

Prediksi Jumlah Pasien Masuk Rumah Sakit Menggunakan Metode Random Forest

Akhmad Sufyan Asaury^{*1}, Abdul Hamid², Gandung Triyono³

^{1,2}Sistem Informasi, Universitas Sapta Mandiri, Indonesia

³Ilmu Komputer, Universitas Budi Luhur, Jakarta, Indonesia

Email: sufyanasaury@univsm.ac.id, hamied@univsm.ac.id, gandung.triyono@budiluhur.ac.id

Abstrak

Kesehatan merupakan kebutuhan utama masyarakat pada masa kini, sehingga peningkatan sarana pelayanan kesehatan menjadi hal yang krusial untuk mencapai derajat kesehatan yang optimal sesuai dengan tujuan pembangunan kesehatan. Rumah sakit sering menghadapi lonjakan pasien setiap harinya, terutama pada instalasi rawat inap. Prediksi jumlah pasien masuk seringkali menggunakan metode sederhana seperti rata-rata historis yang kurang akurat, sehingga menyebabkan ketidakefisienan dalam pengelolaan sumber daya, pengendalian biaya, dan penyediaan layanan berkualitas. Data yang digunakan dalam penelitian ini adalah data bangsal dari Januari 2021 hingga Desember 2022, yang diperoleh dalam format Excel dengan total 730 data selama dua tahun. Penelitian ini bertujuan untuk mengembangkan model prediksi berbasis Random Forest yang dapat memperkirakan jumlah pasien masuk rumah sakit secara akurat dengan mempertimbangkan berbagai faktor, seperti demografi pasien dan variabel waktu. Proses penelitian melibatkan pengumpulan data, preprocessing untuk membersihkan dan mempersiapkan dataset, pelatihan model, serta evaluasi kinerja menggunakan metrik seperti Mean Squared Error (MSE), Mean Absolute Error (MAE), dan Root Mean Squared Error (RMSE). Hasil evaluasi menunjukkan bahwa model prediksi memiliki performa yang baik dengan nilai MSE sebesar 0,64, MAE sebesar 0,60, dan RMSE sebesar 0,80, yang mengindikasikan tingkat kesalahan prediksi yang rendah. Informasi prediksi jumlah pasien ini memberikan dampak signifikan bagi manajemen rumah sakit dalam mengelola kapasitas, seperti alokasi tempat tidur, tenaga medis, dan peralatan medis, sehingga efisiensi operasional dapat ditingkatkan. Dengan demikian, metode prediksi berbasis Random Forest tidak hanya memberikan hasil yang akurat, tetapi juga dapat diterapkan secara langsung untuk mendukung pengambilan keputusan strategis di lingkungan rumah sakit.

Kata kunci: *Machine learning, prediksi pasien, Random Forest, rumah sakit.*

PREDICTION OF THE NUMBER OF PATIENTS ADMITTED TO HOSPITAL USING THE RANDOM FOREST METHOD

Abstract

Health is a fundamental need of society today, making the improvement of healthcare facilities crucial to achieving optimal health status in line with health development goals. Hospitals often face daily surges in patient admissions, especially in inpatient wards. Patient admission predictions frequently rely on simple methods such as historical averages, which are often inaccurate and lead to inefficiencies in resource management, cost control, and the provision of quality services. The data used in this study consists of ward data from January 2021 to December 2022, obtained in Excel format, totaling 730 records over two years. This study aims to develop a prediction model based on Random Forest to accurately forecast hospital patient admissions by considering various factors, such as patient demographics and time variables. The research process includes data collection, preprocessing to clean and prepare the dataset, model training, and performance evaluation using metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), and Root Mean Squared Error (RMSE). The evaluation results indicate that the prediction model performs well, with an MSE of 0.64, MAE of 0.60, and RMSE of 0.80, demonstrating a low prediction error rate. This patient admission prediction information has a significant impact on hospital management in capacity planning, such as the allocation of beds, medical staff, and medical equipment, thereby enhancing operational efficiency. Thus, the Random Forest-based prediction method not only provides accurate results but can also be directly applied to support strategic decision-making in the hospital environment..

Keywords: *Machine learning, patient prediction, Random Forest, hospital.*

1. PENDAHULUAN

Kesehatan merupakan suatu kebutuhan utama masyarakat pada masa sekarang. Oleh karena itu, peningkatan sarana pelayanan dibidang kesehatan harus dilakukan, agar masyarakat mendapatkan tingkat derajat kesehatan yang optimal sesuai dengan tujuan pembangunan kesehatan. Menurut Peraturan Pemerintah RI No 47 Tahun 2021 rumah sakit merupakan institusi pelayanan kesehatan yang menyelenggarakan pelayanan kesehatan perorangan secara paripurna yang menyediakan pelayanan rawat inap, rawat jalan dan gawat darurat. Pada pelaksanaan pelayanan kesehatan, rumah sakit harus mengadakan rekam medis untuk menunjang pelayanan dan pengolahan informasi pasien[1]. Rumah sakit adalah fasilitas kesehatan yang memberikan perawatan medis kepada individu secara menyeluruh, termasuk rawat jalan, rawat inap, dan gawat darurat[2]. Salah satu aspek penting yang harus diperhatikan oleh rumah sakit adalah pelayanan bagi pasien rawat inap. Oleh karena itu, rumah sakit perlu memiliki data peramalan untuk jumlah pasien rawat inap dalam beberapa hari ke depan. Dengan menggunakan data tersebut, rumah sakit dapat mengawasi dan merencanakan kebutuhan pasien di masa mendatang. Dalam melakukan peramalan, penekanan khusus diberikan pada nilai kesalahan yang diperoleh. Untuk mencapai hasil peramalan yang akurat, diperlukan penerapan metode peramalan yang efektif. Semakin kecil nilai kesalahan yang diperoleh, semakin baik pula hasil peramalan yang dihasilkan. Keberhasilan peramalan yang baik dapat memberikan manfaat signifikan bagi rumah sakit dalam meningkatkan dan mengembangkan layanan yang diberikan kepada pasien[3]. Seiring bertambahnya jumlah populasi manusia dan keadaan perekonomian yang semakin maju, maka kesadaran masyarakat terhadap kesehatan semakin meningkat. Hal ini dapat meningkatkan jumlah pengunjung suatu rumah sakit. Meningkatnya kunjungan pasien harus di iringi dengan kesiapan pihak rumah sakit dalam pemenuhan fasilitas dan alat penunjang pelayanan, seperti halnya yang dijelaskan oleh Peraturan Menteri Kesehatan Republik Indonesia Nomor 56 Tahun 2014 tentang Klasifikasi dan Perizinan Rumah Sakit[4]. Rumah sakit mempunyai tugas utama yaitu memberikan pelayanan kesehatan yang berkualitas tinggi, dengan fokus pada upaya kuratif dan rehabilitatif. Untuk meningkatkan mutu pelayanan kesehatan, rumah sakit membutuhkan sistem pencatatan informasi kesehatan yang efektif yang dikelola oleh unit rekam medis terintegrasi. Berkas rekam medis yang terdokumentasi dengan baik dapat dimanfaatkan sebagai statistik data untuk melaporkan kinerja rumah sakit. Pelaporan kinerja ini diperlukan untuk melakukan analisis dan evaluasi yang tepat, sehingga dapat diambil keputusan yang lebih baik dalam meningkatkan proses pelayanan Kesehatan[5].

Berdasarkan data jumlah kunjungan pasien rawat jalan tahun 2017-2019 mengalami peningkatan dan penurunan tahun 2020-2021 dikarenakan virus covid-19 yang menyebabkan pengunjung takut berobat jalan agar tidak terpapar covid-19. Sehingga perlu melakukan prediksi kunjungan pasien rawat jalan. Mengingat pentingnya data kunjungan rawat jalan yang berguna untuk mengetahui sejauh mana perkembangan dari pelayanan yang diberikan kepada pasien dan dapat digunakan sebagai tolak ukur manajemen untuk merancang langkah evaluasi dalam meningkatkan pelayanan pasien rawat jalan[1]. Ketersediaan pelayanan untuk pasien adalah hal yang penting, khususnya di bagian pelayanan rawat inap. Oleh sebab itu, perlu dilakukan prediksi jumlah pasien rawat inap, sehingga rumah sakit mempunyai bayangan kebutuhan pasien kedepannya. Metode peramalan yang digunakan yaitu Triple Exponential Smoothing, metode tersebut cukup sederhana serta memiliki kelebihan mampu menyesuaikan tren perubahan dan komponen musiman[3]. Kepadatan pasien rumah sakit telah menjadi masalah utama yang dihadapi oleh rumah sakit besar. Efek samping bagi rumah sakit meningkatnya kepadatan dan memiliki efek lebih lanjut pada kepuasan pasien, kualitas keperawatan, pengobatan, waktu tunggu, dan lama tinggal. Menentukan secara akurat pasokan dan permintaan tempat tidur rumah sakit untuk penerimaan baru dapat membantu mencegah hasil yang merugikan pasien. Analisis kuantitatif data rekam medis elektronik modern dapat membantu memprediksi penawaran dan permintaan untuk tempat tidur rumah sakit staf yang kosong (SED) dan membantu menghilangkan perkiraan manusia, menstandarkan pekerjaan sehari-hari melalui data yang konkret dan objektif[6].

Satu-satunya variabel independen dalam fenomena kepadatan ini adalah permintaan tempat tidur, yang secara alami berfluktuasi untuk menyatakan musim flu, Musim Liburan, dll. Mempelajari perubahan permintaan tempat tidur rawat inap ini adalah masalah yang dibahas dan dipelajari secara luas yang berdampak pada kemampuan rumah sakit untuk menyediakan perawatan tepat waktu. Dengan merancang dan menerapkan keputusan manajemen kapasitas rumah sakit yang efektif dan keputusan kepegawaian yang efisien merupakan tantangan penting dalam setiap sistem perawatan kesehatan. Secara khusus, ketidaksesuaian dalam kapasitas tempat tidur dengan permintaan tempat tidur dan persyaratan staf klinis yang sesuai dapat terjadi efek negatif pada indikator kinerja utama seperti akses rumah sakit, waktu tunggu, kualitas perawatan serta kepuasan pasien dan karyawan[7].

Manajemen prediksi dan pendeteksian yang efektif dapat membantu mengurangi kepadatan. Sudah berhasil mengusulkan model hybrid yang menggabungkan autoregressive integrated moving average (ARIMA) dan model nonlinear autoregressive neural network (NARNN) dalam studi peramalan penyakit schistosomiasis dan tangan, kaki, dan mulut. Dalam penelitian ini, bertujuan untuk mengeksplorasi penerapan model hybrid ARIMA-NARNN untuk melacak tren pasien rawat inap baru, yang memberikan dasar metodologis untuk mengurangi kepadatan[8].

Menggunakan metode Random Forest dengan perolehan informasi sebagai kriteria karena memiliki keunggulan menghasilkan akurasi yang tinggi, memiliki kesalahan yang rendah serta dapat menangani dataset training yang besar dan missing value pada dataset. Disimpulkan bahwa untuk mencapai kualitas pelayanan medis khususnya efisiensi tempat tidur rumah sakit, penelitian ini akan menerapkan metode random forest untuk memprediksi jumlah tempat tidur rumah sakit menurut kriteria Barber Johnson[9].

Data mining adalah proses menganalisa data dari berbagai sudut pandang dan menghasilkan informasi penting yang dapat digunakan untuk menurunkan biaya, meningkatkan keuntungan, atau bahkan keduanya. Secara teknis, data mining dapat didefinisikan sebagai proses untuk menemukan pola atau korelasi antara ratusan atau ribuan field dalam database yang sangat besar. Menghasilkan informasi yang dapat digunakan untuk meningkatkan daya saing suatu organisasi adalah salah satu keuntungan dari penggunaan data mining. Pada dasarnya, prediksi adalah dugaan atau prediksi tentang suatu kejadian atau peristiwa yang akan datang[10].

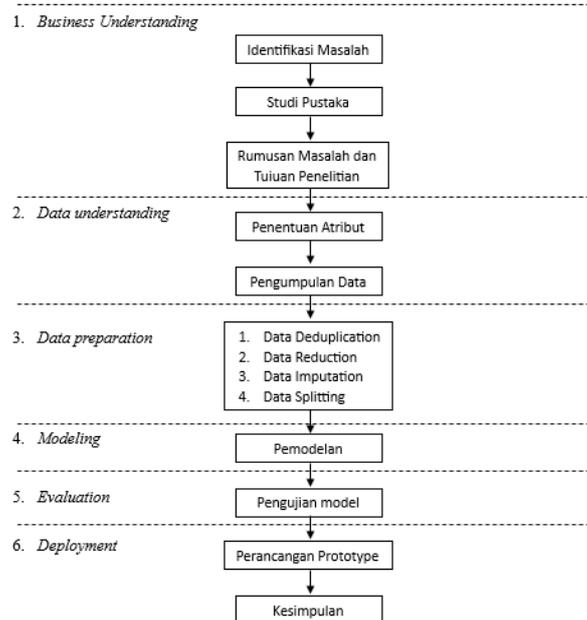
Metode random forest merupakan pengembangan dari decision tree yang menggunakan multiple decision tree. Setiap decision tree dalam random forest dilatih dengan menggunakan sampel individu dan subset acak dari atribut yang dipilih dalam setiap pohon. Saat melakukan klasifikasi, setiap individu diklasifikasikan berdasarkan mayoritas suara yang diperoleh dari kumpulan pohon dalam populasi. Dengan demikian, random forest memanfaatkan voting untuk menghasilkan prediksi berdasarkan suara terbanyak dari pohon-pohon dalam ensemble[11]. Metode Random Forest memiliki kemampuan untuk mengklasifikasikan data yang mengandung atribut yang tidak lengkap. Metode ini dapat digunakan baik untuk klasifikasi maupun regresi, dan mampu menghasilkan tingkat akurasi yang tinggi. Selain itu, Random Forest juga efektif dalam mengatasi jumlah data yang besar dengan cara yang efisien. Metode ini menghasilkan ensemble Decision Tree yang terdiri dari root node, internal node, dan leaf node. Proses pembentukan Decision Tree dilakukan dengan memilih atribut dan data secara acak, mengikuti aturan yang telah ditetapkan[12]. Algoritma random forest efektif dalam menangani nilai yang hilang tetapi rentan terhadap overfitting. Penyesuaian parameter yang tepat dapat diterapkan untuk menghindari overfitting[13]. Algoritma metode Random Forest yang terdiri dari dua bagian. Yang pertama adalah algoritma untuk membuat "n" pohon secara acak untuk membentuk hutan, dan yang kedua adalah algoritma untuk melakukan prediksi dengan menggunakan Random Forest yang telah dibuat sebelumnya[14].

Dalam dua tahun terakhir berdasarkan data internal yang diambil, Rumah Sakit telah mengalami perubahan yang signifikan dalam pola kunjungan pasien di berbagai unit pelayanan. Data statistik menunjukkan bahwa di Instalasi Gawat Darurat (IGD), terjadi variasi yang mencolok dalam jumlah kunjungan pasien. Pada bangsal kebidanan, Rumah Sakit menerima sekitar 2.044 pasien. Dalam konteks penelitian di Rumah Sakit, masalah penggunaan sumber daya, mengendalikan biaya, dan memberikan pelayanan berkualitas sebagai dampak dari prediksi yang tidak akurat terhadap jumlah pasien masuk menjadi semakin relevan. Dengan mengidentifikasi faktor-faktor yang berkontribusi terhadap jumlah pasien masuk dan mengembangkan model prediksi yang akurat, diharapkan dapat memberikan kontribusi penting dalam mengatasi masalah-masalah yang ada dan memastikan bahwa penggunaan sumber daya, mengendalikan biaya, dan memberikan pelayanan berkualitas dapat berjalan dengan lebih baik.

Random Forest dipilih karna merupakan algoritma yang efektif dan fleksibel, terutama dalam menangani nilai yang hilang dan masalah data yang kompleks. Namun, kerentanan terhadap overfitting dapat muncul jika parameter tidak disesuaikan dengan benar. Penyesuaian parameter seperti max_depth, min_samples_split, dan max_features menjadi langkah penting untuk menjaga keseimbangan antara akurasi prediksi dan kemampuan generalisasi model. Dengan penerapan praktik terbaik, Random Forest dapat menjadi alat yang sangat andal dalam prediksi. Penelitian ini bertujuan untuk mengembangkan model prediksi jumlah pasien masuk rumah sakit dengan algoritma Random Forest guna membantu manajemen dalam optimalisasi sumber daya dan tempat tidur pasien.

2. METODE PENELITIAN

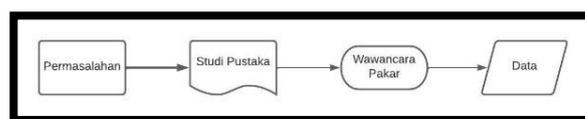
Tahapan penelitian dalam Data Mining menggunakan metodologi CRISP-DM, Suatu standar yang dikenal sebagai Cross-Industry Standard Process for Data Mining (CRISP-DM) dibuat pada tahun 1996 dan dimaksudkan untuk menggunakan proses analisis industri sebagai strategi pemecahan masalah perusahaan atau unit penelitian. Data yang diproses oleh CRISP-DM ini tidak memiliki karakteristik atau ketentuan khusus karena akan diproses kembali dalam fase-fase yang berbeda[15]. Tahapan dalam CRISP-DM adalah sebagai berikut:



Gambar 1. CRISP-DM dalam data mining

2.1. Business Understanding

Tahapan ini merupakan langkah untuk memahami latar belakang permasalahan yang dihadapi di rumah sakit, dengan tujuan menghasilkan pertanyaan penelitian dan tujuan penelitian data mining. Pada tahap ini, penelitian dilakukan melalui observasi, wawancara, dan studi pustaka terkait permasalahan yang ada. Hal ini dilakukan untuk mendapatkan pemahaman yang lebih mendalam mengenai langkah-langkah penyelesaian permasalahan yang akan dilakukan. Wawancara awal dilakukan untuk mengumpulkan informasi terkait kondisi jumlah pasien masuk yang menjadi permasalahan. Selanjutnya dilakukan studi pustaka untuk mengetahui penelitian terkait yang telah dilakukan sebelumnya. Studi pustaka yang terkumpul digunakan untuk landasan acuan sebagai referensi dalam wawancara dengan pakar yaitu bagian Manajemen Rumah Sakit dan Rekam Medis, untuk mengetahui kesinambungan antara atribut penelitian yang telah ada dengan permasalahan yang pada rumah sakit, sehingga dapat memutuskan atribut yang digunakan dalam penelitian ini.



Gambar 2. Penentuan Atribut

2.2. Data Understanding

Pengumpulan data untuk proyek ini melibatkan dataset yang mencakup informasi terkait jumlah pasien masuk di rumah sakit. Hasil wawancara dengan pakar yaitu bagian Manajemen Rumah Sakit dan Rekam Medik. Data yang digunakan adalah data bangsal dari Januari tahun 2021 sampai dengan Desember 2022. Data tersebut diperoleh dalam format excel. Total data yang terkumpul selama dua tahun berjumlah 730 record. Data tersebut mencakup beberapa atribut antara lain tanggal, jumlah pasien awal, pasien masuk ruangan, pasien keluar ruangan, pasien yang masih dirawat, lama dirawat, pasien masuk/keluar pada hari yang sama, jumlah hari perawatan, serta detail mengenai kondisi pasien seperti pasien hidup, dirujuk ke RS lain, pulang paksa, dan lain-lain. Informasi ini memberikan gambaran lengkap tentang arus pasien di rumah sakit, termasuk durasi perawatan, kondisi keluar pasien dan perpindahan pasien antar-ruangan atau rumah sakit lain.

2.3. Data Pre-preparation

Proses tahap ini juga merupakan bagian dari metodologi CRISP-DM. tahap ini merupakan persiapan data sebagai input untuk algoritma tertentu. Terdapat tiga elemen untuk mendefinisikan kualitas data akurasi, lengkap

tidak lengkap (kompli) dan konsisten. Sehingga teknik pemrosesan data yang dipakai untuk mengatasi masalah data untuk mendapatkan data yang valid, nantinya dapat digunakan teknik data mining. Langkah berikutnya adalah melakukan pengecekan terhadap data yang memiliki nilai yang hilang (missing value). Selanjutnya data dibagi menjadi dua bagian, yaitu data training dan data testing. Data training dipakai untuk melatih model, sedangkan pada data testing digunakan dalam menguji performa model yang telah dibuat. Sehingga dapat menerapkan variabel dan model untuk mendapatkan hasil akurasi yang baik.

Linear Regression menghasilkan nilai MSE, MAE, dan RMSE semuanya 0, menandakan prediksi yang sempurna. Namun, ini bisa jadi indikasi adanya overfitting. Decision Tree memiliki performa yang lebih baik dibandingkan dengan Linear Regression. Hal ini ditunjukkan oleh nilai MSE, MAE, dan RMSE yang lebih rendah. Namun, performa Decision Tree masih belum optimal. Random Forest memiliki performa yang terbaik. Hal ini ditunjukkan oleh nilai MSE, MAE, dan RMSE yang paling rendah. Nilai MSE sebesar 0.64 berarti bahwa model Random Forest hanya membuat kesalahan rata-rata sebesar 6%. Nilai MAE sebesar 0.59 berarti bahwa model Random Forest hanya membuat kesalahan rata-rata sebesar 8. Hutan Acak berhasil mengurangi kompleksitas dan overfitting dibandingkan Decision Tree individual. Dapat disimpulkan bahwa model Random Forest memiliki performa yang terbaik dibandingkan dengan model Linear Regression dan Decision Tree.

2.4. Modelling

Konsep dari modeling ini merupakan algoritma klasifikasi model untuk memprediksi katagori dalam satu objek data. Pada penelitian ini menggunakan algoritma Random Forest. Dalam data mining terdapat evaluasi untuk mengetahui tingkat akurasi dari algoritma. Metode Random Forest melakukan pembagian data training dan data testing. Data training digunakan untuk membuat pola. Sedangkan data testing dipakai untuk mencoba pola dengan tujuan mengetahui tingkat akurasi dari algoritma Random Forest. Proses pembuatan model akan menggunakan algoritma Random Forest. Dataset akan diolah dengan tujuan untuk mengoptimalkan kinerja algoritma Random Forest dalam prediksi jumlah pasien masuk.

2.5. Evaluation

Setelah dilakukan algoritma random forest maka akan dievaluasi dan divalidasi, evaluasi model dan validasi prediksi jumlah pasien masuk dengan model machine learning. Beberapa metrik evaluasi model yang dapat digunakan antara lain Error Rate, Mean Squared Error (MSE) dan Mean Absolute Error (MAE).

1. Error rate adalah jumlah prediksi yang salah dibagi dengan total jumlah prediksi. Metrik ini memberikan gambaran tentang tingkat kesalahan model Random Forest dalam melakukan prediksi.
2. Mean Squared Error (MSE) dihitung sebagai akar kuadrat dari rata-rata dari selisih kuadrat antara nilai prediksi dan nilai aktual.
3. Mean Absolute Error (MAE) evaluasi regresi Ini mengukur rata-rata dari selisih absolut antara nilai prediksi dan nilai sebenarnya. MAE memberikan gambaran tentang seberapa dekat prediksi dengan nilai sebenarnya.

Untuk menilai kinerja model machine learning pada data yang belum pernah dilihat sebelumnya. Validasi model dilakukan untuk mengetahui seberapa baik model dapat generalisasi pada data baru. Ada beberapa metode validasi model yang dapat digunakan, contohnya dengan Cross-validation yaitu metode validasi model yang membagi data latih menjadi beberapa bagian.

2.6. Deployment

Pada tahap ini merupakan tahap akhir dalam pembentukan atau membangun model prediksi. Penelitian ini membangun sebuah model untuk prediksi jumlah pasien masuk di rumah sakit dengan tingkat akurasi terbaik. Sehingga model ini dapat memberi dan menjawab dari rumusan masalah dalam memprediksi jumlah pasien masuk.

3. HASIL DAN PEMBAHASAN

3.1 Business Understanding

Analisis Business Understanding menjadi landasan utama dalam menjalankan penelitian ini yang mana melibatkan pemahaman mendalam tentang konteks bisnis, tujuan proyek, serta masalah yang ingin dipecahkan. Tujuan dari proyek ini adalah untuk meramalkan atau memprediksi jumlah pasien masuk di rumah sakit. Dengan demikian, pemahaman akan kebutuhan ini dapat membantu manajemen rumah sakit dalam perencanaan sumber daya, pengelolaan kapasitas, dan penyediaan pelayanan yang lebih efisien. Kriteria keberhasilan proyek ini mungkin melibatkan tingkat akurasi prediksi yang tinggi. Kinerja model Random Forest harus dievaluasi berdasarkan seberapa baik dapat memprediksi jumlah pasien masuk. Selain itu, keberhasilan dapat diukur dari dampak positifnya terhadap efisiensi operasional dan pelayanan pasien. Stakeholder utama mungkin melibatkan

manajemen rumah sakit, staf medis, dan departemen administratif. Pemahaman akan kebutuhan dan harapan mereka terhadap prediksi jumlah pasien masuk, penting untuk mengarahkan analisis dan implementasi model dengan tepat.

3.2. Data Understanding

Pengumpulan data untuk proyek ini melibatkan dataset yang mencakup informasi terkait jumlah pasien masuk di rumah sakit. Hasil wawancara dengan pakar yaitu bagian Manajemen Rumah Sakit dan Rekam Medik. Data yang digunakan adalah data bangsal dari Januari tahun 2021 sampai dengan Desember 2022. Data tersebut diperoleh dalam format excel. Total data yang terkumpul selama dua tahun berjumlah 730 record.

Tgl	Pasien Awal	Jumlah Tempat Tidur	Jumlah Tempat Tidur Kosong	Pasien Masuk Ruangan			Pasien Keluar Ruangan					Pasien Masih Dirawat (4-13)	Lama Dirawat	Pasien Masuk/Keluar pd Hari yg Sama	Jumlah Hari Perawatan (14-21)	
				Pasien Masuk	Pasien Pindahan	Jumlah (1+2+3)	Pasien Dipindahkan	Pasien Keluar								Pasien Mati
								Hidup	Dirujuk Ke RS Lain	Pulang Paksa	Lain-Lain					
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	21	17
1	6	13	7	3	9	4					0	4	5	10		5
2	5	13	8		5	1	1				0	2	3	2		3
3	3	13	10		3						0	1	2	2		2
4	2	13	11	1	3	1					0	1	2	4		2
5	2	13	11	3	5	1					0	1	4	6		4
6	4	13	9	1	5	1					0	1	4	1		4
7	4	13	9		4	2					0	2	2	5		2
8	2	13	11	2	4	1					0	1	3	3		3
9	3	13	10		4	2					0	2	2	4		2
10	2	13	11	2	4	1					0	1	3	1		3
11	3	13	10	2	5						0	0	5			5
12	5	13	8	1	7	3					0	3	4	7		4
13	4	13	9	1	5	3					0	3	2	6		2
14	2	13	11	3	5	2					0	2	3	7		3
15	3	13	10	3	6	1					0	1	5	1		5
16	5	13	8	1	9	1					0	1	5	2		5
17	5	13	8	6	11	1					0	1	10	2		10
18	10	13	3	3	13	3					0	3	10	9		10
19	10	13	3	2	12	1					0	1	11	4		11
20	11	13	2	2	13	4					0	4	9	8		9
21	9	13	4	1	10	5	1				0	6	4	20		4
22	4	13	9	1	5	2					0	2	3	6		5
23	3	13	10	2	5	5					0	0	5	6		5
24	5	13	8	6	11	4					0	4	7	7		7
25	7	13	6	2	9	2					0	2	7	8		7
26	7	13	6	3	10	2					0	2	8	4		8
27	8	13	5	4	12	2					0	2	10	5		10
28	10	13	3	2	12	5					0	5	7	14		7

Gambar 3. Dataset Awal yang berhasil dikumpulkan

Atribut yang digunakan berjumlah 10 atribut yang terdiri dari Bulan, Tanggal, Jumlah Tempat Tidur, Jumlah Tempat Tidur Kosong, Pasien Awal, Pasien Masuk, Pasien Keluar Hidup, Pasien Keluar Mati, Pasien Masih Dirawat, Jumlah Hari Perawatan. Diperoleh atribut sebagai berikut:

Tabel 1. Atribut Keseluruhan

Nama Atribut	Deskripsi Atribut
Bulan	Menunjukkan bulan pada periode tertentu
Tanggal	Menunjukkan tanggal spesifik dalam bulan tertentu
Jumlah Tempat Tidur	Merupakan total kapasitas tempat tidur yang tersedia di suatu fasi
Jumlah Tempat Tidur Kosong	Mengindikasikan jumlah tempat tidur yang masih tersedia dan belum terisi pada suatu waktu
Pasien Awal	Merupakan jumlah pasien yang sudah berada di fasilitas perawatan pada awal periode yang diobservasi
Pasien Masuk	Menunjukkan jumlah pasien yang masuk ke fasilitas perawatan selama periode hari
Pasien Keluar Hidup	Merupakan jumlah pasien yang meninggalkan fasilitas perawatan dalam keadaan hidup selama periode tertentu
Pasien Keluar Mati	Menunjukkan jumlah pasien yang meninggal dunia dan meninggalkan fasilitas perawatan pada suatu periode waktu
Pasien Masih Dirawat	Merupakan jumlah pasien yang masih berada di fasilitas perawatan pada akhir periode waktu yang diobservasi
Jumlah Hari Perawatan	Mengindikasikan total hari perawatan yang diberikan kepada seluruh pasien selama periode tertentu

3.3. Data Preparation

Langkah-langkah pemrosesan data mencakup pembersihan data untuk menghilangkan data yang tidak lengkap atau outlier, penanganan missing values, dan normalisasi data jika diperlukan. Tujuannya adalah memastikan kualitas dan konsistensi data yang akan digunakan dalam analisis. Dengan melakukan pra-pemrosesan data secara cermat, dataset akan menjadi lebih bersih, konsisten, dan siap untuk digunakan dalam pemodelan

algoritma Decision Tree. Ini akan membantu meningkatkan akurasi dan performa model prediksi risiko hipertensi, sehingga hasil penelitian menjadi lebih dapat diandalkan dan relevan untuk pengelolaan kesehatan karyawan PT. Demitra Karsa Perdana, dalam data processing ada beberapa tahapan yaitu :

a. Penggabungan Tabel

Data yang digunakan adalah data bangsal yang diperoleh dari bagian sistem informasi terdiri dari 2 tabel yaitu data bangsal tahun 2021 dan 2022, sehingga didapatkan hasil sebagai berikut :

Tabel 2. Data Contoh Bangsal Tahun 2021-2022

Bulan	Tanggal	Jumlah Tempat Tidur	Jumlah Tempat Tidur Kosong	Pasien Awal	Pasien Masuk	Pasien Keluar Hidup	Pasien Keluar Mati	Pasien Masih Dirawat	Jumlah Hari Perawatan
Januari	1	18	18	0	1	0	0	1	1
Januari	2	18	17	1	0	0	0	1	1
Januari	3	18	17	1	0	0	0	1	1
Januari	4	18	17	1	3	1	0	3	3
Januari	5	18	15	3	3	2	0	4	4
Januari	6	18	14	4	4	2	0	6	6
Januari	7	18	12	6	1	2	0	5	5
Januari	8	18	13	5	2	1	0	6	6

b. Data Transformation

Data yang terkumpul tidak dapat langsung digunakan. Sebagai contoh pada atribut Bulan bertipe kategorikal kemudian ditransformasi menjadi numerik, Pasien Masuk menggabungkan dari atribut pasien masuk dan pasien pindahan, pasien keluar hidup dan pasien keluar matit. Hasil transformasi pada tabel dibawah ini:

a) Atribut Bulan

Tabel 3. Hasil Transformasi data pada atribut bulan

Nilai Atribut Bulan	Hasil Transformasi
Januari	1
Februari	2
Maret	3
April	4
Mei	5
Juni	6
Juli	7
Agustus	8
September	9
Oktober	10
November	11
Desember	12

b) Atribut Pasien Masuk

Tabel 4. Hasil Transformasi data pada atribut Pasien Masuk

Nama Atribut	Digabung Menjadi
Pasien Masuk	Pasien Masuk
Pasien Pindahan	

c) Atribut Pasien Keluar Hidup

Tabel 5. Hasil Transformasi data pada atribut Pasien Keluar Hidup

Nama Atribut	Digabung Menjadi
Pasien dipindahkan	Pasien Keluar Hidup
Hidup	
Dirujuk Ke RS lain	

Pulang paksa
Lain-lain

d) Atribut Pasien Keluar Mati

Tabel 6. Transformasi data pada atribut Pasien Keluar Mati

Nama Atribut	Digabung Menjadi
< 48 jam	Pasien Keluar Mati
> 48 jam	

Setelah dilakukan Transformasi Data maka didapatkan tampilan record dari dataset pada tabel berikut dibawah ini:

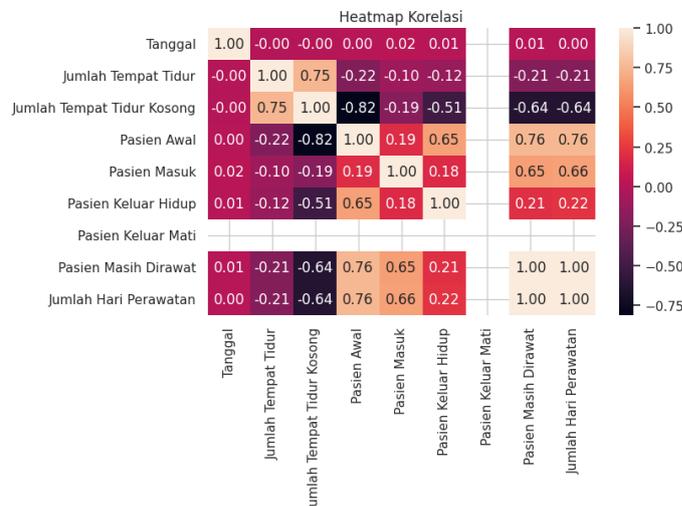
Tabel 7. Tampilan Dataset setelah Data Preprocessing

Bulan	Tanggal	Jumlah Tempat Tidur	Jumlah Tempat Tidur Kosong	Pasien Awal	Pasien Masuk	Pasien Keluar Hidup	Pasien Keluar Mati	Pasien Masih Dirawat	Jumlah Hari Perawatan
1	1	18	18	0	1	0	0	1	1
1	2	18	17	1	0	0	0	1	1
1	3	18	17	1	0	0	0	1	1
1	4	18	17	1	3	1	0	3	3
1	5	18	15	3	3	2	0	4	4
1	6	18	14	4	4	2	0	6	6
1	7	18	12	6	1	2	0	5	5
1	8	18	13	5	2	1	0	6	6
1	9	18	12	6	1	1	0	6	6
1	10	18	12	6	4	2	0	8	8
1	11	18	10	8	1	6	0	3	3
1	12	18	15	3	3	1	0	5	5
1	13	18	13	5	3	2	0	6	6
1	14	18	12	6	2	1	0	7	7
1	15	18	11	7	2	4	0	5	5

3.4. Modelling

1. Analisis Exploratif

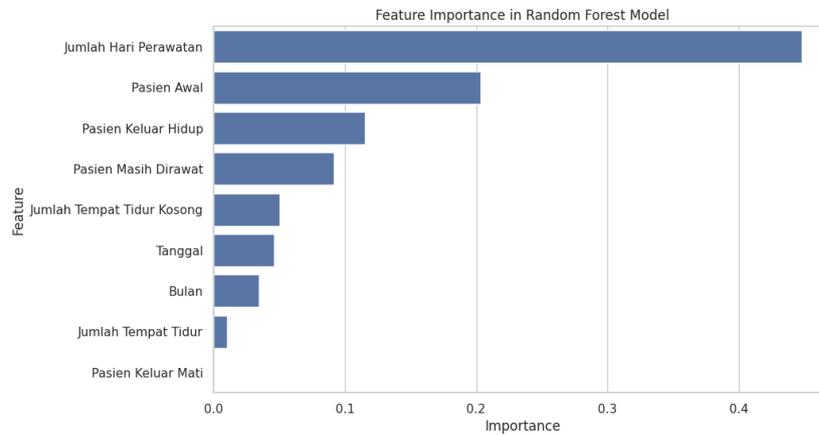
Heatmap ini menggambarkan korelasi antara jumlah pasien rumah sakit dengan berbagai faktor yang terkait dengan rawat inap pada fasilitas bangsal. Gradasi warna berkisar dari merah (korelasi positif) hingga biru (korelasi negatif), dengan putih menunjukkan korelasi tidak signifikan.



Gambar 4. Heatmap Korelasi

2. *Feature Importances In Random Forest Model*

Pentingnya fitur dalam model Random Forest untuk memprediksi jumlah pasien masuk di rumah sakit. Fitur Jumlah Hari Perawatan ini memiliki kepentingan tertinggi, menunjukkan bahwa jumlah hari pasien dirawat di rumah sakit adalah prediktor yang kuat untuk variabel target.

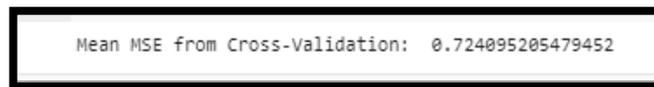


Gambar 5. *Feature Importances*

3. Pra pemrosesan data

Melatih model Random Forest dengan melakukan validasi silang, kinerja Random Forest untuk Memprediksi jumlah pasien masuk di Rumah Sakit menunjukkan pendekatan yang ketat untuk evaluasi model melalui validasi silang. Validasi silang melibatkan pembagian data menjadi beberapa fold, melatih model pada satu fold, dan menguji akurasi model pada fold lainnya. Proses iteratif ini memastikan penilaian yang kuat terhadap generalisasi model di luar data pelatihan.

Mean Squared Error (MSE) yang dilaporkan sebesar 0.724095205479452 menunjukkan kesalahan prediksi rata-rata sebesar magnitudo menunjukkan bahwa model Random Forest memiliki kinerja yang cukup baik dalam memprediksi nilai baru.



Gambar 6. Hasil MSE Validasi Silang

3.5. Evaluation

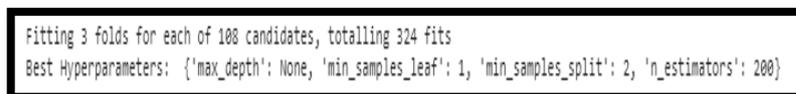
1. *Grid Search untuk Hyperparameter Tuning pada Model Random Forest*

Grid Search adalah metode yang digunakan untuk mencari kombinasi hyperparameter terbaik untuk suatu model. Dalam konteks ini, kita menggunakan Grid Search untuk melakukan tuning pada hyperparameter model Random Forest. Hyperparameter adalah parameter-parameter yang tidak diajarkan oleh model.

Tabel 8. *Grid Hyperparameter Yang Ingin Diuji*

Hyperparameter	Best Value
n_estimators	50, 100, 200
max_depth	None, 10, 20, 30
min_samples_split	2, 5, 10
min_samples_leaf	1, 2, 4

Maka didapat hasil *Best hyperparameter* sebagai berikut:



Gambar 7. *Best Hyperparameter*

2. Hasil Fitting Best Hyperparameter Dijadikan Melatih Model Random Forest

Dengan prediksi yang lebih akurat, rumah sakit dapat mengalokasikan sumber daya secara lebih efektif, mengurangi waktu tunggu pasien, dan mengoptimalkan penggunaan tempat tidur yang tersedia. Berdasarkan hasil evaluasi model, performa algoritma Random Forest menunjukkan keunggulan dibandingkan Decision Tree dalam hal akurasi prediksi. Hal ini terlihat dari nilai Mean Squared Error (MSE) sebesar 0.68, yang lebih rendah dibandingkan Decision Tree sebesar 1.17. Nilai Mean Absolute Error (MAE) juga lebih kecil, yaitu 0.59 untuk Random Forest dibandingkan dengan 0.72 pada Decision Tree. Selain itu, nilai Root Mean Squared Error (RMSE) Random Forest sebesar 0.82 juga lebih baik dibandingkan Decision Tree sebesar 1.08. Sementara itu, model Linear Regression menghasilkan nilai error mendekati nol, yang mengindikasikan performa yang tidak realistis, kemungkinan akibat overfitting atau ketidaksesuaian dengan kompleksitas data. Dari hasil ini, Random Forest menjadi pilihan terbaik untuk memprediksi jumlah pasien masuk rumah sakit, karena mampu menangkap hubungan kompleks antar variabel, menghasilkan prediksi yang andal, dan memberikan manfaat signifikan dalam perencanaan operasional rumah sakit.

Tabel 9. Hasil Pelatihan Model

	Linear Regression	Decision Tree	Random Forest
MSE	0.00	1.17	0.68
MAE	0.00	0.72	0.59
RMSE	0.00	1.08	0.82

3.6. Deployment

Adapun proses Deployment dilakukan setelah diketahui algoritma mana yang memiliki akurasi lebih baik pada penelitian ini, dapat disimpulkan Random Forest yang digunakan dalam prediksi jumlah pasien masuk, berikut tampilan prototyping berbasis web programming:



Gambar 8. Tampilan awal upload dataset

Analisis Eksploratif menggambarkan korelasi antara jumlah pasien rumah sakit dengan berbagai faktor yang terkait dengan rawat inap pada fasilitas bangsal dapat dilihat pada gambar berikut :

Analisis Eksploratif

	Bulan	Tanggal	Jumlah Tempat Tidur	Jumlah Tempat Tidur Kosong	Pasien Awal	Pasien Masuk	Pa
count	730	730	730	730	730	730	
mean	6.526	15.7205	15.0411	8.337	6.7041	2.8	
std	3.4502	8.8023	3.1728	5.3407	3.6519	2.0003	
min	1	1	8	-6	0	0	
25%	4	8	13	5	4	1	
50%	7	16	15	9	6	3	
75%	10	23	18	12	9	4	
max	12	31	18	18	18	10	

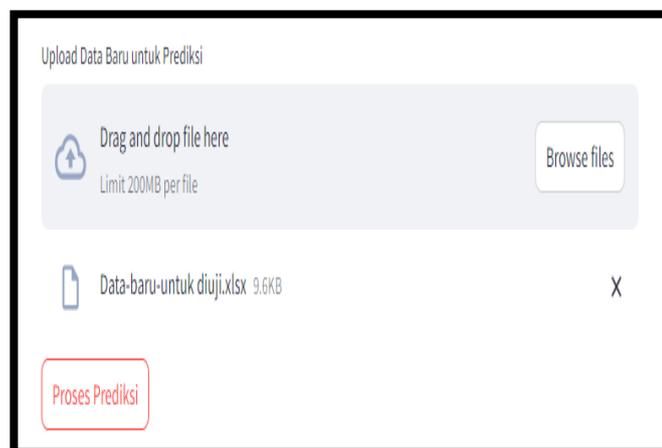
Gambar 9. Halaman Analisis Eksploratif

Memuat Model dan Prediksi menggambarkan tahapan menambahkan model *Random Forest* yang sudah disimpan.



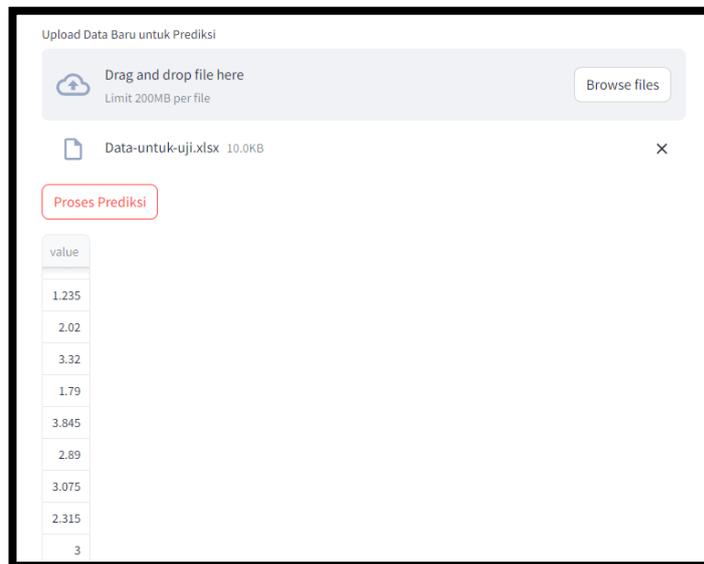
Gambar 10. Halaman Memuat Model dan Prediksi

Upload Data Baru menggambarkan tahapan tambah data baru yang diolah untuk diuji menggunakan model *Random Forest* yang sudah ditambahkan.



Gambar 11. Halaman Upload Data Baru

Proses Prediksi menggambarkan tahapan keluaran hasil prediksi dari model *Random Forest*.



Gambar 12. Halaman Proses Prediksi

4. KESIMPULAN

Berdasarkan permasalahan objek penelitian dan metodologi dalam Pemodelan Prediksi Jumlah Pasien Masuk Rumah Sakit Menggunakan Metode Random Forest, dapat disimpulkan bahwa hasil pengujian Feature Importances menunjukkan bahwa Jumlah Hari Perawatan memiliki kepentingan tertinggi, yang mengindikasikan bahwa durasi pasien dirawat di rumah sakit adalah prediktor yang paling kuat untuk variabel target, sementara hasil pelatihan model menunjukkan performa yang cukup baik dengan Mean Squared Error (MSE) sebesar 0,64, Mean Absolute Error (MAE) sebesar 0,60, dan Root Mean Squared Error (RMSE) sebesar 0,80, sehingga model ini direkomendasikan untuk diintegrasikan ke dalam sistem manajemen rumah sakit guna membantu perencanaan sumber daya dan pengambilan keputusan strategis, dengan pengembangan lanjutan yang dapat mencakup penambahan variabel eksternal, penerapan algoritma hybrid, pengolahan data real-time, analisis faktor penyebab lebih mendalam, dan uji coba pada berbagai rumah sakit untuk meningkatkan akurasi dan generalisasi model.

DAFTAR PUSTAKA

- [1] E. N. Rahmawati and M. A. Laras, "PREDIKSI KUNJUNGAN PASIEN RAWAT JALAN DI RSAU dr. SISWANTO LANUD ADI SOEMARMO KARANGANYAR TAHUN 2022-2026," *Infokes J. Ilm. Rekam Medis dan Inform. Kesehat.*, vol. 13, no. 1, pp. 52–63, 2023, doi: 10.47701/infokes.v13i1.2561.
- [2] 2019 Permenkes RI, No. 30, "PERATURAN MENTERI KESEHATAN REPUBLIK INDONESIA NOMOR 30 TAHUN 2019," *Prog. Retin. Eye Res.*, vol. 561, no. 3, pp. S2–S3, 2019.
- [3] F. Lawalata, E. Sedyono, and H. Purnomo, "Analisis Prediksi Jumlah Pasien Rawat Inap di Rumah Sakit GMIM Siloam Sonder Menggunakan Metode Triple Exponential Smoothing," *Jointer - J. Informatics Eng.*, vol. 2, no. 01, pp. 32–26, 2021, doi: 10.53682/jointer.v2i01.28.
- [4] N. 56 Permenkes RI, "Peraturan Menteri Kesehatan RI No 24 tahun 2022 Tentang Klasifikasi dan Perizinan Rumah Sakit," *Rev. CENIC. Ciencias Biológicas*, vol. 17, no. 3, pp. 1–26, 2014, [Online]. Available: file:///Users/andreataquez/Downloads/guia-plan-de-mejora-institucional.pdf%0Ahttp://salud.tabasco.gob.mx/content/revista%0Ahttp://www.revistaalad.com/pdfs/Guias_ALAD_11_Nov_2013.pdf%0Ahttp://dx.doi.org/10.15446/revfacmed.v66n3.60060.%0Ahttp://www.cenetec.
- [5] Syahbaniar, Desi, Rachmawati, Ervina, Erawantini, and Feby, "Perhitungan Perkiraan Kebutuhan Tempat Tidur Bangsal Kutilang Rsup Dr Kariadi Semarang," *J-REMI J. Rekam Med. dan Inf. Kesehat.*, vol. 2, no. 2, pp. 279–287, 2021, doi: 10.25047/j-remi.v2i2.2200.
- [6] L. Cheng, M. Tapia, K. Menzel, M. Page, and W. Ellis, "Predicting Need for Hospital Beds to Reduce Emergency Department Boarding," *Perm. J.*, vol. 26, no. 4, pp. 14–20, 2022, doi: 10.7812/TPP/21.211.
- [7] M. Tello *et al.*, "Machine learning based forecast for the prediction of inpatient bed demand," *BMC Med.*

-
- Inform. Decis. Mak.*, vol. 22, no. 1, pp. 1–14, 2022, doi: 10.1186/s12911-022-01787-9.
- [8] L. Zhou, P. Zhao, D. Wu, C. Cheng, and H. Huang, “Time series model for forecasting the number of new admission inpatients,” *BMC Med. Inform. Decis. Mak.*, vol. 18, no. 1, pp. 1–11, 2018, doi: 10.1186/s12911-018-0616-8.
- [9] F. W. REYNOLDS, “Early detection of diabetes mellitus,” *N. Y. State J. Med.*, vol. 59, no. 5, pp. 809–810, 2021.
- [10] W. M. Baihaqi, Dianingrum Melia, and Ramadhan Nuzul Aswin Kurnia, “Regresi Linier Sederhana Untuk Memprediksi Kunjungan,” *J. SIMETRIS*, vol. 10, no. 2, pp. 671–680, 2019.
- [11] W. Aser, Samosir Hizkia, and Gantini Tiur, “Analisis Dataset COVID-19 menggunakan Algoritma KNN dan Random Forest,” *J. Strateg. - J. Maranatha*, vol. 4, no. 1, pp. 58–69, 2022, [Online]. Available: <https://mail.strategi.it.maranatha.edu/index.php/strategi/article/view/325>
- [12] C. Wungkana, M. Aror, G. Arther Sandag, J. Arnold Mononutu, A. Bawah, and K. Airmadidi, “Model Prediksi Penderita HCC Menggunakan Algoritma Random Forest,” *J. Sist. Inf.*, vol. 10, no. 1, pp. 1–6, 2022, doi: 10.26418/justin.v10i1.44103.
- [13] Latha, Christalin Beulah, C., and J. C. S., “Improving the accuracy of prediction of heart disease risk based on ensemble classification techniques,” *Informatics Med. Unlocked*, vol. 16, no. November 2018, p. 100203, 2019, doi: 10.1016/j.imu.2019.100203.
- [14] A. R. Haristu, Rosa Prima, and H. Paulina, “Penerapan Metode Random Forest untuk Prediksi Win Ratio Pemain Player Unknown Battleground,” *MEANS (Media Inf. Anal. dan Sist.)*, vol. 4, no. 2, pp. 120–128, 2019, doi: 10.54367/means.v4i2.545.
- [15] M. Manic, “Data Mining,” *Ind. Electron. Handb. - Five Vol. Set*, 2011.