

Classification of Tuberculosis in Rumah Sakit Paru Sumatera Barat Using the C5.0 Algorithm

Meliani Maya Sari, Zilrahmi*, Dony Permana, dan Dwi Sulistiowati

Departemen Statistika, Universitas Negeri Padang, Kota Padang, Negara Indonesia

*Corresponding author: zilrahmi@fmipa.ac.id

Submitted : 23 Desember 2025

Revised : 28 Januari 2026

Accepted : 26 Januari 2026

ABSTRACT

Tuberculosis (TB) remains a serious public health problem, including in West Sumatra Province, where the number of reported cases has continued to increase in recent years. Consequently, effective methods are required to support early detection and accurate classification of TB patients. This study aims to classify the tuberculosis status of patients at Rumah Sakit Paru Sumatera Barat by applying the C5.0 algorithm. The data used in this study consists of secondary data extracted from patient medical records collected from October to December 2024 with a total of 150 patient medical records. The dataset included eight predictor variables representing clinical symptoms and one target variable, namely sputum smear (BTA) examination results. The research process involved data preprocessing, after which the dataset was divided into training and testing subsets using a 70:30 ratio, a classification model was developed using the C5.0 algorithm, and its performance was evaluated using a confusion matrix. The findings indicate that the C5.0 algorithm achieved an accuracy of 91.11%, with a precision of 95.83%, sensitivity of 88.46%, and specificity of 94.74%. Night sweats were identified as the most influential variable in the construction of the decision tree. These findings indicate that the C5.0 algorithm demonstrates excellent performance and can be applied as a decision support method for classifying tuberculosis based on patients' clinical symptoms.

Keywords: Tuberculosis, Data Mining, Classification, Decision Tree, C5.0 Algorithm



This is an open access article under the Creative Commons 4.0 Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. ©2022 by author and Universitas Negeri Padang.

I. PENDAHULUAN

Tuberculosis (TBC) adalah penyakit infeksi yang timbul akibat bakteri *mycobacterium tuberculosis*. TBC tergolong sebagai penyakit infeksi menular yang penyebarannya terjadi melalui udara. Menurut Febriyanto dan Kurniawan (2018), menyatakan bahwa TBC tergolong sebagai penyakit yang berbahaya dan perlu mendapatkan kewaspadaan tinggi karena berpotensi menyebabkan kematian. Penularan penyakit TBC dapat terjadi melalui bakteri yang tersebar ke udara melalui *droplet* (percikan dahak) ketika pasien TBC positif mengalami batuk dan bersin. Infeksi bakteri TBC akan menimbulkan gejala umum yang dialami bagi penderita TBC seperti batuk berkepanjangan yang terjadi selama 3 minggu atau lebih. Gejala lain yang sering muncul meliputi penurunan nafsu makan, berkurangnya berat badan, demam, sesak nafas, nyeri pada dada serta keringat berlebih pada malam hari (Dewi, 2020).

Penularan penyakit TBC dapat terjadi dengan mudah, terutama di daerah dengan mobilitas penduduk yang tinggi serta kualitas kesehatan hidup yang rendah. Selain aspek kepadatan dan persebaran penduduk, kondisi lingkungan juga berperan besar dalam penyebaran bakteri, terutama pada lingkungan tempat yang belum memenuhi standar kesehatan (Girsang dan Tobing, 2010 ; Nuraini, dkk, 2022). Pengaruh dari berbagai faktor risiko tersebut terbukti dengan masih tingginya jumlah kasus TBC ditingkat global. Menurut *World Health Organization* (2023), tercatat sekitar 1,25 juta kematian yang disebabkan penyakit TBC. Pada tahun 2023 tercatat sekitar 10,8 juta penduduk dunia terdiagnosis TBC dengan kasus 6 juta laki-laki, 3,6 juta perempuan dan 1,3 juta anak-anak.

Menurut laporan WHO (2024), Indonesia merupakan negara dengan beban kasus TBC terbanyak di dunia kedua setelah India. Provinsi Sumatera Barat termasuk dalam sepuluh besar provinsi dengan jumlah kasus TBC terbanyak di Indonesia. Badan Pusat Statistik (2024), melaporkan bahwa penemuan kasus TBC di Provinsi Sumatera Barat mengalami peningkatan jumlah kasus setiap tahunnya dari tahun 2021 sampai tahun 2024. Peningkatan yang signifikan terjadi pada tahun 2022 dengan penemuan kasus TBC mencapai 47,8% dan meningkat menjadi 61% pada tahun 2023.

Peningkatan yang terus berlangsung ini menunjukkan bahwa permasalahan tersebut memerlukan perhatian dan penanganan serius dengan melakukan pemeriksaan dan pengelolaan kasus yang lebih efektif.

Rumah Sakit Paru Sumatera Barat merupakan rumah sakit rujukan utama penyakit paru-paru di Provinsi Sumatera Barat yang hadir sebagai solusi layanan kesehatan terpercaya untuk memberikan perawatan kesehatan terbaik khususnya penyakit paru-paru, terutama TBC di Provinsi Sumatera Barat. Sebagai rumah sakit rujukan, Rumah Sakit Paru Sumatera Barat menerima pasien dari berbagai kabupaten dan kota di Provinsi Sumatera Barat, sehingga rumah sakit ini memiliki data pasien dengan karakteristik yang beragam dan dianggap dapat mewakili kondisi kasus TBC di Provinsi Sumatera Barat.

Dalam konteks penanganan penyakit TBC, *data mining* dapat dimanfaatkan sebagai salah satu metode untuk mengolah serta menganalisis data klinis pasien. Menurut Gorunescu (2011), *data mining* dapat diartikan sebagai suatu tahapan sistematis untuk menggali pengetahuan, mengidentifikasi pola, serta menemukan kecenderungan tertentu dari data dalam jumlah besar yang tersimpan pada media penyimpanan, dengan dukungan pendekatan statistik dan matematis. Klasifikasi merupakan metode analisis yang bertujuan membentuk pola atau fungsi tertentu guna mempresentasikan perbedaan antar kelas data, sekaligus digunakan untuk memperkirakan kelas dari data yang telah memiliki label (Setio dkk., 2020).

Penerapan klasifikasi dalam bidang kesehatan menjadi sangat penting khususnya pada penyakit menular seperti TBC. Melalui proses klasifikasi, sistem dapat membantu membedakan antara pasien yang menderita TBC dan pasien yang tidak menderita TBC. Hand dkk. (2001), menjelaskan bahwa salah satu metode klasifikasi adalah pohon keputusan. Pohon keputusan merupakan model berbentuk diagram bercabang yang menyerupai struktur pohon, struktur pohon tersebut berupa *internal node* yang berperan sebagai akar pohon, *internal node* menunjukkan hasil pengujian, dan *leaf node* mempresentasikan kelas tujuan. Salah satu metode yang umum digunakan untuk membangun model pohon keputusan adalah algoritma C5.0 yang merupakan penyempurnaan dari algoritma C4.5 yang sebelumnya juga merupakan hasil pengembangan dari ID3. Sebagai bagian dari metode *decisin tree*, algoritma C5.0 memiliki kemampuan dalam mengolah data menjadi seperangkat aturan yang dapat dimanfaatkan sebagai landasan dalam proses pengambilan keputusan (Pratiwi dkk. 2020 ; Amalda dkk. 2022).

Pada penelitian terdahulu yang dilakukan oleh Yogi (2007), dilakukan perbandingan tiga metode klasifikasi yaitu algoritma C5.0, CART dan CHAID. Hasil penelitian tersebut menunjukkan bahwa algoritma C5.0 menghasilkan tingkat akurasi tertinggi yaitu sebesar 87,72%, dibandingkan dengan CART sebesar 87,28% dan CHAID sebesar 87,15%. Penelitian lain yang dilakukan oleh Sofyan dkk. (2023), mengenai penerapan *data mining* menggunakan algoritma C5.0 dalam memprediksi penyakit stroke menunjukkan bahwa algoritma C5.0 menghasilkan akurasi yang cukup tinggi yaitu sebesar 95%.

Hasil penelitian terdahulu menunjukkan bahwa algoritma C5.0 dapat bekerja dengan baik ketika digunakan untuk mengklasifikasikan data medis dengan hasil akurasi yang tinggi. Dengan tingkat akurasi yang tinggi, algoritma C5.0 dinilai sesuai dan relevan untuk diterapkan dalam penelitian yang berfokus pada klasifikasi penyakit TBC. Berdasarkan pertimbangan tersebut, penelitian ini dilakukan untuk mengklasifikasikan penyakit TBC di Rumah Sakit Paru Sumatera Barat menggunakan algoritma C5.0.

II. METODE PENELITIAN

A. Sumber Data dan Variabel Penelitian

Pada penelitian ini data yang digunakan merupakan data sekunder, dimana data tersebut diperoleh dari Rekam Medis pasien di Rumah Sakit Paru Sumatera Barat pada bulan Oktober 2024 sampai bulan Desember 2024. Data tersebut mencakup berbagai informasi pasien seperti identitas, hasil pemeriksaan klinis, dan diagnosis akhir. Jumlah keseluruhan data yang digunakan sebanyak 150 data pasien. Penelitian ini melibatkan delapan variabel prediktor dan satu variabel target dengan rincian variabel terlihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Keterangan Kategori
(X_1) Jenis Batuk	0 : Batuk Berdahak 1 : Batuk Berdarah 2 : Batuk Kering
(X_2) Lama Batuk	0 : ≥ 3 Bulan 1 : 1 Bulan – < 3 Bulan 2 : < 1 Bulan
(X_3) Penurunan Berat Badan	0 : Ya 1 : Tidak

(X ₄) Nyeri Dada	0 : Ya 1 : Tidak
(X ₅) Sesak Nafas	0 : Ya 1 : Tidak
(X ₆) Keringat Malam	0 : Ya 1 : Tidak
(X ₇) Demam	0 : Ya 1 : Tidak
(Y) Hasil Pemeriksaan BTA	0 : Positif 1 : Negatif

B. Algoritma C5.0

Salah satu metode klasifikasi dalam *data mining* yang secara khusus digunakan pada teknik *decision tree* yaitu Algoritma C5.0. Algoritma C5.0 merupakan algoritma yang dikembangkan sebagai metode lanjutan dari algoritma sebelumnya yang telah dibentuk pada tahun 1987 oleh Ross Quinlan, yaitu C4.5 dan ID3. Perbedaan utama antara algoritma C5.0 dan C4.5 yang dapat dilihat dari struktur pohon keputusan yang dihasilkan, dimana C5.0 mampu menghasilkan *decision tree* yang lebih sederhana dan ringkas dibandingkan dengan Algoritma C4.5. Selain itu, proses pembentukan pohon pada algoritma C5.0 juga dikenal lebih efisien dan cepat dibandingkan algoritma C4.5 (Dunhan, 2018).

Menurut Pratiwi dkk. (2020), keunggulan utama dari Algoritma C5.0 terletak pada kemampuan algoritma tersebut untuk melakukan penghitungan *gain ratio* setelah nilai *gain* diperoleh. Proses pembentukan pohon dimulai dari perhitungan nilai *entropy*, kemudian melakukan perhitungan *gain* dan *gain ratio*. Berbeda dengan ID3 dan C4.5 yang hanya mempertimbangkan nilai *gain*, algoritma C5.0 memungkinkan melakukan penentuan atribut secara lebih optimal dengan mempertimbangkan nilai *gain ratio*. Dengan demikian, atribut yang memiliki nilai terbaik dapat dipilih secara lebih akurat dalam pembentukan pohon keputusan. Adapun rumus untuk menghitung nilai *entropy* sebagai berikut:

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

dengan:

S : Himpunan kasus

n : Jumlah pembagian S

p_i : Jumlah perbandingan seluruh atribut (S_i) pada keseluruhan total (S)

Kemudian melakukan perhitungan nilai *gain* dilakukan menggunakan rumus berikut:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} \times Entropy(S_i) \quad (2)$$

dengan:

$Entropy(S)$: Himpunan kasus

$Entropy(S_i)$: Jumlah pembagian S

A : Atribut

n : Jumlah pembagian S

$|S_i|$: Jumlah kasus atribut

$|S|$: Jumlah kasus pada S

Ketika telah memperoleh nilai *entropy* dan *gain*, tahap selanjutnya adalah menghitung nilai *gain ratio*. Perhitungan *gain ratio* dilakukan untuk mendapatkan atribut terbaik yang akan dijadikan simpul pada pohon keputusan. Rumus untuk menghitung nilai *gain ratio* ini adalah sebagai berikut:

$$Gain Ratio = \frac{Gain(S, A)}{\sum_{i=1}^n Entropy(S_i)} \quad (3)$$

(Pratiwi dkk, 2020)

dengan:

$Gain(S, A)$: Nilai *gain* dari suatu atribut

$\sum_{i=1}^n Entropy(S_i)$: Jumlah nilai *entropy* dalam suatu atribut

Kemudian, proses diulang untuk masing-masing cabang hingga semua kelas pada cabang mendapatkan kelasnya masing-masing.

C. Confusion Matrix

Salah satu metode evaluasi yang umum digunakan untuk menilai kinerja model klasifikasi adalah *Confusion matrix*. Dimana, *Confusion matrix* menyajikan informasi mengenai hasil prediksi model dalam bentuk tabel dua dimensi yang membandingkan antara kelas aktual dengan kelas hasil prediksi. Struktur *confusion matrix* dapat dilihat pada Tabel 2. (Zaki dan Meira 2020).

Tabel 2. Confusion Matrix

Actual Class	Predicted Class	
	Positive (c1)	Negative (c2)
Positive (C ₁)	True Positive (TP)	False Negatif (FN)
Negative (C ₂)	False Positive (FP)	True Negatif (TN)

Keterangan:

- Actual class* : Kelas yang sebenarnya
Predict class : Kelas yang diperoleh dari hasil prediksi pada model klasifikasi
 (TP) : Jumlah data positif yang berhasil diprediksi sebagai positif
 (TN) : Jumlah data negatif yang berhasil diprediksi sebagai negatif
 (FP) : Jumlah data negatif yang salah diprediksi sebagai positif
 (FN) : Jumlah data positif yang salah diprediksi sebagai negatif

Berdasarkan komponen di atas, berbagai ukuran kinerja klasifikasi dapat diturunkan, di antaranya *accuracy*, *precision*, *sensitivity*, *spesificity* berdasarkan rumus berikut:

$$Accuracy = \frac{TP+TN}{TP+FN+FP+TN} \times 100\% \quad (4)$$

$$Precision = \frac{TP}{TP+FP} \times 100\% \quad (5)$$

$$Sensitivity = \frac{TP}{TP+FN} \times 100\% \quad (6)$$

$$Specificity = \frac{TN}{TN+FP} \times 100\% \quad (7)$$

Sumber: (Zaki dan Meira 2020)

D. Teknik Analisis Data

Pada penelitian ini metode yang diterapkan adalah metode klasifikasi dengan memanfaatkan metode *decision tree* algoritma C5.0. Tahapan analisis data yang dilakukan sebagai berikut:

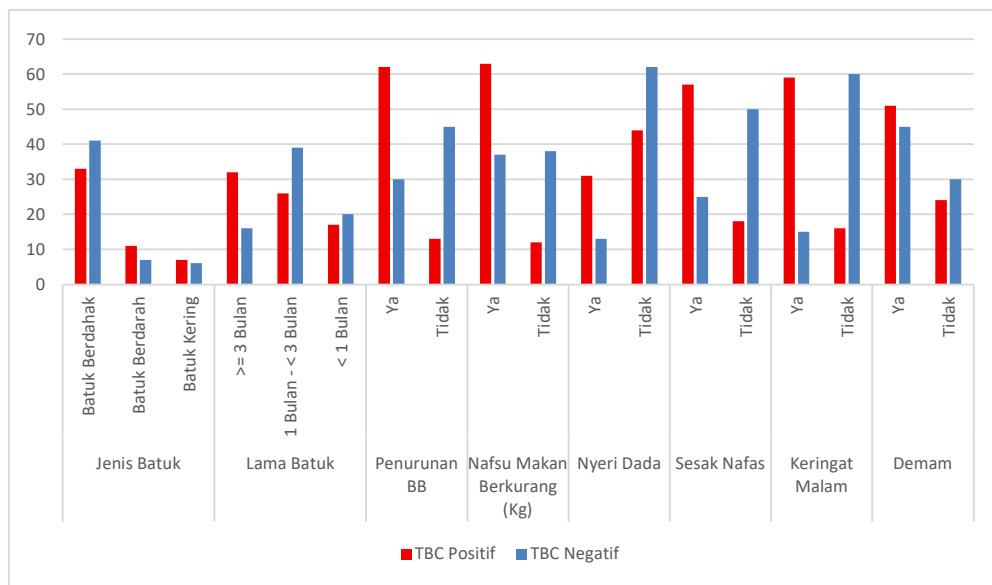
1. Melakukan input data
2. Melaksanakan tahap *preprocessing* data, yang meliputi pembersihan data serta pengelompokan nilai variabel ke dalam kategori tertentu.
3. Membagi dataset menjadi data *training* dan data *testing* dengan proporsi 70% untuk data *training* dan 30% data *testing*.
4. Melakukan proses klasifikasi menggunakan algoritma C5.0 dengan langkah-langkah berikut:
 - a. Menghitung nilai *entropy* untuk masing-masing variabel menggunakan persamaan (1).
 - b. Menghitung nilai *gain* pada masing-masing variabel menggunakan persamaan (2).
 - c. Menghitung nilai *gain ratio* setiap variabel menggunakan persamaan (3).
 - d. Menentukan simpul akar dan cabang pohon berdasarkan nilai *gain ratio* tertinggi.
 - e. Mengulangi proses perhitungan pada langkah sebelumnya hingga seluruh variabel yang digunakan membentuk struktur pohon klasifikasi.
 - f. Mengevaluasi kinerja model dengan *confusion matrix* untuk menghitung nilai *accuracy*, *precision*, *sensitivity*, *spesificity* menggunakan persamaan (4), (5), (6), (7).
5. Menetapkan hasil akhir sebagai dasar pengambilan keputusan.

III. HASIL DAN PEMBAHASAN

A. Eksplorasi Data

Eksplorasi data bertujuan untuk memberikan gambaran distribusi berbagai gejala klinis pasien berdasarkan status TBC positif dan negatif. Visualisasi ini bertujuan untuk menunjukkan perbedaan karakteristik klinis antara kedua kelompok pasien serta memberikan pemahaman awal mengenai pola hubungan antara gejala yang dialami pasien

dengan status TBC. Penjelasan karakteristik pasien yang mengalami TBC positif dan negatif di Rumah Sakit Paru Sumatera Barat terdapat di Gambar 1.

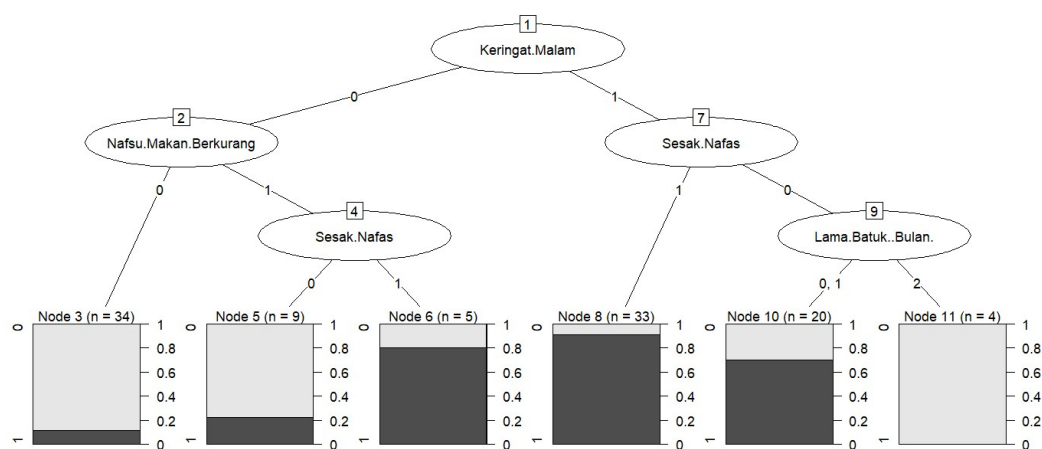


Gambar 1. Hubungan Gejala Klinis dengan Status TBC

Gambar 1 menunjukkan bahwa jenis batuk berdahak merupakan gejala paling dominan dan banyak dialami oleh pasien, baik pada kelompok pasien TBC positif maupun kelompok pasien TBC negatif. Pada gejala lama batuk kelompok pasien TBC positif lebih dominan mengalami durasi lama yaitu ≥ 3 Bulan. Gejala yang lain seperti penurunan BB, nafsu makan berkurang (kg), sesak nafas, keringat malam, dan demam lebih sering ditemukan pada kelompok pasien TBC positif.

B. Hasil Analisis Algoritma C5.0

Berdasarkan analisis yang dilakukan menggunakan algoritma C5.0 didapatkan hasil berupa pohon keputusan (*decision tree*) serta evaluasi model yang meliputi nilai *accuracy*, *precision*, *sensitivity*, dan *specificity*. Pohon keputusan yang dihasilkan menggambarkan struktur aturan klasifikasi yang digunakan oleh model dalam menentukan status TBC pasien berdasarkan variabel gejala klinis, sehingga memudahkan interpretasi hasil klasifikasi. Pohon keputusan tersebut disajikan pada Gambar 2.



Gambar 2. Pohon Keputusan Algoritma C5.0

Berdasarkan Gambar 2 pohon keputusan yang dibentuk menggunakan algoritma C5.0 menampilkan 1 *root node* dengan jenis variabel keringat malam, 4 *internal node* dan 6 *leaf node* sebagai keputusan akhir. Variabel keringat malam menjadi *root node* yang menandakan variabel ini memiliki *gain ratio* tertinggi, setiap *internal node* menjelaskan proses pengambilan keputusan lanjutan, dan *leaf node* ditampilkan berupa balok yang menunjukkan proporsi kelompok pasien TBC positif dan kelompok pasien TBC negatif. Dari hasil pembentukan pohon keputusan algoritma C5.0, jika pasien mengalami gejala keringat malam dan nafsu makan berkurang maka pasien dinyatakan TBC positif.

Evaluasi model yang dilakukan menggunakan *confusion matrix*. Hasil evaluasi model dapat dilihat pada Tabel 3.

Tabel 3. Hasil Perhitungan *Confusion Matrix*

<i>Actual Class</i>	<i>Predicted Class</i>	
	TBC Positif	TBC Negatif
TBC Positif	23	3
TBC Negatif	1	18

Berdasarkan Tabel 3 hasil perhitungan *confusion matrix* yang dihasilkan yaitu jumlah pasien TBC positif yang juga diprediksi TBC positif adalah 23 pasien, sedangkan pasien TBC negatif yang diprediksi TBC negatif adalah 18 pasien. Hasil perhitungan *accuracy*, *precision*, *sensitivity*, dan *specificity* dapat dihitung dengan persamaan (4), (5), (6), (7). Hasil perhitungan disajikan pada Tabel 4.

Tabel 4. Evaluasi Model dengan *Confusion Matrix*

Evaluasi	Nilai
<i>Accuracy</i>	0,9111
<i>Precision</i>	0,9583
<i>Sensitivity</i>	0,8846
<i>Specificity</i>	0,9474

Hasil evaluasi model klasifikasi algoritma C5.0 mendapatkan hasil kinerja sangat baik. Nilai *accuracy* menunjukkan bahwa model dapat mengklasifikasikan status TBC pasien dengan tingkat ketepatan sebesar 91,11%. Nilai *precision* sebesar 95,83% mengindikasikan bahwa sebagian besar pasien yang diprediksi sebagai TBC positif benar-benar merupakan kasus TBC positif, sehingga tingkat kesalahan prediksi positif relatif rendah. Selanjutnya, nilai *sensitivity* sebesar 88,46% menandakan bahwa model memiliki kemampuan yang baik dalam mendeteksi pasien TBC positif, meskipun masih terdapat sebagian kecil kasus positif yang tidak teridentifikasi. Sementara itu, nilai *specificity* sebesar 94,74% menunjukkan bahwa model sangat efektif dalam mengidentifikasi pasien yang tidak menderita TBC. Secara keseluruhan, hasil evaluasi membuktikan algoritma C5.0 mampu menghasilkan performa klasifikasi yang andal dan seimbang dalam membedakan pasien TBC positif dan negatif.

Hasil penelitian ini sejalan dengan penelitian lain yang menerapkan algoritma C5.0 pada kasus kesehatan yang berbeda. Penelitian yang dilakukan oleh Sahputra dkk. (2023), menerapkan algoritma C5.0 pada klasifikasi status gizi ibu hamil di Kota Lhokseumawe, hasil penelitian ini juga menunjukkan performa yang baik dengan nilai *accuracy* sebesar 94%. Hal ini mengindikasikan bahwa algoritma C5.0 memiliki kemampuan klasifikasi yang stabil dan andal pada berbagai permasalahan di bidang kesehatan.

IV. KESIMPULAN

Berdasarkan penelitian yang telah dilakukan, didapatkan hasil algoritma C5.0 dapat mengklasifikasikan status penyakit Tuberculosis (TBC) pada pasien Rumah Sakit Paru Sumatera Barat dengan kinerja yang sangat baik. Model klasifikasi yang dibangun menggunakan 8 variabel gejala klinis menghasilkan tingkat akurasi sebesar 91,11% yang menunjukkan kemampuan model dalam mengklasifikasikan data pasien secara tepat. Hasil evaluasi juga menunjukkan nilai *precision* sebesar 95,83%, *sensitivity* sebesar 88,46%, dan *specificity* sebesar 94,74%, yang menandakan bahwa model tidak hanya efektif dalam mendeteksi pasien TBC positif, namun juga sangat baik untuk mengidentifikasi pasien yang tidak menderita TBC. Variabel keringat malam terbukti menjadi faktor paling dominan dalam pembentukan pohon keputusan, diikuti oleh gejala klinis lainnya seperti nafsu makan berkurang, sesak napas, dan lama batuk. Dengan demikian, algoritma C5.0 dapat dijadikan sebagai metode pendukung pengambilan keputusan yang andal dalam proses klasifikasi dan skrining awal penyakit TBC berbasis data klinis pasien.

DAFTAR PUSTAKA

- Amalda, R. N., Millah, N., & Fitria, I. (2022). Implementasi Algoritma C5.0 Dalam Menganalisa Kelayakan Penerima Keringanan UKT Mahasiswa ITK. *Teorema: Teori Dan Riset Matematika*, 7(1), 101–116.
- Badan Pusat Statistik. (2024). *Kasus Penyakit Menurut Kabupaten/Kota dan Jenis Penyakit di Provinsi Sumatera Barat, 2024*. Badan Pusat Statistik.
- Dewi, L. P. K. (2020). Pemeriksaan Basil Tahan Asam Untuk Membantu Menegakkan Diagnosis Penyakit Tuberkulosis. *International Journal of Applied Chemistry Research*, 1(1), 16. <https://doi.org/10.23887/ijacr.v1i1.28716>
- Dunhan, M. H. (2018). *Data Mining: Introductory and Advanced Topics* (1st Edition). Pearson Education.
- Febriyanto, D., & Kurniawan, Y. I. (2018). Prediksi Penyakit Tuberculosis (TBC) Menggunakan Algoritma C4.5. *Jurnal Ilmiah Sinus (JIS)*, 16(2), 23–36.
- Girsang, M., & Tobing, K. (2010). Karakteristik Demografis dan Hubungannya dengan Penyakit Tuberculosis di Provinsi Jawa Tengah (Analisis Lanjut Riskesdas 2007). *Media Penelitian Kesehatan*, 20.
- Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Springer
- Hand, D., Mannila, H. dan Smyth, P. (2001). *Principles of Data Minig*. MIT Press.
- Nuraini, N., Suhartono, S., & Raharjo, M. (2022). Hubungan Faktor Lingkungan Fisik Dalam Rumah dan Perilaku Kesehatan dengan Kejadian TB Paru di Purwokerto Selatan Banyumas. *Jurnal Kesehatan Lingkungan Indonesia*, 21(2), 210–218. <https://doi.org/10.14710/jkli.21.2.210-218>
- Pratiwi, R., Hayati, M. N., & Prangga, S. (2020). Perbandingan Klasifikasi Algoritma C5.0 Dengan Classification and Regression Tree (Studi Kasus : Data Sosial Kepala Keluarga Masyarakat Desa Teluk Baru Kecamatan Muara Ancalong Tahun 2019). *BAREKENG: Jurnal Ilmu Matematika Dan Terapan*, 14(2), 273–284. <https://doi.org/10.30598/barekengvol14iss2pp273-284>
- Sahputra, I., Mauliza, & Zohra, S. F. A. (2023). Implementasi Algoritma C5 . 0 Pada Klasifikasi Status Gizi Ibu Hamil di Kota Lhokseumawe. *Jurnal Metik*, 7(1). <https://doi.org/10.47002/metik.v7i1.562>
- Setio, P. B. N., Saputro, D. R. S., & Bowo Winarno. (2020). Klasifikasi Dengan Pohon Keputusan Berbasis Algoritme C4.5. *PRISMA, Prosiding Seminar Nasional Matematika*, 3, 64–71.
- Sofyan, F. M. A., Riyandoro, A. P., Maulana, D. F., & Jaman, J. H. (2023). Penerapan Data Mining dengan Algoritma C5.0 Untuk Prediksi Penyakit Stroke. *Jurnal Teknologi Sistem Informasi Dan Sistem Komputer TGD*, 6(2), 619–625.
- World Health Organization. (2024). *Global tuberculosis report 2024*.
- World Health Organization. (2025). *Tuberculosis - World Health Organization (WHO)*. World Health Organization. <https://www.who.int/en/news-room/fact-sheets/detail/tuberculosis>
- Yogi Yusuf, W. (2007). Perbandingan Performasi Algoritma Decision Tree C5.0, CART, dan CHAID: Kasus Prediksi Status Resiko Kredit di Bank X. *Seminar Nasional Aplikasi Teknologi Informasi 2007 (SNATI)*, 0–3.
- Zaki, M. J., & Meira, W. J. (2020). *Data Mining and Machine Learning Fundamental Concepts and Algorithms* (2nd ed.). Cambridge University Press.