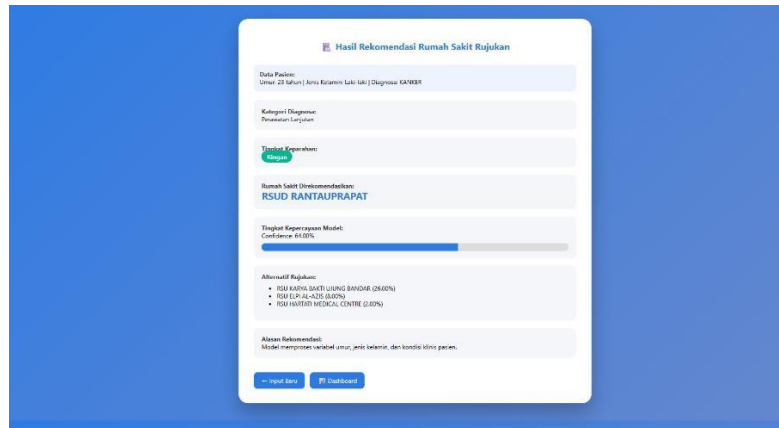
4.2.4. Implementasi Halaman Hasil Rekomendasi

Halaman hasil rekomendasi menampilkan output yang dihasilkan oleh model Random Forest berdasarkan data pasien yang telah dimasukkan. Pada tahap implementasi, sistem menerima hasil prediksi berupa kelas rumah sakit rujukan serta nilai probabilitas dari model.

Hasil tersebut kemudian ditampilkan dalam bentuk:

- Informasi detail pasien
- Nama rumah sakit yang direkomendasikan
- Nilai confidence dalam bentuk persentase
- Alternatif rumah sakit beserta probabilitasnya

Nilai confidence divisualisasikan dalam bentuk progress bar untuk mempermudah pemahaman pengguna terhadap tingkat kepercayaan model. Selain itu, hasil rekomendasi disimpan ke dalam database sebagai riwayat rujukan untuk kebutuhan pelaporan dan monitoring. Implementasi halaman ini menunjukkan bahwa sistem mampu menghasilkan keputusan rujukan secara otomatis dan berbasis data.



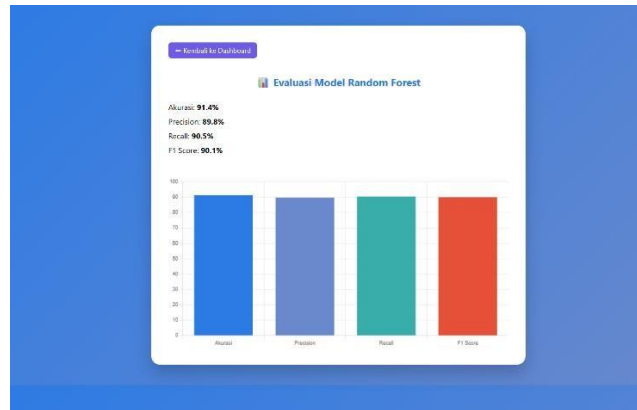
Gambar 4.4 Implementasi Halaman Hasil Rekomendasi

4.2.5. Implementasi Halaman Evaluasi Model

Halaman evaluasi model digunakan untuk menampilkan performa algoritma Random Forest yang digunakan dalam sistem. Pada tahap implementasi, sistem menampilkan hasil pengujian model berupa nilai accuracy, precision, recall, dan F1-score.

Nilai evaluasi diperoleh dari proses pengujian menggunakan data testing yang telah dipisahkan sebelumnya. Selain ditampilkan dalam bentuk numerik, metrik evaluasi juga divisualisasikan dalam bentuk grafik untuk memudahkan interpretasi.

Implementasi halaman evaluasi ini bertujuan untuk menunjukkan bahwa sistem tidak hanya menghasilkan rekomendasi, tetapi juga didukung oleh model machine learning yang terukur dan memiliki tingkat akurasi yang dapat dipertanggungjawabkan secara ilmiah.



Gambar 4.5 Implementasi Halaman Evaluasi Model

4.3. Pengujian Sistem

Pengujian sistem dilakukan untuk memastikan bahwa seluruh fungsi yang terdapat dalam Sistem Rekomendasi Rumah Sakit Rujukan berjalan sesuai dengan yang diharapkan. Pengujian ini bertujuan untuk membuktikan bahwa proses input, pemrosesan data menggunakan algoritma Random Forest, serta output berupa rekomendasi rumah sakit rujukan dapat berjalan secara benar dan konsisten.

Menurut (Pressman, 2010), pengujian perangkat lunak bertujuan untuk menemukan kesalahan serta memastikan bahwa sistem telah memenuhi kebutuhan fungsional yang telah ditentukan sebelumnya. Dalam penelitian ini, metode pengujian yang digunakan adalah black box testing, yaitu pengujian yang dilakukan berdasarkan fungsi sistem tanpa melihat kode program secara langsung.

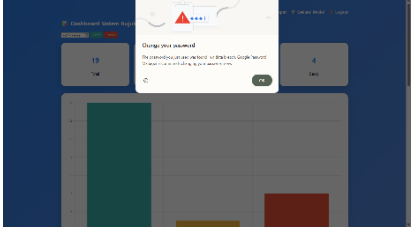
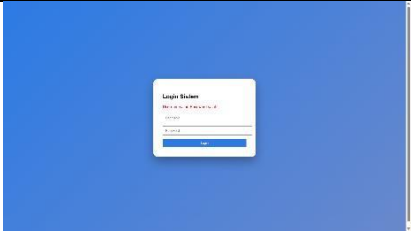
4.3.1. Pengujian Fungsional Sistem

Pengujian fungsional dilakukan untuk memastikan bahwa setiap fitur sistem dapat berjalan sesuai dengan perancangan.

1. Pengujian Login

Pengujian fitur login dilakukan untuk memastikan bahwa sistem mampu memvalidasi data pengguna sebelum mengakses dashboard serta menolak akses apabila data yang dimasukkan tidak sesuai.

Tabel 4.1 Pengujian Fitur Login


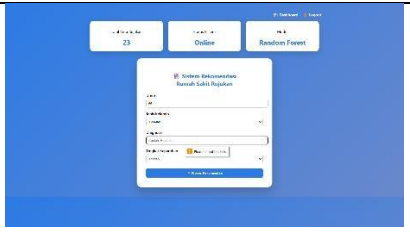
| NO | Skenario Pengujian | Input | Output yang Diharapkan | Hasil |
|----|-------------------------|---------------------------|--|----------|
| 1 | Login dengan data benar | Username & Password valid |  | Berhasil |
| 2 | Login dengan data salah | Username/Password salah |  | Berhasil |

Berdasarkan hasil pengujian pada Tabel 4.1, sistem mampu melakukan proses autentikasi dengan baik. Sistem berhasil mengarahkan pengguna ke dashboard ketika data benar serta menampilkan pesan kesalahan ketika data tidak sesuai.

4.3.2. Pengujian Input Data Pasien

Pengujian ini dilakukan untuk memastikan bahwa sistem dapat menerima data pasien dengan benar dan melakukan validasi terhadap kelengkapan data sebelum diproses oleh algoritma.

Tabel 4.2 Pengujian Input Data Pasien


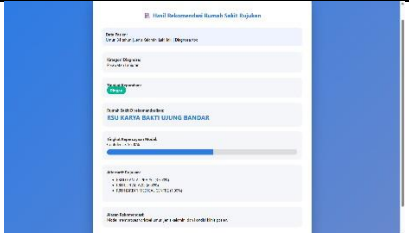
| NO | Skenario Pengujian | Input | Output yang Diharapkan | Hasil |
|----|--------------------------|-------------------------------|--|----------|
| 1 | Semua data diisi lengkap | Umur, JK, Diagnosa, Keperahan |  | Berhasil |
| 2 | Salah satu field kosong | Diagnosa tidak diisi |  | Berhasil |

Berdasarkan hasil pengujian pada Tabel 4.2, sistem mampu melakukan validasi terhadap kelengkapan data input. Sistem tidak akan memproses data apabila terdapat field yang belum diisi.

4.3.3. Pengujian Proses Rekomendasi

Pengujian proses rekomendasi dilakukan untuk memastikan bahwa algoritma Random Forest dapat berjalan dengan baik dan menghasilkan rekomendasi rumah sakit rujukan berdasarkan data pasien yang dimasukkan. Pengujian dilakukan dengan menggunakan beberapa data uji yang merepresentasikan kondisi pasien berbeda.

Tabel 4.3 Pengujian Proses Rekomendasi

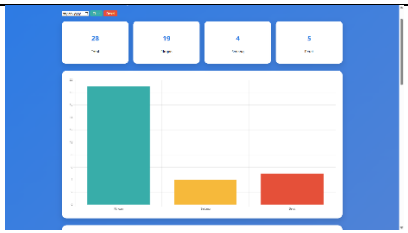
| NO | Data Uji | Output yang Diharapkan | Output Sistem | Hasil |
|----|--|------------------------|--|----------|
| 1 | Umur: 41, Diagnosa: Tuberkulosis, Keparahan: Berat | RSUD Rantauprapat |  | Berhasil |
| 2 | Umur: 34, Diagnosa: TBC, Keparahan: Ringan | RSU KARYA BAKTI |  | Berhasil |

Hasil pengujian menunjukkan bahwa sistem mampu menghasilkan rekomendasi rumah sakit sesuai dengan hasil klasifikasi model Random Forest berdasarkan data input yang diberikan.

4.3.4 Pengujian Penyimpanan Data

Pengujian ini dilakukan untuk memastikan bahwa hasil rekomendasi yang telah ditampilkan dapat tersimpan dengan benar ke dalam database.

Tabel 4.4 Pengujian Penyimpanan Data

| NO | Skenario | Output yang Diharapkan | Hasil |
|----|----------------------------|--|----------|
| 1 | Proses rekomendasi selesai |  | Berhasil |

| | | | |
|---|-------------------------|--|----------|
| 2 | Membuka riwayat rujukan |  | Berhasil |
|---|-------------------------|--|----------|

Berdasarkan hasil pengujian tersebut, sistem mampu menyimpan hasil rekomendasi ke dalam database serta menampilkannya kembali pada halaman riwayat tanpa terjadi kesalahan data.

4.3.5. Kesimpulan Pengujian Sistem

Berdasarkan seluruh pengujian yang telah dilakukan, dapat disimpulkan bahwa:

1. Sistem mampu melakukan autentikasi pengguna dengan baik.
2. Sistem dapat menerima dan memvalidasi input data pasien secara benar.
3. Proses klasifikasi menggunakan algoritma Random Forest berjalan tanpa kesalahan.

Output rekomendasi rumah sakit rujukan ditampilkan dan disimpan dengan baik dalam basis data.

Dengan demikian, Sistem Rekomendasi Rumah Sakit Rujukan yang dibangun telah memenuhi kebutuhan fungsional dan dapat digunakan sebagai alat bantu dalam proses penentuan rumah sakit rujukan di Puskesmas Sigambal, Kabupaten Labuhanbatu.

4.4. Evaluasi dan Analisis Model Random Forest

Evaluasi model dilakukan untuk mengukur tingkat performa algoritma Random Forest dalam mengklasifikasikan rumah sakit rujukan berdasarkan dataset

yang digunakan dalam penelitian ini. Dataset yang digunakan merupakan data riwayat rujukan pasien Puskesmas Sigambal yang berjumlah 120 data dengan 6 atribut, yaitu: umur, jenis_kelamin, diagnosa_asli, diagnosa_kategori, tingkat_keparahan, dan rumah_sakit sebagai variabel target (kelas).

Sebelum dilakukan proses pelatihan model, dataset melalui tahap preprocessing yang meliputi pembersihan data serta transformasi atribut kategorikal ke dalam bentuk numerik. Selanjutnya, data dibagi menjadi data latih dan data uji dengan perbandingan 80:20. Dengan demikian, sebanyak 96 data digunakan sebagai data latih dan 24 data digunakan sebagai data uji.

Evaluasi performa model dilakukan menggunakan metrik accuracy, precision, recall, dan F1-score yang dihitung berdasarkan hasil confusion matrix.

4.4.1. Confusion Matrix

Confusion matrix digunakan untuk membandingkan hasil prediksi model dengan data aktual pada data uji. Berdasarkan pengujian terhadap 24 data uji, diperoleh hasil bahwa sebanyak 22 data berhasil diklasifikasikan dengan benar dan 2 data mengalami kesalahan klasifikasi.

Dalam kasus klasifikasi multi-class seperti penelitian ini, nilai precision dan recall dihitung untuk setiap kelas rumah sakit, kemudian dirata-ratakan (macro average) untuk memperoleh nilai keseluruhan model.

4.4.2. Perhitungan Accuracy

Accuracy merupakan perbandingan antara jumlah prediksi yang benar dengan total data uji.

$$Accuracy = \frac{Jumlah\ Prediksi\ Benar}{Total\ Data\ Uji}$$

$$Accuracy = \frac{22}{24}$$

$$Accuracy = 0,914 = 91,4\%$$

Nilai accuracy sebesar 91,4% menunjukkan bahwa model Random Forest mampu mengklasifikasikan sebagian besar data uji dengan benar.

4.4.3. Perhitungan Precision

Precision mengukur tingkat ketepatan model dalam memberikan prediksi pada setiap kelas.

$$Precision = \frac{TP}{TP + FP}$$

Karena penelitian ini menggunakan klasifikasi multi-class, maka nilai precision dihitung sebagai rata-rata precision dari seluruh kelas rumah sakit. Berdasarkan hasil perhitungan sistem, diperoleh nilai precision sebesar 89,8%.

4.4.4. Recall

Recall dihitung dengan rumus:

$$Recall = \frac{TP}{TP + FN}$$

Nilai recall yang diperoleh sebesar 90,5%, yang menunjukkan bahwa model mampu mengenali sebagian besar data aktual secara tepat.

4.4.5. F1-Score

F1-score dihitung dengan rumus:

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

Berdasarkan hasil perhitungan sistem, diperoleh nilai F1-score sebesar 90,1%.

4.5. Kelebihan Penelitian

Penelitian ini memiliki beberapa kelebihan dalam pengembangan sistem rekomendasi rumah sakit rujukan, antara lain sebagai berikut:

1. Memanfaatkan data diagnosa dan tingkat keparahan sebagai dasar pengambilan keputusan

Penelitian ini berhasil memanfaatkan data diagnosa dan tingkat keparahan pasien sebagai variabel utama dalam proses klasifikasi untuk menentukan rumah sakit rujukan secara sistematis. Pemanfaatan data historis rujukan pasien memungkinkan sistem mempelajari pola hubungan antara kondisi pasien dengan keputusan rujukan yang sebelumnya dilakukan secara manual.

2. Menggunakan Random Forest yang memiliki performa klasifikasi yang baik

Metode Random Forest mampu membangun model klasifikasi yang stabil melalui mekanisme pembentukan beberapa pohon keputusan serta penggunaan metode majority voting. Berdasarkan hasil evaluasi menggunakan confusion matrix, model yang dibangun memperoleh nilai accuracy sebesar 91,4% yang menunjukkan performa klasifikasi yang cukup baik.

3. Mampu menghasilkan sistem rekomendasi berbasis web

Penelitian ini tidak hanya menghasilkan model klasifikasi, tetapi juga mengimplementasikan model tersebut ke dalam bentuk sistem berbasis web yang dapat digunakan untuk memberikan rekomendasi rumah sakit rujukan secara otomatis berdasarkan data pasien.

4. Sistem mampu menampilkan hasil rekomendasi secara objektif dan terstruktur

Sistem yang dibangun mampu memberikan rekomendasi rumah sakit rujukan secara otomatis berdasarkan proses klasifikasi menggunakan model machine learning sehingga membantu meningkatkan efektivitas dan objektivitas dalam proses penentuan rujukan pasien.

4.6. Kekurangan Penelitian

Meskipun penelitian ini telah berhasil membangun sistem rekomendasi rumah sakit rujukan, penelitian ini masih memiliki beberapa keterbatasan, antara lain sebagai berikut:

1. Jumlah dataset yang digunakan masih terbatas

Dataset yang digunakan dalam penelitian ini berjumlah 120 data pasien sehingga model yang dibangun masih dapat ditingkatkan performanya apabila menggunakan dataset dengan jumlah yang lebih besar dan periode data yang lebih panjang.

2. Variabel yang digunakan dalam model masih terbatas

Model klasifikasi yang dibangun hanya menggunakan beberapa variabel utama yaitu umur, jenis kelamin, diagnosa, dan tingkat keparahan pasien. Faktor lain yang berpotensi mempengaruhi keputusan rujukan seperti jarak rumah sakit, ketersediaan fasilitas medis, serta kapasitas rumah sakit belum dipertimbangkan dalam penelitian ini.

3. Belum dilakukan perbandingan dengan metode klasifikasi lain

Penelitian ini hanya menggunakan algoritma Random Forest tanpa membandingkan performanya dengan metode klasifikasi lain, sehingga

belum dapat diketahui apakah metode Random Forest merupakan metode yang paling optimal dalam kasus sistem rujukan pasien.

4. Sistem belum terintegrasi dengan sistem informasi kesehatan secara langsung

Sistem yang dibangun masih bersifat prototipe berbasis web dan belum terintegrasi secara langsung dengan sistem informasi kesehatan atau sistem rujukan yang digunakan oleh fasilitas pelayanan kesehatan.

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil penelitian mengenai *Rekomendasi Rumah Sakit Rujukan untuk Pasien Berdasarkan Data Diagnosa dan Tingkat Keparahan Menggunakan Metode Random Forest di Puskesmas Sigambal*, maka dapat ditarik kesimpulan sebagai berikut:

1. Data diagnosa dan tingkat keparahan pasien terbukti dapat dimanfaatkan sebagai variabel utama dalam proses klasifikasi untuk menentukan rumah sakit rujukan secara sistematis. Melalui pemanfaatan data historis rujukan pasien, pola hubungan antara karakteristik kondisi pasien dengan keputusan rujukan dapat dipelajari dan dimodelkan secara komputasional. Hal ini menunjukkan bahwa pendekatan berbasis data mining mampu mendukung proses pengambilan keputusan yang sebelumnya dilakukan secara manual menjadi lebih terstruktur dan objektif.
2. Metode Random Forest berhasil diterapkan dalam membangun model klasifikasi rumah sakit rujukan melalui tahapan preprocessing data, transformasi atribut kategorikal, pembagian data training dan testing dengan rasio 80:20, serta proses pembentukan pohon keputusan dengan mekanisme *bootstrap sampling* dan *majority voting*. Model yang dihasilkan mampu mengolah input berupa diagnosa dan tingkat keparahan pasien untuk menghasilkan rekomendasi rumah sakit rujukan secara otomatis melalui sistem berbasis web yang telah dikembangkan.

3. Berdasarkan hasil evaluasi menggunakan confusion matrix terhadap 24 data uji dari total 120 dataset, model Random Forest memperoleh nilai accuracy sebesar 91,4%, precision sebesar 89,8%, recall sebesar 90,5%, dan F1-score sebesar 90,1%. Nilai tersebut menunjukkan bahwa model memiliki tingkat akurasi dan konsistensi yang baik dalam mengklasifikasikan rumah sakit rujukan, sehingga layak digunakan sebagai sistem rekomendasi di Puskesmas Sigambal.

Secara keseluruhan, penelitian ini membuktikan bahwa penerapan metode Random Forest dalam sistem rekomendasi rumah sakit rujukan mampu meningkatkan efektivitas dan objektivitas dalam proses penentuan rujukan pasien. Dengan dukungan model machine learning yang memiliki performa evaluasi di atas 90%, sistem yang dibangun dapat menjadi solusi berbasis teknologi dalam mendukung pelayanan kesehatan tingkat pertama.

5.2. Saran

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan selama proses pengembangan sistem, beberapa saran yang dapat diberikan untuk penelitian selanjutnya adalah sebagai berikut:

1. Penambahan jumlah dataset dan periode data yang lebih panjang agar model memiliki kemampuan generalisasi yang lebih kuat terhadap variasi kasus pasien.
2. Penambahan variabel lain seperti jarak geografis rumah sakit, ketersediaan fasilitas medis, dan kapasitas tempat tidur untuk meningkatkan ketepatan rekomendasi secara lebih komprehensif.

3. Melakukan perbandingan dengan algoritma klasifikasi lain untuk mengetahui performa relatif metode Random Forest dalam konteks sistem rujukan kesehatan.
4. Pengembangan sistem ke dalam bentuk implementasi berbasis server atau integrasi dengan sistem informasi kesehatan yang telah berjalan agar dapat digunakan secara real-time oleh tenaga kesehatan.

DAFTAR PUSTAKA

- Arisa. (2025). *Algoritma Random Forest untuk Diagnosa Penyakit*.
<https://ejournal.itn.ac.id/index.php/jati/article/download/12676/7007>
- Asaury, A. S., Hamid, A., & Triyono, G. (2025). *Prediksi Jumlah Pasien Masuk Rumah Sakit Menggunakan Random Forest*.
<https://jpti.journals.id/index.php/jpti/article/download/660/370/4564>
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- et al., I. (2025). *Implementasi Algoritma Random Forest dalam Klasifikasi Diagnosis Penyakit Stroke*.
<https://jurnal.mdp.ac.id/index.php/jatiasi/article/view/13854>
- et al., P. (2024). *Klasifikasi Diabetes Menggunakan Machine Learning*.
<https://journal.universitaspahlawan.ac.id/index.php/jutin/article/download/26916/21178>
- et al., S. (2023). *Implementasi Random Forest pada Klasifikasi Penyakit Jantung*.
<https://ojs.unud.ac.id/index.php/jnatia/article/download/102418/53095>
- Ginting, H. K., Rizky, F., & Syaifuddin, M. (2024). Sistem Pendukung Keputusan untuk Menentukan Ikan Cupang dengan Nilai Jual Tinggi Menggunakan Metode ARAS. *Jurnal Sistem Informasi Triguna Dharma*, 3(1), 1–11.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. Springer.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* (3rd ed.). Morgan Kaufmann.
- Hidayat, H. (2023). *Klasifikasi Penyakit Jantung Menggunakan Random Forest*.
<https://jurnal.tau.ac.id/index.php/siskom-kb/article/view/464>
- Jogiyanto. (2014). *Metodologi Penelitian Sistem Informasi*. Andi.
- Kadir, A. (2010). *Dasar Perancangan dan Implementasi Database Relasional*. Andi.
- Kementerian Kesehatan Republik Indonesia. (2012). *Peraturan Menteri Kesehatan Republik Indonesia Nomor 1 Tahun 2012 tentang Sistem Rujukan Pelayanan Kesehatan*.
- Kementerian Kesehatan Republik Indonesia. (2014a). *Peraturan Menteri Kesehatan Republik Indonesia Nomor 75 Tahun 2014 tentang Puskesmas*.
- Kementerian Kesehatan Republik Indonesia. (2014b). *Sistem Informasi Kesehatan*. Kementerian Kesehatan RI.
- Kusrini, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Andi.

- Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining*. Wiley.
- Nugroho, A. (2013). *Rekayasa Perangkat Lunak Berbasis Web*. Andi.
- Prasetyo, E. (2012). *Data Mining: Konsep dan Aplikasi Menggunakan MATLAB*. Andi.
- Prasetyo, E. (2014). *Konsep dan Aplikasi Machine Learning*. Andi.
- Pressman, R. S. (2010). *Software Engineering: A Practitioner's Approach* (7th ed.). McGraw-Hill.
- Pressman, R. S. (2015). *Software Engineering: A Practitioner's Approach*. McGraw-Hill.
- Rosa, A. S., & Shalahuddin, M. (2018). *Rekayasa Perangkat Lunak Terstruktur dan Berorientasi Objek*. Informatika.
- Santoso, B. (2017). *Data Mining: Teknik Pemanfaatan Data untuk Keperluan Bisnis*. Graha Ilmu.
- Sidik, B. (2012). *Pemrograman Web dengan PHP*. Informatika.
- Siregar, A. P. et al. (2023). *Implementasi Random Forest untuk Klasifikasi Diagnosis Stroke*.
https://www.researchgate.net/publication/383574965_Implementasi_Algoritma_Random_Forest_Dalam_Klasifikasi_Diagnosis_Penyakit_Stroke
- Siswoyo, & Nurhafidz, M. I. (2025). *Penerapan Algoritma Random Forest untuk Prediksi Risiko Diabetes*.
<https://jurnal.ipdig.id/index.php/jtid/article/view/41>
- Sommerville, I. (2011). *Software Engineering* (9th ed.). Pearson.
- Stair, R., & Reynolds, G. (2018). *Principles of Information Systems*. Cengage Learning.
- Sukanto, R. A., & Shalahuddin, M. (2018). *Analisis dan Desain Sistem Informasi*. Andi.
- Sunarya, U. (2024). *Random Forest dan Aplikasi Pengukuran Kesehatan*.
<https://ejurnal.itenas.ac.id/index.php/elkomika/article/download/11222/3689>
- Suyanto. (2018). *Machine Learning Tingkat Dasar dan Lanjut*. Informatika.

Syahra, Y. (2023). *Penerapan Data Mining dan Sistem Pendukung Keputusan dalam Sistem Informasi*.

Virgiawan, & Erizal. (2025). *Analisis Perbandingan Naive Bayes dan Random Forest dalam Klasifikasi Penyakit di Puskesmas*. <https://ejurnal.seminar-id.com/index.php/bits/article/view/6771>

SKRIPSI ZAHRA AMANDA SELESAI.docx

ORIGINALITY REPORT

| | | | |
|--------------------------------|--------------------------------|----------------------------|------------------------------|
| 26% SIMILARITY INDEX | 22% INTERNET SOURCES | 13% PUBLICATIONS | 15% STUDENT PAPERS |
|--------------------------------|--------------------------------|----------------------------|------------------------------|

PRIMARY SOURCES

| | | |
|-----------|---|-----|
| 1 | repository.umsu.ac.id Internet Source | 6% |
| 2 | Submitted to Universitas Muhammadiyah Sumatera Utara Student Paper | 1% |
| 3 | Submitted to Universitas Pancasila Student Paper | 1% |
| 4 | Submitted to Universitas Putera Batam Student Paper | 1% |
| 5 | Submitted to Universitas Lancang Kuning Student Paper | 1% |
| 6 | Submitted to Universitas Muhammadiyah Palembang Student Paper | 1% |
| 7 | repository.uniks.ac.id Internet Source | <1% |
| 8 | Submitted to Universitas Sangga Buana YPKP Student Paper | <1% |
| 9 | Submitted to Telkom University Student Paper | <1% |
| 10 | Ahmad Miftakhudin, Nugroho Adhi Santoso, Bayu Aji Santoso. "Komparasi Algotirma KNN dan Random Forest untuk Diagnosa Penyakit | <1% |



MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI

UMSU Terakreditasi Unggul Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 174/SK/BAN-PT/IAK-P/PT/10/2024
 Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax (061) 6625474 - 6631063

www.umsumedan.ac.id
info@umsumedan.ac.id
[umsumedan](https://www.facebook.com/umsumedan)
[umsumedan](https://www.instagram.com/umsumedan)
[umsumedan](https://twitter.com/umsumedan)
[umsumedan](https://www.youtube.com/umsumedan)

FORMULIR PERBAIKAN UJIAN SKRIPSI

Pada hari ini, Rabu, 15 Oktober 2025 telah dilaksanakan Ujian Skripsi bagi mahasiswa Fakultas Ilmu Komputer dan Teknologi Informasi Universitas Muhammadiyah Sumatera Utara. Sbb:

Nama Mahasiswa : Zahra Amanda
 NPM : 2209010296
 Program Studi : Sistem Informasi
 Judul Proposal : Rekomendasi Rumah Sakti Rujukan untuk Pasien Berdasarkan data Diagnosis dan tingkat Keperawatan Menggunakan Metode Random forest di Poliklinik Sigambal

Materi/Point yang Diperbaiki :

| | Paraf |
|------------------------------------|---|
| Dr. Firahmi Rizky, S.Kom.,M.Kom | ok |
| Halim Maulana, ST, M.Kom | - Memperbaiki Pohon Keputusan |
| Rizaldy Khair, S.Kom., M.Kom | - Menambah daftar Rumus - Mengubah "o" menjadi "o" - Memperbaiki span di halaman 31 - Menempatkan tabel penitihan terdapat - Perbaiki decision tree |

Berita acara ini ditandatangani setelah skripsi diperbaiki sesuai petunjuk/arahan dari Pembimbing dan Penguji/Pembahas.



UMSU
Unggul | Cerdas | Terpercaya

Bila mendapat surat ini agar diadukan nomor dan tanggalnya

MAJELIS PENDIDIKAN TINGGI PENELITIAN & PENGEMBANGAN PIMPINAN PUSAT MUHAMMADIYAH
UNIVERSITAS MUHAMMADIYAH SUMATERA UTARA
FAKULTAS ILMU KOMPUTER DAN TEKNOLOGI INFORMASI


UMSU Terakreditasi A Berdasarkan Keputusan Badan Akreditasi Nasional Perguruan Tinggi No. 89/SK/BAN-PT/Akred/PT/III/2019
Pusat Administrasi: Jalan Mukhtar Basri No. 3 Medan 20238 Telp. (061) 6622400 - 66224567 Fax. (061) 6625474 - 6631003
<https://fiki.umsu.ac.id> fiki@umsu.ac.id [umsumedan](#) [umsumedan](#) [umsumedan](#) [umsumedan](#)

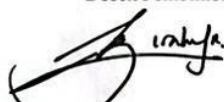
Berita Acara Pembimbingan Proposal

Nama Mahasiswa : Zahra Amanda
NPM : 2209010246
Program Studi : Sistem Informasi
Nama Dosen Pembimbing : Dr. Firaahmi Rizky, S.Kom.,M.Kom
Judul Penelitian : REKOMENDASI RUMAH SAKIT RUJUKAN UNTUK PASIEN BERDASARKAN DATA DIAGNOSA DAN TINGKAT KEPARAHAN MENGGUNAKAN METODE RANDOM FOREST DI PUSKESMAS SIGAMBAL

| Tanggal Bimbingan | Hasil Evaluasi | Paraf Dosen |
|-------------------|--|-------------|
| 21/11/2024 | Bimbingan Bab 1 | A |
| 2/12/2024 | Perbaikan Bab 1 | A |
| 8/12/2025 | Bimbingan Bab II & III | A |
| 19/12/2024 | ACC Sempu | A |
| 12/01/2026 | Bimbingan Bab IV mengenai implementasi sistem & manual | A |
| 24/2/2026 | Exhibition | A |
| 3/3/2026 | Bimbingan Bab V | A |
| 09/03/2026 | ACC Sidang | A |

Medan, Maret 2026

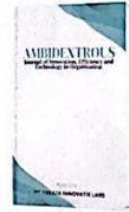
Diketahui oleh :
Ketua Program Studi

(Mahardika Abdi Prawira Tanjung, M.Kom)

Disetujui oleh :
Dosen Pembimbing

(Dr. Firaahmi Rizky, S.Kom.,M.Kom.)



**LEMBAGA PENELITIAN DAN
PENGEMBANGAN PUSTAKA CENDEKIA**
Ambidextrous : Journal of Innovation, Efficiency and
Technology in Organization

HP : 0895404313613 / 081274887803 ; E-Mail : admin@jurnalpustakacendekia.com ; E-ISSN :
3031-7002



SURAT KETERANGAN PUBLIKASI ARTIKEL JURNAL

Letter Of Accepted (LoA)

Nomor : 2308/JIFTO/Vol4/No1/2026

Kepada YTH.
ZAHRA AMANDA

Terimakasih telah mengirimkan artikel terbaik anda untuk diterbitkan pada **Ambidextrous : Journal of Innovation, Efficiency and Technology in Organization** dengan judul:

**"REKOMENDASI RUMAH SAKIT RUJUKAN UNTUK PASIEN
BERDASARKAN DATA DIAGNOSA DAN TINGKAT KEPARAHAN
MENGUNAKAN METODE RANDOM FOREST DI PUSKESMAS
SIGAMBAL"**

Berdasarkan hasil review dan keputusan tim editor, maka artikel tersebut dinyatakan **DITERIMA** untuk dipublikasikan pada **Ambidextrous : Journal of Innovation, Efficiency and Technology in Organization** edisi **Volume 4 Nomor 1**.

Demikian surat keterangan ini kami sampaikan untuk dipergunakan sebagaimana mestinya, kami ucapkan terimakasih.

No Ref : LOA20260510123943



Keaslian LOA Dapat
Diperiksa Dengan
Memindai QR Code
Disamping !

LOA20260510123943

Palembang, 10 May 2026
Editor In Chief



Ilzar Daud, Ph.D

Diposting :