Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id

E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS

https://doi.org/10.35870/jikti.v3i2.2141

Pengembangan Model Prediksi Penyakit Menggunakan Algoritma Machine Learning Berbasis Data Kesehatan

Ade Hastuty [1], Rezky Deswita Ramadhani [2*], Wahyu Arthanugraha [3], Wahyuddin [4]

[1,2*,3,4] Program Studi Teknik Informatika, Fakultas Teknik, Universitas Muhammadiyah Parepare, Kota Parepare, Provinsi Sulawesi Selatan, Indonesia.

Corresponding Email: rezkyrezky270@gmail.com [2*]

Histori Artikel:
Dikirim 24 September 2026; *Diterima dalam bentuk revisi* 10 Oktober 2026; *Diterima* 10 November 2026; *Diterbitkan* 27 Januari 2027. Semua hak dilindungi oleh Lembaga Penelitian dan Pengabdian Masyarakat (LPPM) STMIK Indonesia Banda Aceh.

Abstrak

Penyakit tifus memerlukan penanganan dan identifikasi yang tepat. Identifikasi manual berdasarkan gejala dan data kesehatan memiliki keterbatasan, sehingga diperlukan model prediksi awal yang terstruktur. Penelitian ini bertujuan untuk membangun dan mengevaluasi sistem pendukung keputusan untuk diagnosis awal penyakit tifus berbasis web menggunakan algoritma Random Forest. Data yang digunakan terdiri atas 486 rekam medis sekunder dari Puskesmas Baranti, yang meliputi 383 sampel positif tifus dan 103 sampel negatif. Pembagian data dilakukan dengan metode Stratified Random Split dengan rasio 80:20, yaitu 388 sampel untuk data latih dan 98 sampel untuk data uji. Model dibangun menggunakan 500 decision trees dengan indeks Gini Impurity, kemudian diintegrasikan ke dalam aplikasi web menggunakan framework Flask dan basis data MySQL. Hasil pengujian menunjukkan akurasi sebesar 94,9%, presisi 95,0%, recall 98,7%, F1-score 96,8%, dan AUC 97,7%. Analisis Feature Importance menggunakan metode Mean Decrease Gini menunjukkan bahwa bintik merah dengan nilai kepentingan sebesar 35,5% dan mimisan dengan nilai 34,4% merupakan fitur yang digunakan dalam proses prediksi model. Dengan demikian, model Random Forest dapat diimplementasikan dalam aplikasi berbasis web untuk mendukung prediksi awal penyakit tifus berdasarkan gejala dan data kesehatan pengguna.

Kata Kunci: Penyakit Tifus; Website; Sistem Pendukung Keputusan; Machine Learning; Random Forest.

Abstract

Typhoid fever requires precise handling and identification. Manual identification based on symptoms and health data has limitations, thus necessitating a structured early prediction model. This study aims to build and evaluate a web-based decision support system for the early diagnosis of typhoid fever using the Random Forest algorithm. The data used consists of 486 secondary medical records from Baranti Health Center, including 383 positive typhoid samples and 103 negative samples. The data was divided using the Stratified Random Split method with an 80:20 ratio, allocating 388 samples for training data and 98 samples for testing data. The model was constructed using 500 decision trees with Gini Impurity index and subsequently integrated into a web application using the Flask framework and MySQL database. The testing results indicate an accuracy of 94.9%, precision of 95.0%, recall of 98.7%, F1-score of 96.8%, and AUC of 97.7%. Feature Importance analysis using the Mean Decrease Gini method reveals that the presence of red spots, with an importance value of 35.5%, and nosebleeds, with a value of 34.4%, are features utilized in the model's prediction process. Therefore, the Random Forest model can be implemented in a web-based application to support the early prediction of typhoid fever based on symptoms and user health data.

Keyword: Typhoid Fever; Website; Decision Support System; Machine Learning; Random Forest.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

1. Pendahuluan

Dalam era perkembangan teknologi yang semakin pesat, pemanfaatan teknologi informasi semakin luas, termasuk dalam bidang kesehatan. Salah satu teknologi yang berkembang adalah Machine Learning, yaitu metode yang mampu mengolah dan menganalisis data untuk menemukan pola tertentu (Hamilton, 2026). Dalam bidang kesehatan, teknologi ini dapat dimanfaatkan untuk membantu identifikasi dan prediksi penyakit berdasarkan data serta gejala pasien. Pemanfaatannya menjadi penting karena data kesehatan memerlukan pengolahan yang terstruktur dan efisien (Akbar *et al.*, 2025). Salah satu penyakit yang masih ditemukan di Indonesia adalah tifus atau demam tifoid, yang disebabkan oleh bakteri *Salmonella Typhi*. Penyakit ini dapat menimbulkan berbagai gejala yang muncul secara bersamaan, sehingga identifikasi berdasarkan gejala awal memerlukan ketelitian. Oleh karena itu, diperlukan pendekatan yang mampu mengolah berbagai karakteristik gejala untuk mendukung prediksi awal penyakit secara terstruktur (Imarah, 2021). Puskesmas Baranti merupakan salah satu fasilitas pelayanan kesehatan yang melayani masyarakat di wilayah Baranti (Asmila *et al.*, 2025). Data rekam medis pasien yang tersedia dapat dimanfaatkan sebagai sumber data untuk membangun model prediksi penyakit tifus. Penggunaan data dari fasilitas kesehatan tingkat pertama juga memungkinkan pengembangan model berdasarkan karakteristik data lokal yang tersedia (Ramadhan *et al.*, 2026). Salah satu pendekatan yang dapat diterapkan adalah pengembangan sistem prediksi berbasis web. Sistem ini memungkinkan pengguna untuk memasukkan data diri dan gejala yang dialami melalui antarmuka yang disediakan, kemudian memproses data tersebut untuk menghasilkan prediksi awal penyakit. Dalam penelitian ini, algoritma Machine Learning Random Forest digunakan untuk mengklasifikasikan data berdasarkan pola yang dipelajari dari rekam medis pasien (Ardin & Salam, 2025). Random Forest merupakan algoritma klasifikasi yang menggabungkan sejumlah decision tree untuk menghasilkan prediksi berdasarkan hasil pembelajaran dari data. Algoritma ini digunakan untuk mengolah fitur gejala pasien dan menentukan hasil klasifikasi berdasarkan pola yang terbentuk selama proses pelatihan (Riansah *et al.*, 2025). Selain menghasilkan prediksi, model juga dapat dianalisis melalui feature importance untuk mengidentifikasi fitur yang berkontribusi terhadap hasil klasifikasi. MySQL merupakan sistem manajemen basis data yang digunakan untuk menyimpan dan mengelola data secara terstruktur. MySQL digunakan sebagai basis data untuk mendukung pengelolaan data pada aplikasi, yang berkaitan dengan proses sistem. Selain MySQL, pengembangan model juga melibatkan bahasa pemrograman Python sebagai bagian dari proses pengolahan data dan pembentukan model Machine Learning (Yunita *et al.*, 2024). Python merupakan bahasa pemrograman yang berfungsi untuk melakukan pengolahan data, membangun model prediksi, serta menerapkan algoritma (Suryati *et al.*, 2025). Proses pembentukan model menggunakan pustaka Machine Learning Scikit-learn, sedangkan model yang telah terbentuk disimpan dalam bentuk berkas agar dapat digunakan kembali oleh backend aplikasi tanpa melakukan pelatihan ulang. Selanjutnya, model tersebut dihubungkan dengan sistem berbasis web untuk memproses input gejala dan menghasilkan prediksi penyakit secara real-time.

Penelitian terdahulu yang dilakukan oleh Alfarisi *et al.* (2025) berjudul "Pengembangan Model Machine Learning untuk Deteksi Penyakit Diabetes Menggunakan Analisis Gini Importance" bertujuan untuk mengembangkan model klasifikasi Machine Learning untuk mendeteksi penyakit diabetes secara dini dengan memanfaatkan analisis kontribusi fitur. Metode yang digunakan melibatkan lima algoritma klasifikasi, yaitu Random Forest, Support Vector Machine (SVM), Logistic Regression, Linear Discriminant Analysis (LDA), dan Artificial Neural Network (ANN). Dataset yang digunakan berasal dari Kaggle dengan jumlah 100.000 data. Penelitian tersebut melakukan tahapan preprocessing, hybrid resampling, feature importance menggunakan Gini Importance, serta pembagian data 80% untuk pelatihan dan 20% untuk pengujian dengan tambahan 10-fold cross validation. Hasil penelitian menunjukkan bahwa Random Forest memperoleh akurasi tertinggi sebesar 87%, dengan nilai recall 79%, precision 82%, F1-score 81%, dan AUC sebesar 0,85. Analisis Gini Importance juga menunjukkan bahwa blood_glucose, bmi, dan age merupakan tiga fitur yang paling berkontribusi terhadap prediksi diabetes. Namun, penelitian tersebut masih berfokus pada deteksi penyakit diabetes menggunakan dataset sekunder

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

dari Kaggle dan belum secara khusus membahas prediksi penyakit tifus berdasarkan data rekam medis dari fasilitas kesehatan tingkat pertama. Selain itu, penelitian tersebut lebih menitikberatkan pada perbandingan performa beberapa algoritma dan seleksi fitur menggunakan Gini Importance, sedangkan penerapan model dalam bentuk sistem berbasis web untuk input gejala pasien secara langsung bukan menjadi fokus utama penelitian. Sementara itu, penelitian yang saya lakukan berfokus pada pengembangan model prediksi penyakit tifus menggunakan algoritma Random Forest dengan memanfaatkan data kesehatan dari Puskesmas Baranti. Model yang dikembangkan tidak hanya digunakan untuk proses klasifikasi, tetapi juga diintegrasikan ke dalam aplikasi berbasis web sehingga pengguna dapat memasukkan data diri dan gejala yang dialami untuk memperoleh hasil prediksi awal. Dengan demikian, penelitian ini memiliki perbedaan pada jenis penyakit, sumber data, karakteristik fitur, serta bentuk penerapan model, yaitu pengembangan sistem prediksi tifus berbasis web yang memanfaatkan data kesehatan lokal dari Puskesmas. Kondisi tersebut menunjukkan bahwa penelitian ini penting dilakukan karena diperlukan suatu pendekatan yang dapat memanfaatkan data kesehatan yang tersedia untuk membantu proses prediksi awal penyakit tifus. Pemanfaatan data rekam medis dari Puskesmas Baranti memberikan gambaran yang lebih sesuai dengan karakteristik data kesehatan pada lokasi penelitian. Selain itu, penerapan algoritma Random Forest ke dalam aplikasi berbasis web diharapkan dapat memberikan hasil prediksi secara lebih praktis berdasarkan data dan gejala yang dimasukkan pengguna. Berdasarkan uraian tersebut, penelitian ini bertujuan untuk mengembangkan model prediksi penyakit tifus menggunakan algoritma Machine Learning Random Forest berbasis data kesehatan dari Puskesmas Baranti. Model yang dikembangkan selanjutnya diterapkan pada sistem berbasis web agar data gejala yang dimasukkan pengguna dapat diproses secara otomatis dan menghasilkan prediksi awal.

2. Metode Penelitian

Penelitian ini menggunakan metode pengembangan model prediksi penyakit yang dilaksanakan di Puskesmas Baranti, Sulawesi Selatan. Penelitian ini memanfaatkan perangkat keras dan perangkat lunak yang mendukung melalui tahapan studi literatur, desain, perancangan dan pembuatan aplikasi, pengujian, serta penulisan laporan.

2.1 Tahapan Penelitian

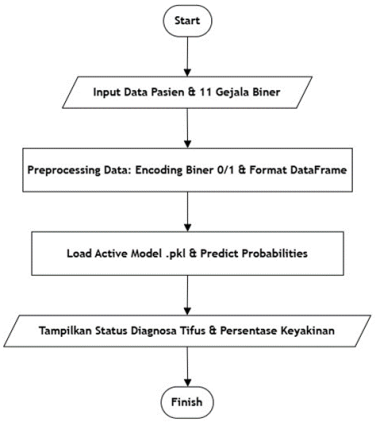
Dalam penelitian ini, tahapan yang dilakukan meliputi studi literatur, di mana referensi yang berkaitan dengan penyakit tifus, Machine Learning, algoritma Random Forest, dan pengembangan aplikasi berbasis web dipelajari sebagai dasar penelitian. Selanjutnya, data rekam medis pasien dikumpulkan dari Puskesmas Baranti, dengan seleksi data yang sesuai dengan kebutuhan penelitian. Proses preprocessing data dilakukan untuk memeriksa dan mengolah data, termasuk menangani data yang tidak sesuai, data kosong, dan duplikat sebelum digunakan dalam pemodelan. Setelah itu, model prediksi penyakit tifus dibangun menggunakan algoritma Random Forest berdasarkan data yang telah diproses. Model tersebut kemudian diintegrasikan ke dalam aplikasi berbasis web menggunakan Python, Flask, dan MySQL, sehingga pengguna dapat memasukkan data dan gejala untuk memperoleh hasil prediksi. Terakhir, pengujian dan analisis dilakukan untuk mengevaluasi kinerja model menggunakan metrik seperti akurasi, presisi, recall, F1-score, dan AUC, serta untuk menguji fungsi aplikasi agar sistem berjalan sesuai dengan rancangan.

2.2 Rancangan Sistem

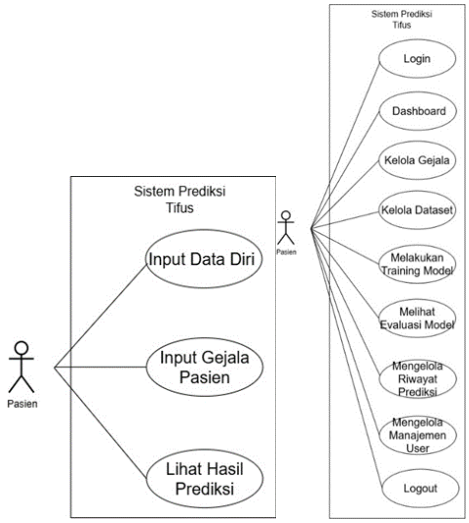
Alur prediksi Tifus dimulai dari input data diri dan 11 gejala biner oleh pasien, kemudian data divalidasi dan diproses oleh backend Flask dalam format Pandas *DataFrame*. Selanjutnya, *model Random Forest* (.pkl) digunakan untuk menghasilkan prediksi dan probabilitas, lalu hasilnya disimpan ke basis data dan ditampilkan pada halaman laporan diagnosa. *Flowchart* ini menggambarkan *Main*

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

Success Scenario (Happy Path), dengan validasi input dilakukan melalui *HTML5/JavaScript* sebelum data diproses.



Gambar 1. *Flowchart* Penelitian



Gambar 2. *Use Case* Diagram Pasien dan admin

Tabel 1. Deskripsi *Usecase* Pasien dan Admin

No	Use Case	Aktor	Deskripsi
1.	Input Data Diri Pasien	Pasien	Pasien menginput data diri terlebih dahulu
2.	Input Gejala Pasien	Pasien	Pasien menginput keluhan atau gejala awal yang dirasakan untuk memulai proses pelayanan medis.
3.	Lihat Hasil Prediksi	Pasien	Pengguna dapat melihat tingkat keyakinan prediksi penyakit
4.	Login	Admin	Proses autentikasi pengguna untuk mengakses sistem dengan memasukkan username dan *password* yang valid sesuai hak aksesnya.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

5.	Dashboard	Admin	Halaman utama yang menampilkan ringkasan informasi, mengenai dataset serta riwayat prediksi
6.	Kelola Gejala		Admin mengelola gejala seperti menambah, menghapus serta mengedit gejala.
7.	Kelola Dataset	Admin	Admin mengelola dataset seperti upload dan proses dataset, mengosongkan dataset, sert export dataset.
8.	Melakukan *Training Model*	Admin	Melakukan *training model Random Forest*
9.	Melihat *Evaluasi Model*	Admin	Admin melihat detail dari *evaluasi training model,* melihat model yang aktif serta melihat riwayat model yang telah di *training*
10.	Mengelola riwayat prediksi	Admin	Admin dapat melihat riwayat pasien yang telah melakukan prediksi serta dapat mengelolanya
11.	Mengelola Manajemen User	Admin	Admin dapat melihat dan menghapus daftar pengguna sistem

Tabel 2. Rancangan Output

No	Jenis Output	Deskripsi	Bentuk Output Akhir
1	Halaman input data pasien	Fitur untuk memasukkan data diri dan gejala awal yang dialami pasien.	Form input pada aplikasi web
2	Halaman prediksi penyakit	Menampilkan hasil prediksi risiko Tifus berdasarkan model Random Forest.	Tampilan hasil prediksi (text output)
3	Integrasi model machine learning	Aplikasi memanggil file model .pkl/.joblib untuk melakukan prediksi tanpa perlu training ulang.	Endpoint pemanggilan model di server
4	Laporan hasil prediksi	Informasi prediksi yang dapat ditampilkan atau disimpan sementara.	Output berupa teks hasil prediksi
5	Antarmuka Pengguna (UI)	Desain tampilan yang sederhana, mudah digunakan, dan responsif.	Prototype UI dalam bentuk halaman web

Rancangan output aplikasi ini menunjukkan komponen utama yang dibangun untuk mendukung proses prediksi penyakit secara otomatis. Setiap bagian aplikasi dirancang agar mampu mengolah data input pasien, menjalankan model prediksi yang telah dilatih, serta menyajikan hasil prediksi secara jelas dan mudah dipahami oleh pengguna.

2.3 Teknik Pengumpulan Data

Data penelitian terdiri atas 486 rekam medis pasien, yang mencakup 383 sampel berlabel positif tifus dan 103 sampel berlabel negatif tifus. Data ini diperoleh melalui dokumentasi rekam medis di Puskesmas Baranti pada periode Mei hingga Agustus 2026. Pemilihan data dilakukan berdasarkan kriteria inklusi dan eksklusi. Kriteria inklusi mencakup rekam medis yang memiliki label diagnosis serta informasi variabel penelitian yang relevan untuk pembentukan model. Sementara itu, kriteria eksklusi mencakup rekam medis duplikat, data dengan label diagnosis yang tidak jelas, dan data yang tidak memenuhi kebutuhan atribut penelitian. Penerapan kriteria tersebut bertujuan untuk memastikan bahwa data yang digunakan sesuai dengan kebutuhan pemodelan. Selain itu, identitas pasien dijaga kerahasiaannya selama proses pengolahan data.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

Metode pengumpulan data dalam penelitian ini menggunakan beberapa teknik. Pertama, dokumentasi dilakukan dengan mengumpulkan data dari rekam medis Puskesmas, yang mencakup gejala, hasil diagnosis dokter, dan jenis penyakit. Data ini digunakan sebagai dataset untuk melatih, menguji, dan mengevaluasi model Machine Learning, dengan tetap memperhatikan izin, etika, dan kerahasiaan data pasien. Kedua, wawancara terstruktur atau semi-terstruktur dilakukan untuk memperoleh informasi mengenai proses dokter dalam menilai gejala dan menentukan diagnosis. Hasil wawancara ini digunakan untuk menentukan variabel klinis yang relevan dalam model prediksi. Ketiga, studi pustaka dilakukan dengan mengkaji jurnal ilmiah, buku, dan dokumen terkait Machine Learning, khususnya metode Random Forest, serta penyakit tifus sebagai landasan teori penelitian. Terakhir, pengujian sistem dilakukan melalui Black Box Testing untuk memastikan bahwa setiap fitur dan fungsi aplikasi berjalan sesuai dengan rancangan dan kebutuhan pengguna, di mana pengujian ini dilakukan dengan menguji input dan output sistem tanpa memeriksa struktur atau kode program di dalamnya.

2.4 Variabel Penelitian

Fitur yang digunakan untuk mengidentifikasi gejala penyakit tifus terdiri atas sebelas gejala, yaitu bintik merah, mimisan, lidah kotor, nyeri ulu hati, mual, sakit kepala, lemas, muntah, batuk, demam, dan sakit perut. Kesebelas gejala ini berfungsi sebagai indikator dalam proses identifikasi awal untuk mengetahui kemungkinan seseorang mengalami penyakit tifus berdasarkan gejala yang dialami.

2.5 Teknik Analisis Data

Teknik analisis data pada penelitian ini menggunakan pendekatan Machine Learning dengan algoritma Random Forest untuk membangun model prediksi penyakit berdasarkan atribut kesehatan pasien. Analisis dimulai dengan tahap preprocessing, yang meliputi pembersihan data, penanganan data kosong dan duplikasi, serta transformasi data agar sesuai dengan kebutuhan model. Selanjutnya, dilakukan feature selection untuk memilih atribut yang relevan. Dataset kemudian dibagi menjadi data latih dan data uji, di mana data latih digunakan untuk membangun model Random Forest, sedangkan data uji digunakan untuk mengevaluasi kinerja model. Model Random Forest beroperasi dengan membentuk beberapa decision tree dan menggabungkan hasil prediksi melalui mekanisme majority voting. Kinerja model dievaluasi menggunakan metrik seperti precision, recall, dan F1-score. Model yang telah memenuhi hasil evaluasi kemudian diimplementasikan ke dalam sistem untuk mendukung prediksi awal penyakit tifus.

2.6 Recall

Recall, atau Sensitivity, merupakan metrik evaluasi yang digunakan untuk mengukur kemampuan model dalam mendeteksi data positif secara benar. Dalam konteks penelitian prediksi penyakit, Recall menunjukkan kemampuan model dalam mengenali pasien yang benar-benar mengalami tifus. Semakin tinggi nilai Recall, semakin baik kemampuan model dalam mendeteksi kasus positif dan mengurangi jumlah kasus yang terlewat (Nababan *et al.*, 2025). Rumus untuk menghitung Recall adalah:

$$\text{Recall}=\text{TPTP+FNRecall}=\mathit{TP}+\mathit{FNTP}$$

Keterangan:

- TP (True Positive): jumlah pasien sakit yang diprediksi dengan benar sebagai sakit.
- FN (False Negative): jumlah pasien sakit yang diprediksi sebagai sehat.

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

Interpretasi dari nilai Recall berada pada rentang 0 hingga 1. Nilai yang mendekati 1 menunjukkan kemampuan model yang baik dalam mendeteksi kasus positif, sementara nilai yang rendah menunjukkan masih banyak kasus positif yang tidak terdeteksi (false negative).

2.7 F1-Score

F1-Score merupakan metrik evaluasi yang digunakan untuk mengukur keseimbangan antara Precision dan Recall. Metrik ini sangat berguna ketika terdapat ketidakseimbangan jumlah data (imbalanced data), karena mempertimbangkan ketepatan prediksi serta kemampuan model dalam menemukan seluruh data positif (Kurnia Phan, 2026). Rumus untuk menghitung F1-Score adalah:

$$\text{F1-Score}=2\times(\text{Precision}\times\text{Recall})\text{Precision}+\text{RecallF1-Score}=\text{Precision}+\text{Recall}2\times(\text{Precision}\times\text{Recall})$$

Dengan definisi Precision sebagai:

$$\text{Precision}=\text{TPTP}+\text{FPPrecision}=TP+FPTP$$

Keterangan:

- TP (True Positive): jumlah data positif yang diprediksi dengan benar sebagai positif.
- FP (False Positive): jumlah data negatif yang diprediksi salah sebagai positif.
- Recall: kemampuan model dalam mendeteksi seluruh data positif.

Interpretasi dari nilai F1-Score berada pada rentang 0 hingga 1. Nilai yang mendekati 1 menunjukkan keseimbangan yang baik antara Precision dan Recall, sementara nilai yang rendah menunjukkan bahwa Precision, Recall, atau keduanya masih belum optimal.

3. Hasil dan Pembahasan

3.1 Hasil

3.1.1 Perhitungan Sistem

Berdasarkan metodologi yang telah dirancang, proses implementasi dan pengujian sistem pendukung keputusan diagnosis awal penyakit Tifus (Demam Tifoid) menggunakan algoritma Random Forest berbasis web di diuraikan secara bertahap dalam sub-bab berikut:

Tabel 3. Karakteristik Distribusi Sebaran Kelas Target

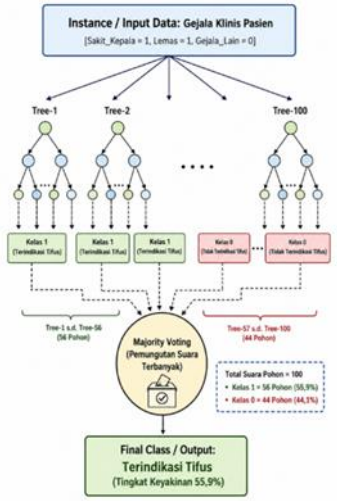
No	Kelas Target (Diagnosa_Biner)	Interpretasi Diagnosis Klinis	Jumlah Sampel (Pasien)	Persentase(%)
1.	1	Terindikasi Penyakit Tifus (Positif)	383	78,81%
2.	0	Tidak Terindikasi Tifus (Negatif)	103	21,19%
Total Keseluruhan Dataset			486	100,00%

Untuk menjamin validitas evaluasi model, sebanyak 486 data dibagi menggunakan teknik *Stratified Random Split* dengan rasio 80% data latih dan 20% data uji. Teknik stratifikasi digunakan untuk mempertahankan proporsi kelas positif dan negatif pada kedua subset data. Rincian pembagian data disajikan pada tabel berikut:

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

Tabel 4. Hasil Pembagian Data (Data Splitting)

Subset Data	Proporsi Rasio	Jumlah Kelas Positif (1)	Jumlah Kelas Negatif (0)	Total Baris Data
Data Latih (Training Data)	80%	306	82	388
Data Uji (Testing Data)	20%	77	21	98
Total Keseluruhan		383	103	486



Gambar 4. Arsitektur Pembentukan Model Random Forest Prediksi Tifus

Cara kerja model dalam memproses data gejala klinis pasien hingga menghasilkan keputusan akhir dimulai dengan memasukkan sebelas fitur gejala biner (0/1) ke dalam model sesuai dengan urutan fitur pada data pelatihan. Selanjutnya, algoritma Random Forest yang terdiri atas 500 decision trees memproses data secara paralel dengan melakukan percabangan berdasarkan kombinasi fitur dan kriteria Gini Impurity yang telah dipelajari pada tahap pelatihan. Setiap pohon menghasilkan prediksi berupa kelas 1 (Terindikasi Tifus) atau kelas 0 (Tidak Terindikasi Tifus). Hasil prediksi dari seluruh pohon kemudian dikumpulkan menggunakan mekanisme majority voting, di mana kelas dengan jumlah suara terbanyak ditetapkan sebagai hasil prediksi akhir yang ditampilkan pada antarmuka aplikasi. Tingkat keyakinan dihitung berdasarkan proporsi jumlah pohon yang memilih kelas prediksi terhadap seluruh pohon.

Tabel 5. Hasil Pengujian *Confusion Matrix* Model *Random Forest*

Matriks Kekacauan (N = 98)		Kelas Prediksi Sistem	
		Terindikasi Tifus (1)	Tidak Terindikasi Tifus (0)
Kelas Aktual	Aktual Tifus (1)	True Positive (TP): 76	False Negative (FN): 1
	Aktual Tidak Tifus (0)	False Positive (FP): 4	True Negative (TN): 17

Hasil klasifikasi pada data uji sebanyak 98 data menunjukkan bahwa terdapat 76 True Positive (TP), 17 True Negative (TN), 4 False Positive (FP), dan 1 False Negative (FN). Model berhasil mengidentifikasi 76 pasien yang terindikasi tifus dan 17 pasien yang tidak terindikasi tifus secara tepat. Sementara itu, terdapat 4 data False Positive, yaitu pasien yang diprediksi terindikasi tifus tetapi kondisi aktualnya negatif, serta 1 data False Negative, yaitu pasien yang sebenarnya terindikasi tifus tetapi diprediksi negatif oleh model. Jumlah tersebut sesuai dengan keseluruhan data uji, yang dapat dirumuskan sebagai berikut:

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

TP+FN+FP+TN=76+1+4+17=98*TP*+*FN*+*FP*+*TN*=76+1+4+17=98

Dengan demikian, total data pada confusion matrix sesuai dengan jumlah 98 data uji, yang merupakan 20% dari 486 data keseluruhan.

1) Akurasi (*Accuracy*)
Akurasi menunjukkan proporsi prediksi yang benar terhadap seluruh data uji.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \times 100\%$$
$$= \frac{76 + 17}{76 + 17 + 4 + 1} \times 100\% = \frac{93}{98} \times 100\% = \mathbf{94,9\%}$$

2) *Recall* / Sensitivitas
Recall mengukur kemampuan model dalam mendeteksi pasien yang benar-benar menderita Tifus.

$$Recall = \frac{TP}{TP + FN} \times 100\%$$
$$= \frac{76}{76 + 1} \times 100\% = \mathbf{98,7\%}$$

3) Presisi (*Precision*)
Precision menunjukkan ketepatan model dalam memprediksi pasien sebagai Tifus.

$$Precision = \frac{TP}{TP + FP} \times 100\%$$
$$= \frac{76}{76 + 4} \times 100\% = \mathbf{95,0\%}$$

4) *F1-Score*
F1-Score merupakan metrik yang menggambarkan keseimbangan antara *Precision* dan *Recall*.

$$F1\text{-}Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$
$$= 2 \times \frac{0,95 \times 0,987}{0,95 + 0,987} = \mathbf{96,8\%}$$

5) *Specificity*
Specificity menunjukkan kemampuan model dalam mengenali pasien yang sebenarnya tidak menderita Tifus.

$$Specificity = \frac{TN}{TN + FP} \times 100\%$$

Tabel 6 Nilai Feature Importance Gejala Penyakit Tifus

No	Atribut Gejala Klinis Pasien	Nilai Bobot Kepentingan (Feature Importance)	Peringkat Kontribusi
1	Bintik Merah	0,355 (35,5%)	1 (Tertinggi)
2	Mimisan	0,344 (34,4%)	2
3	Lidah Kotor	0,078 (7,9%)	3
4	Nyeri Ulu Hati	0,056 (5,6%)	4
5	Mual	0,044 (4,4%)	5
6	Sakit Kepala	0,044 (4,4%)	6

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

7	Lemas	0,039 (3,9%)	7
8	Muntah	0,028 (2,8%)	8
9	Batuk	0,010 (1,0%)	9
10	Demam	0,001 (0,1%)	10
11	Sakit Perut	0,000 (0,0%)	11 (Terendah)

Berdasarkan tabel di atas, analisis *Feature Importance* menggunakan metode *Mean Decrease Gini* menunjukkan kontribusi masing-masing dari 11 gejala klinis terhadap keputusan model dalam memprediksi Tifus. Gejala dengan kontribusi terbesar adalah Bintik Merah sebesar 35,5% dan Mimisan sebesar 34,4%, yang secara keseluruhan menyumbang hampir 70% dari nilai kepentingan fitur. Selanjutnya, kontribusi fitur terdiri atas Lidah Kotor 7,9%, Nyeri Ulu Hati 5,6%, Mual 4,4%, Sakit Kepala 4,4%, Lemas 3,9%, Muntah 2,8%, Batuk 1,0%, Demam 0,1%, dan Sakit Perut 0,0%. Nilai *Feature Importance* tersebut menunjukkan seberapa besar kontribusi relatif setiap gejala dalam proses pembentukan keputusan pohon pada model *Random Forest*, bukan menunjukkan bahwa suatu gejala secara klinis lebih penting atau menjadi penyebab Tifus. Secara komputasi, nilai kepentingan fitur diperoleh berdasarkan perubahan ketidakmurnian (*impurity*) pada node pohon menggunakan kriteria Gini Impurity

Tabel 7 Matriks Data Latih 11 Gejala

Gejala	Kelompok Ya	Positif Tifus/ Non-Tifus	Kelompok Tidak	Positif tifus/ Non-Tifus	Total Data
Bintik Merah	31	0/31	357	306/51	388
Mimisan	35	0/35	353	306/47	388
Lidah Kotor	280	250/30	108	56/52	388
Nyeri Ulu Hati	260	230/30	128	76/52	388
Mual	295	250/45	93	56/37	388
Sakit Kepala	310	260/50	78	46/32	388
Lemas	320	265/55	68	41/27	388
Muntah	240	200/40	148	106/42	388
Batuk	190	155/35	198	151/47	388
Demam	370	295/75	18	11/7	388
Sakit Perut	368	306/82	0	0/0	388

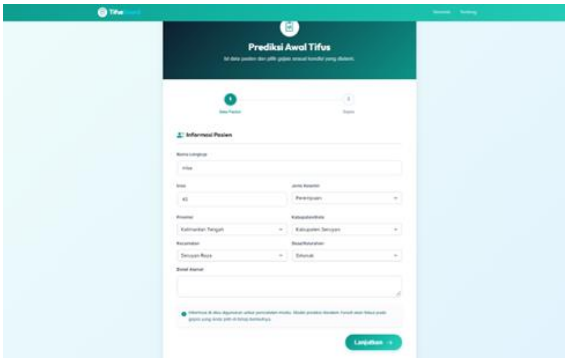


Gambar 5 Halaman Utama

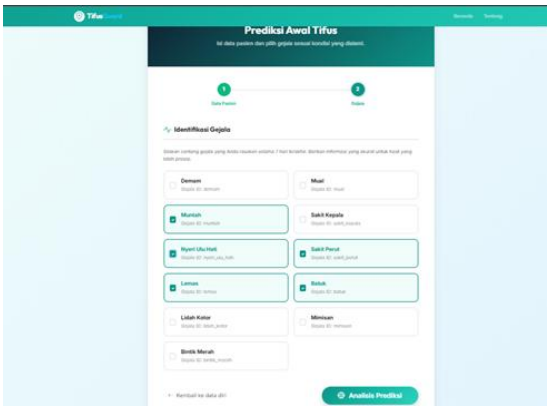
Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id

E - ISSN: 3047-6674 | P - ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS

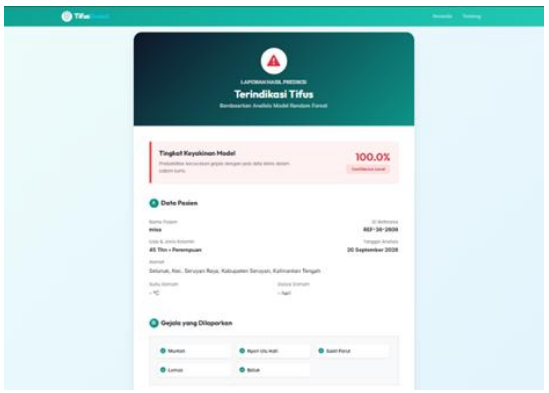
https://doi.org/10.35870/jikti.v3i2.2141



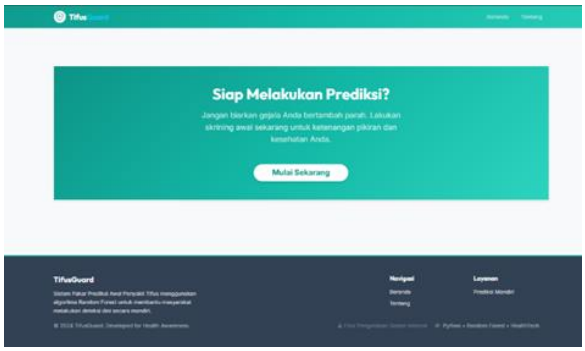
Gambar 6 Formulir Input Data Diri



Gambar 7 Formulir Input Gejala Pasien



Gambar 8 Laporan Hasil Prediksi

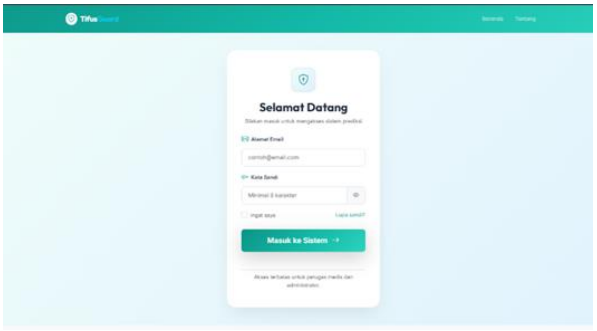


Gambar 9 Akses Administrator Internal

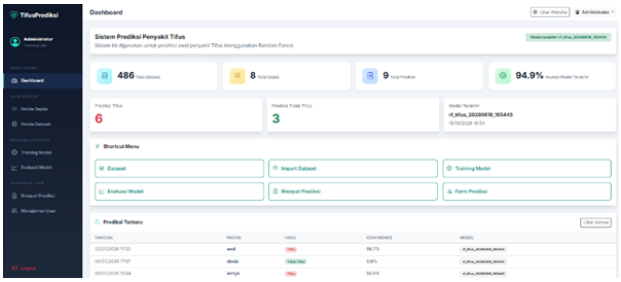
Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id

E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS

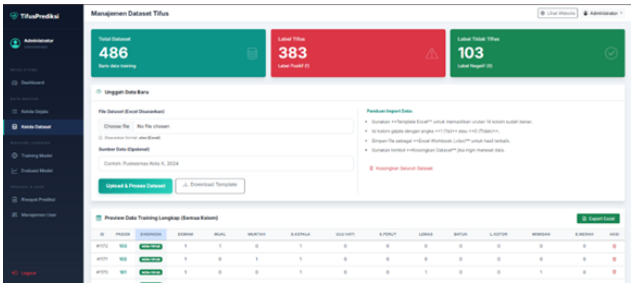
https://doi.org/10.35870/jikti.v3i2.2141



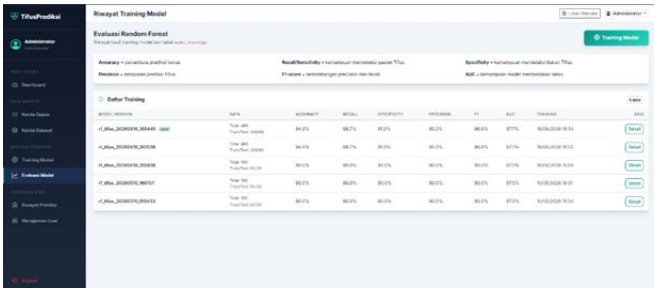
Gambar 10 Halaman Login Administrator



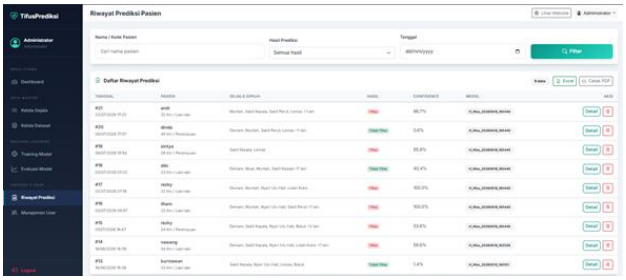
Gambar 11 Dashboard Administrator



Gambar 12 Kelola Dataset

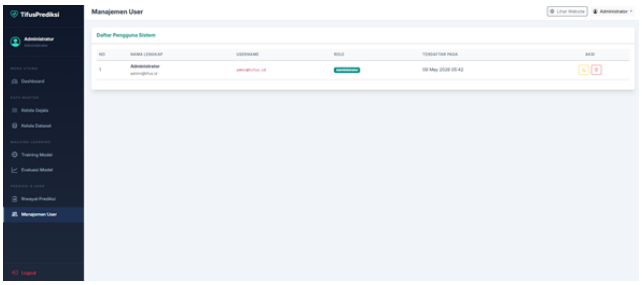


Gambar 13 Evaluasi Model



Gambar 14 Riwayat Prediksi

Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141



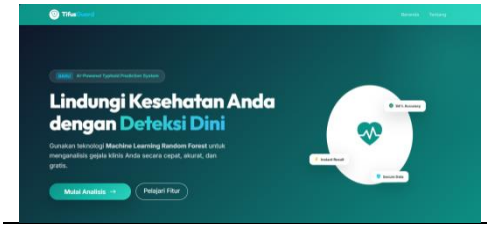
Gambar 15 Manajemen User

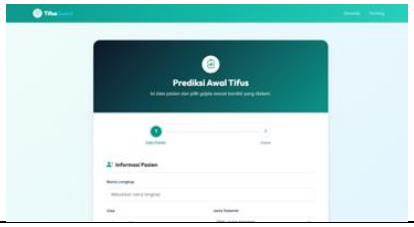
3.1.2 Pengujian Sistem BlackBox

Black box testing adalah metode pengujian perangkat lunak yang berfokus sepenuhnya pada fungsionalitas sistem tanpa memedulikan struktur kode internal aplikasi. Penguji bertindak layaknya pengguna akhir dengan memberikan *input* tertentu ke dalam sistem, lalu memeriksa apakah *output* yang dihasilkan sudah sesuai dengan dokumen spesifikasi atau belum (Permatasari *et al.*, 2023).

Tabel 8 BlackBox Halaman Utama

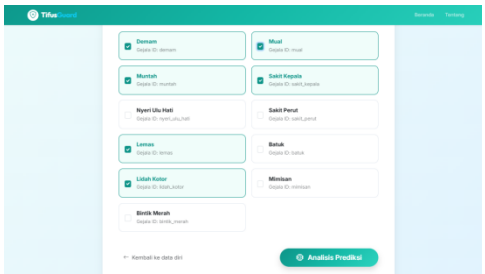
Tes Factor	Hasil	Keterangan
Mengakses URL website dan menekan tombol "Mulai Analisis"	Berhasil	Berhasil, sistem menampilkan antarmuka utama dan mengarahkan pengguna ke formulir input data diri
Screenshot		

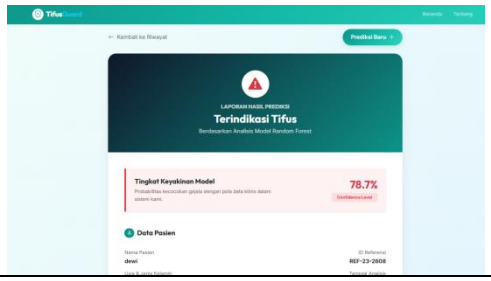




Tabel 9 BlackBox Halaman Identifikasi Gejala

Tes Factor	Hasil	Keterangan
Mencentang gejala klinis dan menekan tombol "Analisis Prediksi"	Berhasil	Berhasil, sistem mengirimkan data indikator gejala ke model *Random Forest* dan menampilkan hasil klasifikasi
Screenshot		





Tabel 10 BlackBox Halaman Login Untuk Administrator

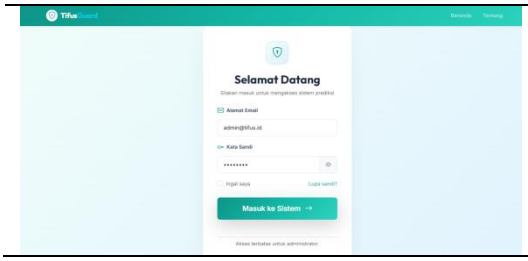
Tes Factor	Hasil	Keterangan
Menginput email dan kata sandi yang benar	Berhasil	Berhasil, sistem dapat memvalidasi email dan menampilkan halaman *dashboard*
Screenshot		

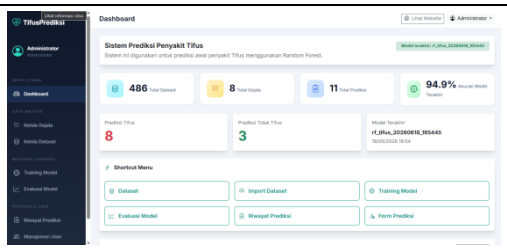
Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id

E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari

OPEN ACCESS

https://doi.org/10.35870/jikti.v3i2.2141

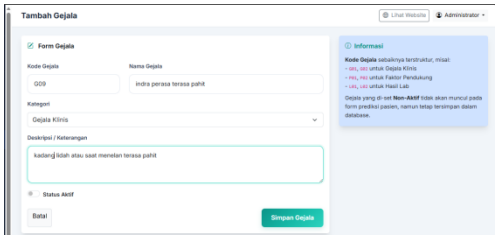


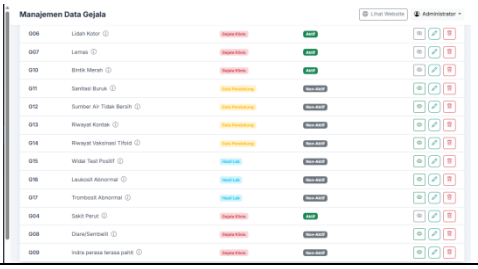


Tabel 11 BlackBox Tambah Gejala

Tes Factor	Hasil	Keterangan
Tambah Gejala	Berhasil	Data Gejala berhasil ditambahkan dan tampil pada manajemen data gejala

Screenshot

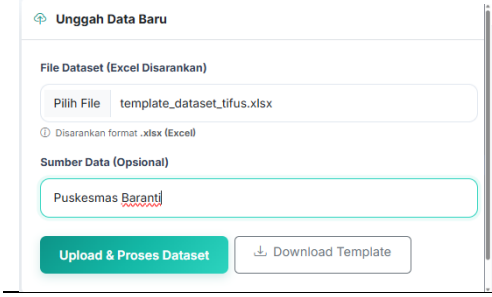




Tabel 12 BlackBox Unggah Dataset

Tes Factor	Hasil	Keterangan
Unggah dataset	Berhasil	Dataset berhasil di unggah dan tampil pada preview data training

Screenshot

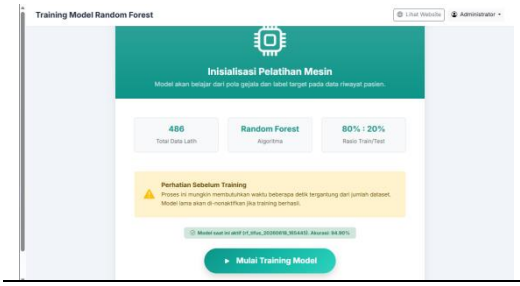




Tabel 13 BlackBox Training Model Random Forest

Tes Factor	Hasil	Keterangan
Training Model Random Forest	Berhasil	Training Random Forest berhasil dan hasilnya bisa dilihat di detail evaluasi model

Screenshot

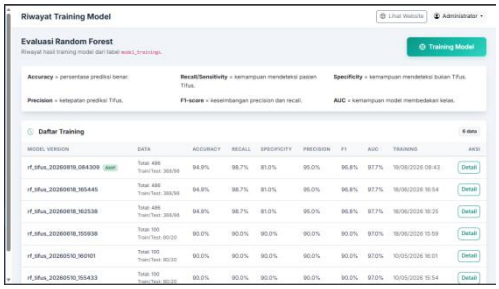




Jurnal Ilmu Komputer dan Teknologi Informasi
https://journal.stmiki.ac.id
E - ISSN: 3047-6674 | P - ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

Tabel 14 BlackBox Riwayat Training Model Random Forest

Tes Factor	Hasil	Keterangan
Riwayat *Training Model*	Berhasil	Riwayat *training model* berhasil dilihat detailnya
Screenshot		





3.2 Pembahasan

Hasil klasifikasi model menunjukkan bahwa algoritma Random Forest berhasil mengidentifikasi gejala tifus dengan akurasi yang cukup tinggi, di mana 76 dari 98 data uji terklasifikasi sebagai True Positive. Hasil ini mencerminkan kemampuan model dalam mendeteksi pasien yang terindikasi tifus secara tepat, dengan tingkat False Positive yang relatif rendah (4 kasus) dan hanya 1 kasus False Negative. Hal ini sejalan dengan penelitian oleh Alfarisi *et al.* (2025), yang menemukan bahwa penggunaan algoritma Random Forest dalam deteksi penyakit diabetes juga menunjukkan kinerja yang baik, dengan akurasi mencapai 87%. Penelitian tersebut menekankan pentingnya pemilihan fitur yang relevan dan preprocessing data yang tepat untuk meningkatkan kinerja model. Selain itu, penelitian oleh Kurnia Phan (2026) yang mengeksplorasi penerapan teknik Machine Learning pada diagnosis penyakit infeksi juga menunjukkan bahwa model berbasis Random Forest dapat secara signifikan meningkatkan deteksi dini penyakit, mirip dengan hasil yang diperoleh dalam penelitian ini. Dengan demikian, temuan ini menegaskan potensi algoritma Random Forest sebagai alat yang efektif dalam mendukung diagnosis awal penyakit tifus, serta pentingnya pengembangan lebih lanjut dalam sistem pendukung keputusan berbasis teknologi untuk meningkatkan layanan kesehatan.

4. Kesimpulan dan Saran

Berdasarkan hasil pengujian yang telah dilakukan, dapat disimpulkan bahwa aplikasi model prediksi awal penyakit tifus menggunakan algoritma Machine Learning telah berhasil dikembangkan. Sistem prediksi penyakit tifus berbasis web ini dibangun dan diuji menggunakan 486 data rekam medis pasien dari Puskesmas. Kinerja model menunjukkan performa yang baik pada data uji, dengan tingkat akurasi mencapai 94,9%, presisi 95,0%, F1-Score 96,8%, dan nilai AUC 97,7%. Analisis Feature Importance mengidentifikasi dua gejala klinis yang paling berpengaruh dalam prediksi awal penyakit tifus, yaitu variabel bintik merah dengan kontribusi sebesar 35,5% dan mimisan sebesar 34,4%. Selain itu, validasi sistem melalui pengujian Black-Box menunjukkan bahwa seluruh skenario pengujian yang dilakukan berhasil sesuai dengan keluaran yang diharapkan. Temuan ini menegaskan efektivitas model dalam mendukung diagnosis awal penyakit tifus dan potensi penerapan teknologi Machine Learning dalam bidang kesehatan. Penelitian selanjutnya disarankan menggunakan jumlah data yang lebih banyak dan beragam untuk meningkatkan keandalan model Random Forest. Pengembangan sistem juga dapat dilakukan dengan menambahkan fitur yang relevan serta membandingkan kinerja beberapa algoritma klasifikasi.

E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

5. Daftar Pustaka

Akbar, I., Supriadi, F., & Junaedi, D. I. (2025). Pemanfaatan machine learning di bidang kesehatan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(1), 1744-1749. https://doi.org/10.36040/jati.v9i1.12663.

Alfarisi, M. A., Sumanto, B., & Rachmawan, I. P. F. (2025). Pengembangan Model Machine Learning untuk Deteksi Penyakit Diabetes Menggunakan Analisis Gini Importance. *Journal of Internet and Software Engineering*, *6*(2), 127-134. https://doi.org/10.22146/jise.v6i2.16551.

Hamilton, K. (2026). ANALISIS PERKEMBANGAN ARTIFICIAL INTELLIGENCE DALAM DIAGNOSIS PENYAKIT SERTA DAMPAKNYA TERHADAP PELAYANAN KESEHATAN DI ERA MODERN. *Jurnal Riset Sistem Informasi*, *3*(3), 83-93. https://doi.org/10.69714/1vr6d731.

Imara, F. (2020, August). Salmonella typhi bakteri penyebab demam tifoid. In *Prosiding Seminar Nasional Biologi* (Vol. 6, No. 1, pp. 1-5). https://doi.org/10.24252/psb.v6i1.14264.

Permatasari, I., Adhania, F., Putri, S. A., & Nursari, S. R. C. (2023). Pengujian Black Box Menggunakan Metode Analisis Nilai Batas pada Aplikasi DANA. *KONSTELASI: Konvergensi Teknologi dan Sistem Informasi*, *3*(2), 373-387. https://doi.org/10.24002/konstelasi.v3i2.8289.

Ramadhan, A. A., Hartono, R., Mualim, I., & Subarwanti, Y. (2026). Sistem Informasi Tingkat Kekeruhan Air Layak Guna Menggunakan Sensor Turbidity dan Wemos D1. *Jurnal Kajian dan Penelitian Umum*, *4*(1), 10-26. https://doi.org/10.47861/jkpu-nalanda.v4i1.2088.

Riansah, A., Nurdiawan, O., & Herdiana, R. (2025). Penerapan Algoritma Random Forest dan Decision Tree untuk Meningkatkan Akurasi Klasifikasi Penjualan pada Toko Bangunan. *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(3), 4242-4249. https://doi.org/10.36040/jati.v9i3.13622.

Siregar, Y. S., Sembiring, B. O., Rahayu, E., Hasdiana, H., & Franchitika, R. (2024). Pemanfaatan aplikasi MySQL untuk membantu siswa SMK Swasta Nur Azizi dalam pengolahan data. *Jurnal Pengabdian Masyarakat*, *3*(2), 229-240. https://doi.org/10.70340/japamas.v3i2.185.

Suryati, M., Aldi, T., Saputra, D., & Ikasari, I. H. (2025). Penggunaan Python dalam analisis data dengan machine learning. *JRIIN: Jurnal Riset Informatika dan Inovasi*, *2*(10), 1959-1968.

Trenggono, P. H., & Bachtiar, A. (2023). Peran artificial intelligence dalam pelayanan kesehatan: A systematic review. *Jurnal Ners*, *7*(1), 444-451.

Apriliah, W., Kurniawan, I. B., Baydhowi, M., & Haryati, T. (2021). Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest. Sistemasi, 10(1), 163. https://doi.org/10.32520/stmsi.v10i1.1129.

Desiani, A., Irmeilyana, I., & Amran, A. (2022). Implementasi Algoritma Naïve Bayes dan Support Vector Machine (SVM) Pada Klasifikasi Penyakit Kardiovaskular. Jurnal Teknik Elektro Dan Komputasi (Elkom), 4(2), 207–214. https://doi.org/10.32528/elkom.v4i2.7691.

Erlin, E., Marlim, Y. N., Junadhi, J., Suryati, L., & Agustina, N. (2022). Deteksi Dini Penyakit Diabetes Menggunakan Machine Learning dengan Algoritma Logistic Regression. Jurnal

E-ISSN: 3047-6674 | P-ISSN: 3047-6682
Vol. 4 No. 1 (2027) | Januari
OPEN ACCESS
https://doi.org/10.35870/jikti.v3i2.2141

Nasional Teknik Elektro Dan Teknologi Informasi (Jnteti), 11(2), 88–96. https://doi.org/10.22146/jnteti.v11i2.3586.

Febrianti, A., & Hidayat, H. (2024). Aplikasi DIATECT Untuk Prediksi Penyakit Diabetes Menggunakan SVM Berbasis Web. Jurnal Tekno Kompak, 18(1), 150. https://doi.org/10.33365/jtk.v18i1.3643.

Masruriyah, A., Novita, H. Y., Sukmawati, C., Ramadhan, A., Arif, S., & Dermawan, B. A. (2024). Pengukuran Kinerja Model Klasifikasi dengan Data Oversampling pada Algoritma Supervised Learning untuk Penyakit Jantung. Computer Science (Co-Science), 4(1), 62–70. https://doi.org/10.31294/coscience.v4i1.2389.

Prabowo, A. S., & Kurniadi, F. I. (2023). Analisis Perbandingan Kinerja Algoritma Klasifikasi dalam Mendeteksi Penyakit Jantung. Jurnal Siskom-Kb (Sistem Komputer Dan Kecerdasan Buatan), 7(1), 56–61. https://doi.org/10.47970/siskom-kb.v7i1.468.

Purwono, P., Dewi, P., Wibisono, S. K., & Dewa, B. P. (2022). Model Prediksi Otomatis Jenis Penyakit Hipertensi dengan Pemanfaatan Algoritma Machine Learning Artificial Neural Network. Insect (Informatics and Security) Jurnal Teknik Informatika, 7(2), 82–90. https://doi.org/10.33506/insect.v7i2.1828.

Septiani, W. D. (2020). OPTIMASI ALGORITMA C4.5 MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI PENYAKIT HEPATITIS. Inti Nusa Mandiri, 15(1), 59–64. https://doi.org/10.33480/inti.v15i1.1421.

Utami, D., Dwiatmoko, F., & Novsagita, A. (2026). Komparasi Algoritma Machine Learning dengan Hyperparameter Tuning pada Prediksi Level Obesitas. *Modem: Jurnal Informatika dan Sains Teknologi.*, *4*(3), 93-105. https://doi.org/10.62951/modem.v4i3.955.

Vira, S. (2021). Pengelompokkan Provinsi di Indonesia berdasarkan Penyakit Tidak Menular Menggunakan Metode Partisi, Hierarki, dan Fuzzy Clustering. Jurnal Msa (Matematika Dan Statistika Serta Aplikasinya), 9(1). https://doi.org/10.24252/msa.v9i1.17942.