**TIN: Terapan Informatika Nusantara**
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

# Perbandingan Kinerja Algoritma SVM dan Random Forest dalam Klasifikasi Tuberkulosis di Puskesmas Sandar Angin

**Rinda Desfourtheen, Lastri Widya Astuti, Muhammad Haviz Irfani***

Fakultas Ilmu Komputer dan Sains, Program Studi Teknik Informatika, Universitas Indo Global Mandiri, Palembang, Indonesia
Email: [1]2022110132@students.uigm.ac.id, [2]lastriwidya@uigm.ac.id, [3,*]m.haviz@uigm.ac.id
Email Penulis Korespondensi: m.haviz@uigm.ac.id

**Abstrak–**Tuberkulosis (TBC) merupakan penyakit menular yang masih menjadi permasalahan kesehatan di Indonesia sehingga diperlukan pendekatan yang mampu mendukung proses diagnosis secara akurat. Pemanfaatan *machine learning* dapat membantu klasifikasi berdasarkan karakteristik klinis pasien, tetapi perbedaan algoritma dapat menghasilkan kinerja yang berbeda. Penelitian ini bertujuan membandingkan kinerja *Support Vector Machine* (SVM) dan *Random Forest* serta menentukan algoritma dengan kinerja terbaik dalam klasifikasi TBC menggunakan data pasien Puskesmas Sandar Angin. Penelitian menggunakan data primer 201 pasien periode Januari 2023 hingga Oktober 2025 di Puskesmas Sandar Angin dengan 14 variabel, yaitu no, jenis kelamin, umur, nyeri dada, batuk berdahak, lama batuk, sesak napas, kelelahan, penurunan berat badan, suhu tubuh, keringat dingin, nafsu makan berkurang, sputum, dan diagnosis. Evaluasi model dilakukan melalui skema pembagian data dengan rasio 80:20 serta metode *5-Fold Cross Validation*. Pada skenario pembagian data 80:20, algoritma SVM menghasilkan *accuracy* sebesar 98%, *precision* 94%, *recall* 100%, dan *F1-score* 97%, sedangkan *Random Forest* memperoleh *accuracy* 93%, *precision* 88%, *recall* 94%, dan *F1-score* 91%. Berdasarkan hasil validasi *5-Fold Cross Validation*, SVM mencatat *accuracy* tertinggi sebesar 90%, sementara *Random Forest* memperoleh *accuracy* tertinggi sebesar 88%. Analisis hasil pengujian memperlihatkan bahwa SVM menunjukkan keunggulan dalam proses klasifikasi jika dibandingkan dengan *Random Forest*, sehingga dipilih sebagai algoritma yang memberikan kinerja terbaik dalam klasifikasi penyakit TBC pada data pasien di Puskesmas Sandar Angin.

**Kata Kunci:** Tuberkulosis; *Support Vector Machine*; *Random Forest*; Klasifikasi*; Machine Learning*

**Abstract–**Tuberculosis (TBC) remains a major infectious disease and public health problem in Indonesia, requiring approaches that can support accurate diagnosis. Machine learning can assist TBC classification based on patients' clinical characteristics, but different algorithms may produce different performance. This study aims to compare the performance of Support Vector Machine (SVM) and Random Forest and determine the best-performing algorithm for TBC classification using patient data from Sandar Angin Public Health Center, Pagar Alam. The study used primary data from 201 patients recorded from January 2023 to October 2025, consisting of 14 variables: number, gender, age, chest pain, productive cough, cough duration, shortness of breath, fatigue, weight loss, body temperature, cold sweats, decreased appetite, sputum, and diagnosis. Model evaluation was conducted using an 80:20 data-splitting scheme and 5-Fold Cross Validation. In the 80:20 split, SVM achieved 98% accuracy, 94% precision, 100% recall, and 97% F1-score, while Random Forest achieved 93% accuracy, 88% precision, 94% recall, and 91% F1-score. Based on 5-Fold Cross Validation, SVM achieved the highest accuracy of 90%, while Random Forest achieved 88%. Overall, SVM demonstrated superior classification performance and was selected as the best-performing algorithm for TBC classification using patient data from Sandar Angin Public Health Center, Pagar Alam.

**Keywords**: Tuberculosis; Support Vector Machine; *Random Forest*; Classification; Machine Learning

## 1. PENDAHULUAN

Laporan World Health Organization (2024) menyebutkan bahwa jumlah kasus tuberkulosis (TBC) secara global pada tahun 2023 diperkirakan mencapai 10,8 juta orang. Delapan negara tercatat menyumbang lebih dari dua pertiga dari total kasus global, dan Indonesia menjadi salah satunya dengan kontribusi sekitar 10% dari seluruh kasus dunia. Di tingkat nasional, menurut (Kementerian Kesehatan Republik Indonesia, 2024), Indonesia mencatat sekitar 885.000 kasus TBC sepanjang tahun 2024. Kasus tersebut meliputi 496.000 penderita laki-laki, 359.000 penderita perempuan, dan 135.000 penderita anak pada kelompok usia 0–14 tahun. Di tingkat fasilitas Kesehatan, seperti Pusat Kesehatan Masyarakat (Puskesmas) Sandar Angin yang berlokasi di Jl. Persirah Yohan, Kecamatan Dempo Utara Kota Pagar Alam, hasil observasi dan wawancara menunjukkan bahwa penyakit yang saat ini mengalami peningkatan dan banyak diderita oleh masyarakat yang datang berobat ke Puskesmas Sandar Angin adalah TBC.

Seseorang yang tinggal serumah dengan penderita TBC lebih rentan terinfeksi karena penyakit ini sangat mudah menular. Ini terjadi ketika seseorang yang terinfeksi tidak sengaja menghirup percikan ludah atau *droplet* ketika mereka bersin atau batuk (Wahyudi & Setyawan, 2022). Keberadaan gejala klinis, seperti batuk berdahak yang berlangsung berkepanjangan, batuk berdarah, demam, penurunan berat badan, serta gejala lainnya, menjadi indikator penting dalam proses diagnosis. Namun, proses diagnosis sering kali menghadapi tantangan karena memerlukan waktu yang lama dan data yang kompleks (Zahra et al., 2025). Oleh karena itu, pemanfaatan teknik *machine learning* memberikan peluang untuk membantu proses klasifikasi penyakit TBC sehingga dapat melakukan diagnosis penyakit TBC secara lebih dini.

*Support Vector Machine* (SVM) dan *Random Forest* adalah dua algoritma *machine learning* yang telah banyak diimplementasikan pada berbagai penelitian di bidang kesehatan, khususnya untuk klasifikasi penyakit. SVM dikenal mampu mengolah data berdimensi tinggi dengan membentuk *hyperplane* yang memberikan pemisahan antarkelas secara optimal. Karakteristik tersebut memungkinkan SVM menghasilkan tingkat akurasi yang baik, termasuk pada dataset yang berukuran relatif keci (Mutmainah et al., 2024). Selain itu, SVM juga mampu menangani data *non-linear* melalui penggunaan fungsi kernel (Du et al., 2024). Sementara itu, *Random Forest* menerapkan pendekatan *ensemble* dengan membentuk beberapa *decision tree*, kemudian mengombinasikan prediksi yang dihasilkan untuk memperoleh

**TIN: Terapan Informatika Nusantara**
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

model klasifikasi yang lebih stabil dan tidak mudah mengalami *overfitting* (M. W. Nugroho, 2025). SVM dan *Random Forest* sama-sama algoritma klasifikasi *supervised learning* yang sering digunakan pada data medis dan memiliki performa yang baik, namun keduanya memiliki mekanisme kerja yang berbeda (Ihsan, 2025).

Berbagai penelitian sebelumnya telah mengimplementasikan SVM dan *Random Forest* untuk menyelesaikan permasalahan klasifikasi maupun prediksi pada bidang kesehatan. Dalam penelitian Manurung et al. (2025) yang membandingkan kinerja SVM dan *Random Forest* dalam memprediksi penyakit hipertensi. Hasil penelitian menunjukkan bahwa *Random Forest* memperoleh *accuracy* sebesar 98,92%, sedangkan SVM memperoleh *accuracy* sebesar 83,91%. Meskipun penelitian menunjukkan bahwa algoritma *Random Forest* memiliki kinerja lebih baik dalam memprediksi penyakit hipertensi, penelitian tersebut masih belum membandingkan algoritma pada *dataset* pasien TBC. Sementara itu, penelitian yang dilakukan oleh Triloka & Sugianto (2025) menggunakan algoritma SVM dan *Random Forest* untuk memprediksi hasil pengobatan TBC di Indonesia. Evaluasi model menghasilkan nilai *accuracy* 99,61% untuk SVM, sementara *Random Forest* mencapai *accuracy* sebesar 81,80%. Meskipun SVM menunjukkan kinerja lebih baik, penelitian tersebut berfokus pada prediksi luaran pengobatan, yaitu pasien sembuh atau meninggal, dengan variabel yang mencakup karakteristik pasien, gejala klinis, pemeriksaan, terapi, kepatuhan pengobatan, riwayat terapi, status gizi, dan status kesembuhan. Penggunaan variabel tersebut berkaitan dengan kondisi dan proses pengobatan pasien TBC, sementara penggunaan variabel klinis berupa gejala pasien dan hasil sputum untuk klasifikasi diagnosis TBC di Puskesmas Sandar Angin masih belum menjadi fokus penelitian.

Studi yang dilakukan oleh Aini et al. (2026) mengevaluasi performa *Random Forest*, SVM, *XGBoost*, dan *K-Nearest Neighbor* (KNN) untuk klasifikasi penyakit jantung berdasarkan data klinis dari 1.888 pasien. Hasil penelitian menunjukkan bahwa *Random Forest* memperoleh *accuracy* tertinggi sebesar 97,88%, diikuti *XGBoost* sebesar 94,71%, SVM sebesar 92,59%, dan KNN sebesar 90,21%. Meskipun *Random Forest* menunjukkan kinerja lebih baik dibandingkan SVM pada klasifikasi penyakit jantung, penelitian tersebut masih menggunakan *dataset* penyakit jantung dan belum menerapkan kedua algoritma tersebut pada *dataset* pasien TBC. Selain itu, penelitian yang dilakukan oleh Azhari et al. (2021) mengevaluasi kinerja algoritma C4.5, *Random Forest*, SVM, dan *Naive Bayes* pada klasifikasi kelayakan peserta Jogja International Scout Camp (JISC) menggunakan 200 data. Berdasarkan hasil evaluasi, SVM memperoleh nilai *accuracy* tertinggi sebesar 95,00%, diikuti *Naive Bayes* dan C4.5 yang masing-masing mencapai 86,67%, sedangkan *Random Forest* memperoleh *accuracy* sebesar 83,33%. Meskipun penelitian menunjukkan bahwa algoritma SVM memiliki kinerja lebih baik dalam klasifikasi kelayakan peserta, penelitian tersebut diterapkan pada permasalahan seleksi peserta kegiatan dan bukan pada klasifikasi penyakit TBC.

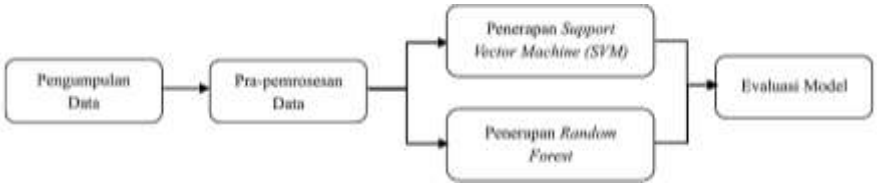
Berbagai penelitian terdahulu menunjukkan bahwa SVM dan *Random Forest* memiliki kinerja yang berbeda bergantung pada objek, karakteristik data, dan tujuan klasifikasi yang digunakan. Namun, penelitian sebelumnya yang membandingkan kedua algoritma tersebut masih berfokus pada objek dan permasalahan yang berbeda, seperti klasifikasi hipertensi, penyakit jantung, prediksi luaran pengobatan TBC, serta klasifikasi kelayakan peserta kegiatan. Kondisi tersebut menunjukkan bahwa penerapan kedua algoritma untuk mengklasifikasikan status diagnosis TBC berdasarkan data klinis pasien masih belum banyak dikaji. Selain itu, penerapan SVM dan *Random Forest* pada fasilitas pelayanan kesehatan tingkat pertama, khususnya Puskesmas Sandar Angin Pagar Alam, belum menjadi fokus penelitian terdahulu. Perbedaan objek, tujuan klasifikasi, dan karakteristik data tersebut menunjukkan adanya *research gap* yang mendasari penelitian ini. Oleh karena itu, penelitian ini bertujuan untuk membandingkan kinerja algoritma SVM dan *Random Forest* dalam mengklasifikasikan diagnosis TBC berdasarkan data klinis pasien di Puskesmas Sandar Angin Pagar Alam serta menentukan algoritma yang memberikan kinerja terbaik.

## 2. METODOLOGI PENELITIAN

### 2.1 Tahapan Penelitian

Penelitian ini menerapkan pendekatan kuantitatif dengan melakukan analisis komparatif terhadap algoritma SVM dan *Random Forest* pada klasifikasi penyakit TBC menggunakan data klinis pasien. Analisis tersebut bertujuan mengidentifikasi algoritma yang menunjukkan kinerja terbaik. Analisis dilakukan menggunakan data primer yang berasal dari Puskesmas Sandar Angin yang terdiri atas 201 data pasien yang dikumpulkan dalam kurun waktu Januari 2023 hingga Oktober 2025. Data tersebut terdiri atas empat belas variabel yang meliputi no, jenis kelamin, umur, nyeri dada, batuk berdahak, lama batuk (minggu), sesak nafas, kelelahan, penurunan berat badan (kg), suhu tubuh (°C), keringat dingin, nafsu makan berkurang, sputum, serta diagnosis sebagai label klasifikasi. Sebelum model dikembangkan, dilakukan serangkaian proses *pre-pemrosesan* yang terdiri atas penghapusan atribut yang tidak digunakan, pemeriksaan keberadaan *missing value* dan data duplikat, konversi data kategorikal melalui *encoding*, serta normalisasi atribut numerik menggunakan *StandardScaler*. Selanjutnya, *dataset* dipisahkan menjadi data pelatihan (*training*) dan data pengujian (*testing*) dengan proporsi 80:20. Penilaian terhadap model dilakukan menggunakan *confusion matrix* dengan menghitung metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Untuk memperoleh gambaran performa yang lebih representatif, penelitian ini juga menerapkan *5-Fold Cross Validation* guna mengevaluasi konsistensi hasil klasifikasi pada setiap *fold*. Gambar 1 menyajikan alur penelitian.

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943




**Gambar 1.** *Flowchart* Tahapan Penelitian

### 2.2 Pengumpulan Data

Sebagai sumber data penelitian, digunakan data primer yang berasal dari pasien TBC di Puskesmas Sandar Angin dengan cakupan periode Januari 2023 hingga Oktober 2025. Data yang diperoleh berjumlah 201 data dengan 14 atribut yang meliputi No, Jenis Kelamin, Umur, Nyeri Dada, Batuk Berdahak, Lama Batuk (minggu), Sesak Nafas, Kelelahan, Penurunan Berat Badan (kg), Suhu Tubuh (°C), Keringat Dingin, Nafsu Makan Berkurang, Sputum, dan Diagnosis. Tabel 1 berikut menyajikan deskripsi atribut yang digunakan dalam *dataset* penelitian.

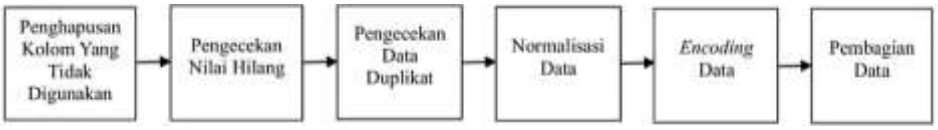
**Tabel 1.** Deskripsi Atribut

| No | Atribut | Deskripsi |
|---|---|---|
| 1 | No | Nomor urut data pasien yang digunakan sebagai identitas setiap *record*. |
| 2 | Jenis Kelamin | Jenis kelamin pasien yang terdiri atas laki-laki dan perempuan. |
| 3 | Umur | Usia pasien dalam satuan tahun. |
| 4 | Nyeri Dada | Menunjukkan apakah pasien mengalami gejala nyeri dada. |
| 5 | Batuk Berdahak | Menunjukkan apakah pasien mengalami batuk berdahak. |
| 6 | Lama Batuk (Minggu) | Lama waktu pasien mengalami batuk yang dinyatakan dalam satuan minggu. |
| 7 | Sesak Nafas | Menunjukkan apakah pasien mengalami gejala sesak napas. |
| 8 | Kelelahan | Menunjukkan apakah pasien mengalami gejala kelelahan. |
| 9 | Penurunan Berat Badan (Kg) | Besarnya penurunan berat badan pasien dalam satuan kilogram. |
| 10 | Suhu Tubuh (°C) | Suhu tubuh pasien yang diukur dalam satuan derajat Celcius. |
| 11 | Keringat Dingin | Menunjukkan apakah pasien mengalami gejala keringat dingin. |
| 12 | Nafsu Makan Berkurang | Menunjukkan apakah pasien mengalami penurunan nafsu makan. |
| 13 | Sputum | Hasil pemeriksaan dahak (sputum) pasien yang terdiri atas Positif dan Negatif. |
| 14 | Diagnosis | Label kelas yang menunjukkan hasil diagnosis pasien, yaitu TBC dan Tidak TBC. |

Atribut yang tercantum pada Tabel 1 berperan sebagai representasi karakteristik dan kondisi pasien dalam dataset penelitian. Atribut tersebut mencakup informasi yang relevan dengan kondisi klinis pasien dan hasil pemeriksaan, sedangkan diagnosis digunakan sebagai variabel target untuk membedakan pasien TBC dan tidak TBC.

### 2.3 Pra-pemrosesan Data

Pra-pemrosesan data merupakan proses persiapan dataset yang dilakukan sebelum tahap analisis dan pembentukan model untuk memastikan kualitas data yang digunakan memenuhi kebutuhan penelitian (Irfani & Gasim, 2024). Pada tahap ini, data yang belum optimal akan diperbaiki dengan menangani berbagai permasalahan, seperti adanya nilai yang hilang, data yang berlebihan, outlier, serta format data yang tidak seragam (Luthfi & Fauzi, 2024). Permasalahan tersebut dapat memengaruhi kualitas hasil pengolahan data, khususnya dalam proses klasifikasi. Dalam penelitian ini, pra-pemrosesan data dilakukan untuk meningkatkan kesiapan *dataset* sebelum memasuki proses pemodelan sehingga kualitas data yang digunakan dapat mendukung performa model klasifikasi (Rahman et al., 2024). Rangkaian tahapan pra-pemrosesan data disajikan pada Gambar 2.




**Gambar 2.** Tahapan Pra-pemrosesan Data

Tahap pra-pemrosesan data yang ditunjukkan pada Gambar 2 dilakukan sebelum proses klasifikasi. Tahap ini terdiri atas beberapa proses sebagai berikut:

1. Penghapusan Kolom yang Tidak Digunakan
   Penghapusan kolom merupakan proses menghilangkan atribut yang tidak relevan terhadap tujuan analisis dapat mengurangi kompleksitas data sekaligus meningkatkan efisiensi komputasi selama pemodelan (Oktaviani et al., 2024). Proses pemilihan fitur menjadi salah satu tahapan penting dalam klasifikasi karena tidak seluruh atribut pada dataset memberikan kontribusi terhadap pembentukan model, sehingga pemilihan atribut yang relevan dapat meningkatkan hasil klasifikasi (Astuti et al., 2022). Pada penelitian ini, atribut No, Jenis Kelamin, dan Umur dihapus karena tidak digunakan sebagai fitur dalam proses klasifikasi penyakit TBC. Sementara itu, atribut yang berkaitan dengan gejala TBC tetap dipertahankan sebagai variabel masukan dalam pembentukan model klasifikasi.

**TIN: Terapan Informatika Nusantara**
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

2. Pengecekan Nilai Hilang
   Pengecekan nilai hilang (*missing value*) dilakukan untuk memverifikasi kelengkapan data pada setiap atribut sebelum memasuki tahap pemodelan. Keberadaan nilai hilang dapat memengaruhi kualitas data dan menurunkan performa model sehingga perlu diidentifikasi sejak tahap awal (Biddinika et al., 2024).
3. Pengecekan Data Duplikat
   Pengecekan data duplikat bertujuan untuk mengidentifikasi data yang memiliki nilai identik pada seluruh atribut. Data duplikat dapat menyebabkan informasi menjadi berulang, meningkatkan risiko bias, serta memengaruhi hasil pelatihan model (Rindu et al., 2026).
4. *Encoding* Data
   *Encoding* merupakan tahapan pra-pemrosesan data yang mengonversi atribut kategorikal ke dalam bentuk numerik sehingga dapat dikenali dan diolah oleh algoritma *machine learning* (Sakmar et al., 2026). Pada penelitian ini, atribut Nyeri Dada, Batuk Berdahak, Sesak Nafas, Kelelahan, Keringat Dingin, dan Nafsu Makan Berkurang dikonversi dari nilai "Ya" dan "Tidak" menjadi 1 dan 0. Selain itu, atribut Sputum diubah dari "Positif" dan "Negatif" menjadi 1 dan 0, sedangkan atribut Diagnosis diubah dari "TBC" menjadi 1 dan "Tidak TBC" menjadi 0.
5. Normalisasi Data
   Normalisasi merupakan tahapan pra-pemrosesan data yang bertujuan menyamakan skala pada atribut numerik agar perbedaan rentang nilai antarfitur tidak memengaruhi proses pembentukan model (Sahona et al., 2026). Penelitian ini menggunakan metode *StandardScaler* untuk mentransformasikan atribut Lama Batuk (Minggu), Penurunan Berat Badan (Kg), dan Suhu Tubuh (°C). Atribut hasil *encoding* tidak dinormalisasi karena telah memiliki skala biner yang sama.
6. Pembagian Data
   Pembagian data merupakan tahapan untuk memisahkan *dataset* menjadi data *training* dan data *testing*. Data pelatihan digunakan untuk membangun model dan mempelajari pola pada *dataset*, sedangkan data pengujian dimanfaatkan untuk menilai kemampuan model dalam melakukan prediksi terhadap data yang tidak terlibat selama proses pelatihan (Putra et al., 2025). Dalam penelitian ini digunakan proporsi pembagian sebesar 80% untuk pelatihan dan 20% untuk pengujian. Untuk meningkatkan keandalan hasil penilaian, proses validasi juga dilakukan menggunakan metode *5-Fold Cross Validation*, sehingga evaluasi performa menjadi lebih representatif.

### 2.4 Penerapan *Support Vector Machine*

*Support Vector Machine* (SVM) merupakan salah satu algoritma *machine learning* berbasis *supervised learning* yang banyak digunakan dalam permasalahan klasifikasi maupun regresi. Pada proses klasifikasi, SVM bekerja dengan menentukan batas pemisah atau *hyperplane* yang optimal untuk memisahkan data ke dalam kelas yang berbeda. Penentuan *hyperplane* dilakukan dengan mempertimbangkan jarak atau *margin* antara batas pemisah dengan data terdekat dari masing-masing kelas, sehingga diperoleh pemisahan kelas yang optimal (Gasim et al., 2026). Data yang berada paling dekat dengan *hyperplane* disebut *support vector* dan memiliki peranan penting dalam menentukan posisi batas pemisah antar kelas. Selain mampu menangani data dengan jumlah atribut yang relatif banyak, SVM juga memiliki kemampuan generalisasi yang baik sehingga dapat menghasilkan klasifikasi terhadap data baru secara efektif (Mutmainah et al., 2024). Fungsi keputusan SVM dinyatakan pada Persamaan (1).

$$f(x) = w^T x + b \tag{1}$$

Dengan *w* melambangkan vektor bobot, *x* menyatakan vektor data masukan, dan *b* melambangkan konstanta bias.

### 2.5 Penerapan *Random Forest*

Sebagai salah satu metode *ensemble*, *Random Forest* memanfaatkan kumpulan *decision tree* untuk menghasilkan keputusan klasifikasi yang lebih andal. Pada tahap pelatihan, setiap pohon dibangun dari sampel data yang diperoleh melalui *bootstrap sampling* atau pengambilan data secara acak dengan pengembalian (*sampling with replacement*). Setelah seluruh pohon terbentuk, kelas akhir ditentukan berdasarkan hasil *majority voting*, yaitu kelas yang paling banyak dipilih oleh setiap *decision tree* (Wiyanto et al., 2025) . Selain menggunakan data pelatihan yang berbeda, algoritma ini juga memilih sebagian fitur secara acak pada setiap pemisahan *node*, sehingga terbentuk *decision tree* dengan karakteristik yang beragam. Setelah seluruh *decision tree* menghasilkan prediksi, kelas akhir ditentukan berdasarkan hasil *majority voting*, yaitu kelas yang paling banyak dipilih oleh seluruh pohon keputusan. Strategi tersebut membuat *Random Forest* menghasilkan prediksi yang lebih konsisten sekaligus meningkatkan ketahanan model terhadap *overfitting* yang umum dijumpai pada *decision tree* tunggal. Dengan kemampuan generalisasi yang baik terhadap data baru, algoritma ini banyak dimanfaatkan untuk menyelesaikan berbagai permasalahan klasifikasi, termasuk pada aplikasi di bidang kesehatan (A. Nugroho & Harini, 2024). Prediksi akhir pada algoritma *Random Forest* dinyatakan pada Persamaan (2).

$$\hat{y} = mode\left(h_1(x), h_2(x), \ldots, h_n(x)\right) \tag{2}$$

Dimana $h(x)$ menyatakan prediksi yang dihasilkan oleh *decision tree* ke-*i*, sedangkan *n* menunjukkan banyaknya *decision tree* yang membentuk model.

**TIN: Terapan Informatika Nusantara**
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

### 2.6 Evaluasi Model

Tahap evaluasi model bertujuan untuk mengetahui tingkat kinerja model klasifikasi melalui analisis *confusion matrix*. Metode ini dilakukan dengan mencocokkan hasil prediksi model terhadap label aktual sehingga diperoleh nilai *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). Keempat komponen tersebut menjadi dasar perhitungan *accuracy*, *precision*, *recall*, dan *F1-score*, yang digunakan untuk menilai ketepatan, kemampuan identifikasi, serta keseimbangan performa model dalam proses klasifikasi (Prasetyo, 2022). *Accuracy* menunjukkan persentase prediksi yang benar dari keseluruhan data yang dievaluasi, *precision* digunakan untuk menilai seberapa tepat model ketika memprediksi kelas positif, sementara *recall* mengukur kemampuan model dalam mengidentifikasi seluruh kasus positif yang tersedia. Di sisi lain, *F1-score* mengombinasikan nilai *precision* dan *recall* untuk memberikan gambaran performa model secara lebih seimbang (Adventino Gulo et al., 2025). Rumus yang digunakan ditunjukkan pada Persamaan (3) hingga Persamaan (6).

$$\text{Accuracy} = \frac{\text{TP+TN}}{\text{TP+TN+FP+FN}} \tag{3}$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP+FP}} \tag{4}$$

$$\text{recall} = \frac{\text{TP}}{\text{TP+FN}} \tag{5}$$

$$\text{F1 score} = 2\left(\frac{\text{recall} \cdot \text{presisi}}{\text{recall+presisi}}\right) \tag{6}$$

Dimana TP merupakan jumlah data positif yang diprediksi dengan benar sebagai positif, TN merupakan jumlah data negatif yang diprediksi dengan benar sebagai negatif, FP merupakan jumlah data negatif yang diprediksi sebagai positif, dan FN merupakan jumlah data positif yang diprediksi sebagai negatif.

## 3. HASIL DAN PEMBAHASAN

### 3.1 Pra-pemrosesan Data

*Dataset* yang digunakan dalam penelitian ini merupakan data primer yang berasal dari Puskesmas Sandar Angin sebanyak 201 data pasien dengan 14 atribut yang terdiri atas No, Jenis Kelamin, Umur, Nyeri Dada, Batuk Berdahak, Lama Batuk (Minggu), Sesak Nafas, Kelelahan, Penurunan Berat Badan (Kg), Suhu Tubuh (°C), Keringat Dingin, Nafsu Makan Berkurang, Sputum, dan Diagnosis. Proses klasifikasi diawali dengan pelaksanaan pra-pemrosesan data pada *dataset* sebagai langkah untuk memperbaiki kualitas data sekaligus memastikan bahwa data telah berada dalam kondisi yang layak untuk digunakan pada tahap pemodelan. Tahap pra-pemrosesan data diawali dengan penghapusan atribut No, Jenis Kelamin, dan Umur karena ketiga atribut tersebut tidak digunakan sebagai fitur dalam proses klasifikasi. Atribut No hanya berfungsi sebagai identitas atau penomoran data, sedangkan Jenis Kelamin dan Umur tidak digunakan dalam penelitian ini. Sementara itu, atribut yang berkaitan dengan gejala TBC dipertahankan karena memiliki keterkaitan dengan proses klasifikasi diagnosis dan digunakan sebagai variabel masukan dalam pembentukan model. Setelah proses seleksi atribut, dataset terlebih dahulu diverifikasi dengan memeriksa keberadaan *missing value* sebagai langkah untuk memastikan kelengkapan data pada setiap atribut. Berikutnya, dilakukan identifikasi terhadap data duplikat guna menghindari penggunaan entri yang sama lebih dari satu kali, sehingga potensi bias dalam proses klasifikasi dapat diminimalkan. Tahap pemeriksaan kualitas data memperlihatkan bahwa dataset terbebas dari *missing value* dan data duplikat sehingga dapat langsung digunakan pada proses pembentukan model tanpa dilakukan tahap *data cleaning* tambahan. Informasi mengenai atribut yang dipertahankan setelah proses seleksi disajikan pada Tabel 2.

**Tabel 2.** *Dataset* Setelah Seleksi Atribut

| Nyeri Dada | Batuk Berdahak | Lama Batuk (Minggu) | Sesak Nafas | Kelelahan | Penurunan Berat Badan (Kg) | Suhu Tubuh ( °C) | Keringat Dingin | Nafsu Makan Berkurang | Sputum | Diagnosis |
|---|---|---|---|---|---|---|---|---|---|---|
| Ya | Tidak | 5 | Ya | Ya | 4 | 36.8 | Ya | Ya | Negatif | TBC |
| Ya | Tidak | 4 | Ya | Tidak | 0 | 36.5 | Tidak | Tidak | Positif | Tidak TBC |
| Tidak | Tidak | 1 | Tidak | Ya | 1 | 37.2 | Tidak | Tidak | Negatif | Tidak TBC |
| Ya | Ya | 6 | Ya | Ya | 3 | 37.8 | Ya | Ya | Positif | TBC |
| Tidak | Ya | 2 | Ya | Tidak | 1 | 37.0 | Tidak | Ya | Positif | Tidak TBC |
| Ya | Ya | 7 | Ya | Ya | 6 | 38.5 | Ya | Ya | Positif | TBC |
| Tidak | Tidak | 1 | Tidak | Ya | 0 | 36.8 | Tidak | Tidak | Negatif | Tidak TBC |
| Ya | Ya | 3 | Ya | Ya | 4 | 38.0 | Ya | Ya | Negatif | TBC |
| Tidak | Ya | 2 | Ya | Tidak | 1 | 37.3 | Tidak | Tidak | Negatif | Tidak TBC |
| Tidak | Tidak | 2 | Tidak | Ya | 0 | 37.4 | Ya | Tidak | Negatif | Tidak TBC |

**TIN: Terapan Informatika Nusantara**
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

Berdasarkan Tabel 2, dataset yang telah diperoleh setelah seleksi atribut dan pemeriksaan kualitas data selanjutnya diproses melalui tahap *encoding*. Tahap ini dilakukan untuk mengubah data kategorikal menjadi representasi numerik sehingga dapat digunakan dalam proses klasifikasi. Dalam penelitian ini, atribut Nyeri Dada, Batuk Berdahak, Sesak Nafas, Kelelahan, Keringat Dingin, dan Nafsu Makan Berkurang dikodekan dengan memberikan nilai 1 untuk kategori "Ya" dan nilai 0 untuk kategori "Tidak". Selain itu, atribut Sputum diubah dari kategori "Positif" dan "Negatif" menjadi nilai 1 dan 0, sedangkan variabel Diagnosis ditransformasikan menjadi label biner, yaitu TBC bernilai 1 dan Tidak TBC bernilai 0 sebagai target klasifikasi. Hasil proses *encoding* menghasilkan seluruh atribut kategorikal dalam bentuk numerik sehingga siap digunakan pada tahap pemodelan. Setelah proses *encoding*, dilakukan normalisasi data menggunakan metode *StandardScaler* pada atribut numerik Lama Batuk (Minggu), Penurunan Berat Badan (Kg), dan Suhu Tubuh (°C). Melalui metode *StandardScaler*, setiap nilai ditransformasikan sehingga memiliki nilai *mean* sebesar 0 dan *standard deviation* sebesar 1, yang membuat seluruh atribut numerik berada pada skala yang seragam. Sementara itu, atribut hasil encoding tidak dinormalisasi karena telah memiliki skala biner yang sama, yaitu 0 dan 1, sehingga tidak memerlukan transformasi lebih lanjut. Setelah seluruh tahapan pra-pemrosesan data selesai dilakukan, dataset telah memenuhi kebutuhan untuk digunakan dalam pelatihan dan pengujian model klasifikasi. Adapun hasil *encoding* dan normalisasi disajikan pada Tabel 3.

**Tabel 3.** Hasil *Encoding* dan Normalisasi

| Nyeri Dada | Batuk Berdahak | Lama Batuk (Minggu) | Sesak Nafas | Kelelahan | Penurunan Berat Badan (Kg) | Suhu Tubuh ( °C) | Keringat Dingin | Nafsu Makan Berkurang | Sputum | Diagnosis |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 0 | 0.728272 | 1 | 1 | 0.321502 | -0.599283 | 1 | 1 | 0 | 1 |
| 1 | 0 | 0.226962 | 1 | 0 | -1.083324 | -0.900926 | 0 | 0 | 1 | 0 |
| 0 | 0 | -1.276971 | 0 | 1 | -0.732117 | -0.197093 | 0 | 0 | 0 | 0 |
| 1 | 1 | 1.229583 | 1 | 1 | -0.029704 | 0.406192 | 1 | 1 | 1 | 1 |
| 0 | 1 | -0.775660 | 1 | 0 | -0.732117 | -0.398188 | 0 | 1 | 1 | 0 |
| 1 | 1 | 1.730894 | 1 | 1 | 1.023916 | 1.110025 | 1 | 1 | 1 | 1 |
| 0 | 0 | -1.276971 | 0 | 1 | -1.083324 | -0.599283 | 0 | 0 | 0 | 0 |
| 1 | 1 | -0.274349 | 1 | 1 | 0.321502 | 0.607287 | 1 | 1 | 0 | 1 |
| 0 | 1 | -0.775660 | 1 | 0 | -0.732117 | -0.096546 | 0 | 0 | 0 | 0 |
| 0 | 0 | -0.775660 | 0 | 1 | -1.083324 | 0.004002 | 1 | 0 | 0 | 0 |

Hasil *encoding* dan normalisasi pada Tabel 3 menunjukkan bahwa seluruh atribut yang digunakan dalam penelitian telah berada dalam bentuk numerik dan siap digunakan dalam proses klasifikasi. Tahap selanjutnya adalah membagi *dataset* menjadi data *training* dan data *testing* dengan rasio 80:20. Berdasarkan jumlah keseluruhan 201 data, sebanyak 160 data digunakan sebagai data latih, sedangkan 41 data digunakan sebagai data uji. Pembagian data dilakukan secara acak dengan mempertahankan proporsi kelas pada masing-masing subset agar distribusi kelas TBC dan Tidak TBC tetap representatif. Data latih selanjutnya digunakan untuk membangun model SVM dan *Random Forest*, sedangkan data uji digunakan untuk mengukur kinerja kedua model berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*.

### 3.2 Penerapan *Support Vector Machine*

Setelah proses pembagian data menjadi data *training* dan data *testing* selesai dilakukan, tahap selanjutnya adalah membangun model klasifikasi menggunakan algoritma SVM. Pada penelitian ini, SVM digunakan untuk mengklasifikasikan data pasien ke dalam dua kelas, yaitu TBC dan Tidak TBC. Model dibangun menggunakan data *training* yang telah melalui tahap pra-pemrosesan, kemudian digunakan untuk melakukan klasifikasi terhadap data *testing*. Implementasi SVM dilakukan menggunakan pustaka *scikit-learn* dengan kelas SVC. Kernel yang digunakan adalah linear kernel dengan *random state* sebesar 42. Penggunaan linear kernel bertujuan untuk membentuk *hyperplane* yang dapat memisahkan data dari kedua kelas berdasarkan kombinasi atribut yang digunakan. Parameter *random state* ditetapkan untuk menjaga konsistensi proses komputasi sehingga hasil yang diperoleh dapat direproduksi. Model SVM dilatih menggunakan data *training* melalui proses *fitting*. Setelah model terbentuk, data *testing* dimasukkan ke dalam model untuk menghasilkan prediksi kelas. Hasil prediksi tersebut selanjutnya digunakan pada tahap evaluasi untuk mengukur kinerja model berdasarkan *accuracy, precision, recall,* dan *F1-score.*

### 3.3 Penerapan *Random Forest*

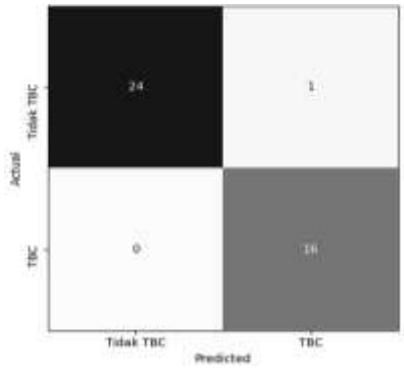
Sebagai metode pembanding, penelitian ini juga menerapkan algoritma *Random Forest* untuk melakukan klasifikasi diagnosis TBC. *Random Forest* merupakan metode *ensemble learning* yang menghasilkan prediksi dengan menggabungkan sejumlah *decision tree*. Pendekatan tersebut memungkinkan model menghasilkan klasifikasi yang lebih stabil dengan memanfaatkan hasil prediksi dari beberapa pohon keputusan. Implementasi *Random Forest* dilakukan menggunakan pustaka *scikit-learn* melalui kelas *RandomForestClassifier*. Model dikonfigurasi dengan jumlah 100 *decision tree* melalui *parameter n_estimators=100* dan *random state* sebesar 42. Penetapan jumlah pohon bertujuan untuk membentuk model *ensemble* yang dapat digunakan dalam proses klasifikasi, sedangkan *random state*

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

digunakan untuk menjaga konsistensi hasil pada proses pemodelan. Model *Random Forest* dilatih menggunakan data *training* yang telah dipersiapkan pada tahap sebelumnya. Setelah proses pelatihan selesai, model digunakan untuk memprediksi kelas pada data *testing*. Hasil prediksi kemudian digunakan untuk mengevaluasi kinerja model dengan *accuracy*, *precision*, *recall*, dan *F1-score*. Hasil evaluasi SVM dan *Random Forest* selanjutnya dibandingkan untuk menentukan algoritma yang memberikan kinerja klasifikasi terbaik pada dataset penelitian.

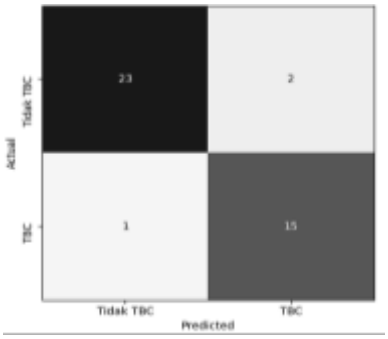
### 3.4 Evaluasi Model

Setelah model SVM dan *Random Forest* dibangun dan menghasilkan prediksi pada data *testing*, tahap selanjutnya adalah melakukan evaluasi untuk mengukur kinerja masing-masing model dalam mengklasifikasikan pasien TBC dan Tidak TBC. Tahap evaluasi dilakukan dengan memanfaatkan *confusion matrix* untuk membandingkan kelas prediksi dengan label aktual pada data pasien. Melalui pendekatan ini dapat diketahui distribusi hasil klasifikasi yang benar maupun yang salah, sekaligus diperoleh nilai TP, TN, FP, dan FN. Komponen tersebut selanjutnya digunakan untuk menghitung *accuracy*, *precision*, *recall*, dan *F1-score* sebagai dasar penilaian terhadap kinerja model. Hasil perhitungan setiap metrik kemudian menjadi acuan dalam mengevaluasi dan membandingkan performa algoritma SVM serta *Random Forest* pada klasifikasi pasien TBC dan Tidak TBC. Visualisasi *confusion matrix* untuk model SVM ditampilkan pada Gambar 3.




**Gambar 3.** *Confusion Matrix* SVM

Gambar 3 memperlihatkan bahwa algoritma SVM menghasilkan nilai TN sebesar 24, FP sebesar 1, FN sebesar 0, dan TP sebesar 16. Hasil tersebut menunjukkan bahwa SVM mampu mengklasifikasikan seluruh data pasien TBC dengan benar tanpa menghasilkan *False Negative* yang berarti tidak terdapat pasien TBC yang diprediksi sebagai Tidak TBC. Tidak ditemukannya nilai FN menunjukkan bahwa seluruh pasien TBC pada data pengujian berhasil dikenali sebagai kasus positif sehingga tidak ada pasien TBC yang salah diklasifikasikan sebagai Tidak TBC. Meskipun demikian, masih terdapat satu data pasien Tidak TBC yang diprediksi sebagai TBC, sehingga menghasilkan satu nilai *False Positive*. Temuan tersebut menandakan bahwa masih terdapat sejumlah kecil sampel pada kategori Tidak TBC yang belum dapat diidentifikasi secara tepat oleh model. Terlepas dari hal tersebut, hasil evaluasi melalui *confusion matrix* memperlihatkan bahwa SVM mampu memberikan prediksi yang benar pada sebagian besar data pengujian. Kinerja tersebut didukung oleh capaian *accuracy*, *precision*, *recall*, dan *F1-score* yang tinggi. Visualisasi *confusion matrix* untuk algoritma *Random Forest* disajikan pada Gambar 4.




**Gambar 4.** *Confusion Matrix Random Forest*

Gambar 4 memperlihatkan bahwa algoritma *Random Forest* menghasilkan nilai TN sebanyak 23, FP sebanyak 2, FN sebanyak 1, dan TP sebanyak 15. Secara keseluruhan, *Random Forest* mampu menghasilkan prediksi yang tepat pada mayoritas data pengujian. Meski demikian, model masih keliru dalam mengidentifikasi tiga sampel, yang terdiri atas dua sampel Tidak TBC yang diklasifikasikan sebagai TBC dan satu sampel TBC yang diklasifikasikan sebagai Tidak TBC. Kondisi ini menunjukkan bahwa performa model belum sepenuhnya bebas dari kesalahan klasifikasi pada kedua kategori. Jika dibandingkan dengan algoritma SVM, jumlah kesalahan klasifikasi pada *Random Forest* lebih

**TIN: Terapan Informatika Nusantara**
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

banyak karena masih terdapat FP dan FN, sedangkan SVM hanya menghasilkan satu FP tanpa FN. Tabel 4 menyajikan hasil evaluasi algoritma SVM dan *Random Forest* berdasarkan nilai *accuracy*, *precision*, *recall*, dan *F1-score*.

**Tabel 4.** Hasil Pengukuran Kinerja SVM dan *Random Forest*

| Algoritma | *Accuracy* | *Precision* | *Recall* | *F1-Score* |
|---|---|---|---|---|
| SVM | 98% | 94% | 100% | 97% |
| *Random Forest* | 93% | 88% | 94% | 91% |

Dari hasil evaluasi pada Tabel 4 terlihat bahwa SVM menghasilkan capaian metrik yang melampaui *Random Forest* pada *accuracy*, *precision*, *recall*, dan *F1-score*. Nilai *recall* sebesar 100% mengindikasikan bahwa seluruh kasus TBC pada data pengujian berhasil teridentifikasi tanpa ada kasus positif yang terlewat. Keunggulan pada *precision* dan *F1-score* juga menunjukkan bahwa SVM mampu menghasilkan prediksi yang lebih andal dengan keseimbangan yang baik antara ketepatan identifikasi dan cakupan deteksi kasus positif. Berdasarkan evaluasi dengan skema pembagian data 80:20, algoritma SVM menunjukkan hasil klasifikasi yang lebih unggul dibandingkan *Random Forest* pada dataset penelitian ini. Untuk menilai konsistensi performa model secara lebih menyeluruh, dilakukan evaluasi lanjutan menggunakan metode *5-Fold Cross Validation*. Ringkasan hasil evaluasi tersebut disajikan pada Tabel 5.

**Tabel 5.** Hasil 5-*Fold Cross Validation*

| *K-Fold* | *Accuracy* | | *Precision* | | *Recall* | | *F1-Score* | |
|---|---|---|---|---|---|---|---|---|
| | SVM | RF | SVM | RF | SVM | RF | SVM | RF |
| *Fold k=1* | 0.83 | 0.83 | 0.67 | 0.67 | 0.92 | 0.92 | 0.77 | 0.77 |
| *Fold k=2* | 0.90 | 0.88 | 0.94 | 0.88 | 0.83 | 0.83 | 0.88 | 0.86 |
| *Fold k=3* | 0.88 | 0.88 | 0.78 | 0.78 | 0.70 | 0.70 | 0.74 | 0.74 |
| *Fold k=4* | 0.88 | 0.85 | 1.00 | 1.00 | 0.72 | 0.67 | 0.84 | 0.80 |
| *Fold k=5* | 0.87 | 0.88 | 1.00 | 1.00 | 0.72 | 0.72 | 0.84 | 0.84 |
| *Best Fold* | 0.90 | 0.88 | 0.94 | 0.88 | 0.83 | 0.83 | 0.88 | 0.86 |

Evaluasi menggunakan skema *5-Fold Cross Validation* sebagaimana disajikan pada Tabel 5 menunjukkan bahwa kedua algoritma mampu mempertahankan kinerja yang relatif stabil pada seluruh *fold*. Meskipun terdapat sedikit perbedaan nilai pada setiap pengujian, kedua algoritma mampu mempertahankan performa yang stabil selama proses validasi. SVM mencatat *accuracy* terbaik sebesar 0,90 pada *Fold* ke-2, sedangkan *Random Forest* memperoleh nilai tertinggi sebesar 0,88 yang muncul pada *Fold* ke-2, *Fold* ke-3, dan *Fold* ke-5. Hasil serupa juga terlihat pada metrik *precision*, *recall*, dan *F1-score*, di mana SVM secara konsisten menunjukkan capaian yang lebih unggul daripada *Random Forest*. Berdasarkan keseluruhan hasil validasi, *Fold* ke-2 dipilih sebagai acuan evaluasi karena menghasilkan kombinasi nilai *accuracy*, *precision*, *recall*, dan *F1-score* yang paling optimal pada algoritma SVM maupun *Random Forest*. Pemilihan *fold* terbaik ini dilakukan untuk memberikan gambaran yang lebih representatif terhadap kemampuan masing-masing algoritma pada proses klasifikasi. Hasil performa terbaik dari kedua algoritma berdasarkan Fold ke-2 selanjutnya dirangkum pada Tabel 6.

**Tabel 6.** Hasil Terbaik *5-fold Cross Validation*

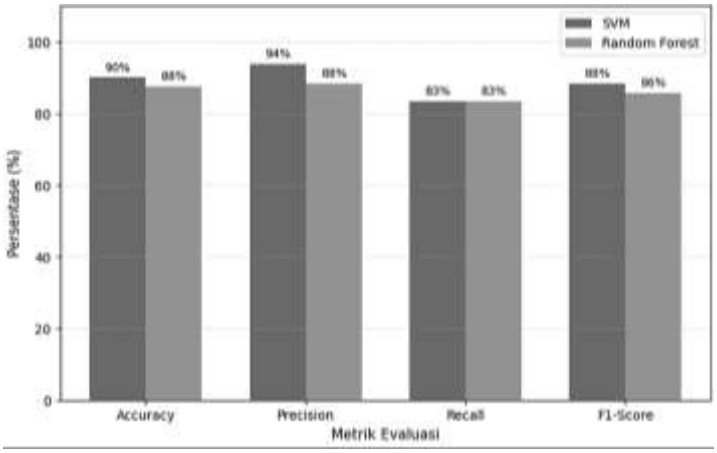
| Algoritma | *Accuracy* | *Precision* | *Recall* | *F1-Score* |
|---|---|---|---|---|
| SVM | 90% | 94% | 83% | 88% |
| *Random Forest* | 88% | 88% | 83% | 86% |

Dari hasil evaluasi yang disajikan pada Tabel 6, SVM menjadi algoritma dengan capaian terbaik pada *fold* terpilih dalam *5-Fold Cross Validation*, ditinjau dari metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Keunggulan tersebut menunjukkan bahwa SVM lebih efektif dalam menangkap karakteristik data sehingga mampu menghasilkan klasifikasi yang lebih andal dibandingkan *Random Forest*. Meskipun demikian, *Random Forest* juga memperoleh hasil evaluasi yang relatif tinggi, sehingga kedua algoritma tetap memiliki potensi untuk diterapkan pada klasifikasi penyakit TBC. Untuk mempermudah interpretasi hasil serta memperjelas perbandingan performa kedua algoritma, hasil evaluasi pada *fold* terbaik kemudian divisualisasikan dalam bentuk grafik batang yang membandingkan metrik *accuracy*, *precision*, *recall*, dan *F1-score*. Penyajian grafik tersebut disampaikan pada Gambar 5.

Dari grafik pada Gambar 5 terlihat bahwa SVM unggul atas *Random Forest* pada sebagian besar parameter evaluasi. SVM mencatat *accuracy* sebesar 90%, sedangkan *Random Forest* mencapai 88%. Pola yang sama juga terlihat pada metrik *precision*, dengan capaian masing-masing sebesar 94% dan 88%. Namun, kedua algoritma menghasilkan nilai *recall* yang sama, yaitu 83%, sehingga keduanya memiliki tingkat keberhasilan yang setara dalam mendeteksi kasus TBC pada *fold* terbaik. Meskipun nilai *recall* kedua algoritma sama, perbedaan pada nilai *accuracy* dan *precision* menyebabkan nilai *F1-score* SVM lebih tinggi, yaitu 88%, sedangkan *Random Forest* memperoleh 86%. Temuan tersebut mengindikasikan bahwa SVM mampu menghasilkan prediksi dengan tingkat ketepatan yang lebih tinggi sekaligus mempertahankan keseimbangan yang lebih baik antara nilai *precision* dan *recall* Visualisasi pada Gambar 5 memperlihatkan bahwa SVM menjadi algoritma dengan hasil evaluasi terbaik pada *fold* terpilih dalam *5-Fold Cross*

TIN: Terapan Informatika Nusantara
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

*Validation*, sehingga SVM dinilai lebih sesuai untuk diterapkan pada proses klasifikasi penyakit TBC menggunakan *dataset* pada penelitian ini.




**Gambar 5.** Perbandingan Metrik Evaluasi SVM dan *Random Forest* pada *Fold* Terbaik

### 3.5 Pembahasan

Berdasarkan evaluasi yang dilakukan melalui skema pembagian data 80:20 dan 5-Fold Cross Validation, algoritma SVM memberikan hasil klasifikasi yang lebih unggul daripada Random Forest pada data pasien TBC di Puskesmas Sandar Angin . Keunggulan tersebut tercermin dari kemampuan SVM dalam mempelajari karakteristik data sehingga memperoleh capaian accuracy, precision, recall, dan F1-score yang lebih baik. Hasil penelitian ini konsisten dengan temuan (Triloka & Sugianto, 2025) yang melaporkan bahwa algoritma SVM mencapai tingkat accuracy sebesar 99,61%, lebih tinggi dibandingkan *Random Forest* yang memperoleh 81,80% pada prediksi hasil pengobatan pasien TBC. Walaupun objek kajian kedua penelitian berbeda, yaitu prediksi hasil pengobatan dan klasifikasi penyakit TBC, keduanya mengindikasikan bahwa SVM mampu memberikan kinerja yang unggul pada data yang berkaitan dengan TBC.

Temuan penelitian ini memberikan hasil yang berbeda dibandingkan penelitian (Manurung et al., 2025) dan (Aini et al., 2026), di mana pada kedua penelitian tersebut Random Forest dilaporkan sebagai algoritma dengan performa paling baik. Perbedaan tersebut dapat disebabkan oleh variasi pada komposisi dataset, jumlah sampel, keseimbangan distribusi kelas, maupun karakteristik atribut yang digunakan selama proses pemodelan. Penelitian (Aini et al., 2026), misalnya, menggunakan 1.888 data pasien, sedangkan penelitian ini hanya menggunakan 201 data pasien, sehingga karakteristik data yang dipelajari oleh model juga berbeda. Dukungan terhadap hasil penelitian ini juga terlihat pada penelitian (Azhari et al., 2021) yang melaporkan keunggulan SVM dengan accuracy sebesar 95,00% dibandingkan Random Forest yang memperoleh 83,33% pada konteks klasifikasi yang berbeda. Kondisi tersebut menguatkan bahwa tingkat keberhasilan algoritma dapat berubah sesuai dengan karakteristik data dan konteks permasalahan yang dihadapi. Oleh sebab itu, SVM dinilai lebih sesuai untuk diterapkan pada klasifikasi penyakit TBC menggunakan data pasien Puskesmas Sandar Angin, dengan tetap mempertimbangkan pengaruh karakteristik *dataset* dan parameter model terhadap hasil yang diperoleh.

## 4. KESIMPULAN

Penelitian ini menunjukkan bahwa algoritma SVM dan *Random Forest* dapat diterapkan untuk melakukan klasifikasi penyakit TBC berdasarkan data klinis pasien di Puskesmas Sandar Angin Pagar Alam. Berdasarkan hasil pengujian dan validasi yang telah dilakukan, SVM menunjukkan kinerja yang lebih unggul dibandingkan *Random Forest* dalam mengklasifikasikan pasien ke dalam kategori TBC dan Tidak TBC. Keunggulan SVM terlihat dari capaian metrik evaluasi yang secara umum lebih tinggi serta kemampuannya mengenali seluruh kasus TBC pada data pengujian tanpa menghasilkan *False Negative*. Hasil tersebut menunjukkan bahwa SVM memiliki kemampuan yang lebih baik dalam membedakan karakteristik kedua kelas pada *dataset* yang digunakan. Meskipun demikian, hasil penelitian tidak dapat secara langsung digeneralisasikan pada populasi yang lebih luas karena dataset yang digunakan masih terbatas dan hanya berasal dari satu fasilitas pelayanan kesehatan. Perbedaan karakteristik pasien, jumlah data, distribusi kelas, serta kondisi pelayanan kesehatan pada lokasi lain berpotensi memengaruhi kinerja model. Dengan demikian, pemilihan algoritma klasifikasi perlu mempertimbangkan karakteristik data dan tujuan penerapannya, bukan hanya berdasarkan kinerja pada satu dataset. Secara keseluruhan, hasil penelitian memberikan indikasi bahwa SVM merupakan pendekatan yang lebih sesuai untuk klasifikasi penyakit TBC pada dataset penelitian ini dan dapat menjadi dasar bagi pengembangan sistem klasifikasi berbasis data klinis. Penelitian selanjutnya disarankan menggunakan dataset dengan jumlah dan variasi yang lebih besar dari beberapa fasilitas pelayanan kesehatan, melakukan optimasi parameter model, serta membandingkan metode klasifikasi lainnya untuk memperoleh model yang lebih robust dan memiliki kemampuan generalisasi yang lebih baik.


**TIN: Terapan Informatika Nusantara**
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

## REFERENCES

Adventino Gulo, S., Amelia Pertiwi, A., Putri Syaifullah Nasution, S., & Syahputra, H. (2025). Deteksi Deepfake Dalam Citra Menggunakan Convolutional Neural Network (Cnn). *JATI (Jurnal Mahasiswa Teknik Informatika)*, *9*(5), 8655–8660. https://doi.org/10.36040/jati.v9i5.14896

Aini, S. A., Fitriani, D., & Awwalin, M. A. (2026). Klasifikasi penyakit jantung berdasarkan faktor klinis menggunakan algoritma XGBoost, Random Forest, Support Vector Machine (SVM) dan K-Nearest Neighbor (KNN). *JATI (Jurnal Mahasiswa Teknik Informatika), 10(2), 3449–3454. https://doi.org/10.36040/jati.v10i2.17644*

Astuti, L. W., Saluza, I., Yulianti, E., & Dhamayanti. (2022). Feature Selection Menggunakan Binary Wheal Optimizaton Algorithm ( BWOA ) pada Klasifikasi Penyakit Diabetes. *Jurnal Ilmiah Informatika Global*. 13(4), 1–6. https://doi.org/10.36982/jiig.v13i1.2057

Azhari, M., Situmorang, Z., & Rosnelly, R. (2021). Perbandingan Akurasi , Recall , dan Presisi Klasifikasi pada Algoritma. *Jurnal Media Informatika Budidarma*. *5*(April), 640–651. https://doi.org/10.30865/mib.v5i2.2937

Biddinika, M. K., Masitha, A., & Fatimah, F. A. N. (2024). Machine Learning Techniques for Heart Disease Prediction Using a Multi Algorithm Approach. 12(2), 149–158. https://doi.org/10.30595/juita.v12i2.24153

Du, K. L., Jiang, B., Lu, J., Hua, J., & Swamy, M. N. S. (2024). Exploring Kernel Machines and Support Vector Machines: Principles, Techniques, and Future Directions. *Mathematics*, *12*(24), 1–58. https://doi.org/10.3390/math12243935

Gasim, Heriansyah, R., Irfani, M. H., & Cahyani, S. (2026). *Buku Ajar Pengenalan Pola Pattern Recognition* (1st ed.). Repository Global Aksara Pers.

Ihsan, C. (2025). *Dasar-Dasar Machine Learning Teori, Algoritma, dan Implementasi*. Greenbook Publisher.

Irfani, M. H., & Gasim. (2024). *Segmentasi teks pada citra tulisan tangan kalimat menggunakan metode Median Filtering dan Otsu*. *18*(April), 88–97. https://doi.org/10.24252/teknosains.v18i1.44307

Kementerian Kesehatan Republik Indonesia. (2024). *Laporan Kinerja Kementerian Kesehatan Republik Indonesia Tahun 2024*.

Luthfi, A. M., & Fauzi, F. (2024). *Perbandingan Klasifikasi Random Forest ,* Support Vector Machines , dan LGBM Pada Klasifikasi Kualitas Udara di Jakarta. *JUSTINDO (Jurnal Sistem dan Teknologi Informasi Indonesia) 9*(2), 99–108. https://doi.org/10.32528/justindo.v9i2.1912

Manurung, S. Z. Y. B., Aldo, D., & Sa'adah, A. (2025). Perbandingan Algoritma Support Vector Machine Dan Algoritma Random Forest Dalam Prediksi Hipertensi. *E-Proceeding of Engineering*, *12*(4), 6934.

Mutmainah, M., Cipta, S. P., Mambang, M., Zulfadhilah, M., Naparin, H., & Syapotro, U. (2024). Analisis Sentimen Terhadap Aplikasi Parak Acil Online Berdasarkan Ulasan Masyarakat Menggunakan Metode Support Vector Machine (SVM). *Jurnal Nasional Komputasi Dan Teknologi Informasi (JNKTI)*, *7*(5), 1042–1049. https://doi.org/10.32672/jnkti.v7i5.7962

Nugroho, A., & Harini, D. (2024). Teknik Random Forest untuk Meningkatan Akurasi Data Tidak Seimbang. *JSITIK: Jurnal Sistem Informasi Dan Teknologi Informasi Komputer*, *2*(2), 128–140. https://doi.org/10.53624/jsitik.v2i2.379

Nugroho, M. W. (2025). Analisis Performa Algoritma Random Forest dalam Mengatasi Overfitting pada Model Prediksi. *Jurnal JTIK (Jurnal Teknologi Informasi Dan Komunikasi)*, *9*(4), 1562–1571. https://doi.org/10.35870/jtik.v9i4.4236

Oktaviani, V., Rosmawarni, N., & Muslim, M. P. (2024). *Perbandingan Kinerja Random Forest Dan Smote Random Forest Dalam Mendeteksi Dan Mengukur Tingkat Stres Pada Mahasiswa Tingkat Akhir*. *4221*(April), 43–49. https://doi.org/10.52958/iftk.v20i1.9158

Prasetyo, E. (2022). *Data Mining: Konsep dan Aplikasi menggunakan MATLAB*. ANDI.

Putra, R. S., Izhari, F., Syekh, U. I. N., Hasan, A., & Addary, A. (2025). Analisis Komparasi Algoritma Random Forest dan Support Vector Machine untuk Deteksi Intrusi Jaringan. *TECHSI: Jurnal Teknik Informatika*, 16(2), 74–87. https://doi.org/10.29103/techsi.v16i2.25811

Rahman, F. D., Zulfa, M. I., & Taryana, A. (2024). Clustering dan Klasifikasi Data Cuaca Kota Cilacap Menggunakan K-Means dan Random Forest. 1(April), 90–97. https://doi.org/10.61124/sinta.v1i2.15

Rindu, A. A. C., Astriratma, R., & Zaidiah, A. (2026). K-Means Algorithm Implementation for Project Health Clustering. 5(158), 1064–1076. https://doi.org/10.29207/resti.v7i5.5181

Sahona, M. A., Astuti, L. W., & Irfani, M. H. (2026). *Klasifikasi Usia Pengguna Berdasarkan Nilai Guna Perangkat Seluler Menggunakan Metode Support Vector Machine*. *7*(3), 815–821. https://doi.org/https://doi.org/10.55338/jumin.v7i3.8798

Sakmar, M., Kadir, N. T., Shofo, P. A., & Darmawan, A. (2026). Efektivitas XGBoost, LightGBM, dan CatBoost pada Dataset Imbalanced Predictive Maintenance. *3*, 36–44. https://doi.org/10.61124/sinta.v3i1.145

Triloka, J., & Sugianto, D. (2025). Prediction of Tuberculosis Treatment Outcomes in Indonesia Using Support Vector Machine and Random Forest. *Journal of Applied Informatics and Computing*, *9*(5), 2478–2485. https://doi.org/10.30871/jaic.v9i5.10018

Wahyudi, A., & Setyawan, D. (2022). *Manajemen Pelayanan Kesehatan*. Deepublish.

Wiyanto, A., Puspasari, S., Astuti, L. W., Komputer, I., Informatika, T., Indo, U., & Mandiri, G. (2025). Perbandingan Kinerja Algoritma Support Vector Machine dan Random Forest dalam Analisis Sentimen Ulasan Hotel di Kota

**TIN: Terapan Informatika Nusantara**
Vol 7, No 3, August 2026, page 1642-1652
ISSN 2722-7987 (Media Online)
Website https://ejurnal.seminar-id.com/index.php/tin
DOI 10.47065/tin.v7i3.10943

Palembang pada Google Map. Jurnal Teknik Informatika dan Teknologi Informasi, 2, 679–691. https://doi.org/https://doi.org/10.55606/jutiti.v5i2.5802

World Health Organization. (2024). *Global tuberculosis report 2024*. https://iris.who.int/handle/10665/379339

Zahra, A., Azzahirah, A., & Ahmad, A. M. (2025). *Penerapan Support Vector Machine untuk Klasifikasi Tuberkulosis Paru dari Citra Rontgen Dada*. *1*(1), 17–25.