

**KLASIFIKASI PENYAKIT *TUBERCULOSIS* DI RUMAH SAKIT
PARU SUMATERA BARAT MENGGUNAKAN ALGORITMA C5.0**



**Oleh
MELIANI MAYA SARI
NIM. 21337080**

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG
2026**

**KLASIFIKASI PENYAKIT *TUBERCULOSIS* DI RUMAH SAKIT
PARU SUMATERA BARAT MENGGUNAKAN ALGORITMA C5.0**

SKRIPSI

*Diajukan sebagai salah satu persyaratan memperoleh gelar
Sarjana Statistika*



Oleh

MELIANI MAYA SARI

NIM. 21337080

**PROGRAM STUDI SARJANA STATISTIKA
DEPARTEMEN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
UNIVERSITAS NEGERI PADANG**

2026

PERSETUJUAN SKRIPSI

KLASIFIKASI PENYAKIT *TUBERCULOSIS* DI RUMAH SAKIT PARU SUMATERA BARAT MENGGUNAKAN ALGORITMA C5.0

Nama : Meliani Maya Sari
NIM : 21337080
Program Studi : S1 Statistika
Jurusan : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Mengetahui:
Kepala Departemen Statistika



Dr. Yenni Kurniawati, S.Si., M.Si.
NIP. 198402232010122005

Padang, 3 Februari 2026
Disetujui Oleh:
Pembimbing



Zilrahmi, S.Pd., M.Si.
NIP. 198911062019032009

PENGESAHAN LULUS UJIAN SKRIPSI

Nama : Meliani Maya Sari
NIM : 21337080
Program Studi : S1 Statistika
Jurusan : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

KLASIFIKASI PENYAKIT *TUBERCULOSIS* DI RUMAH SAKIT PARU SUMATERA BARAT MENGGUNAKAN ALGORITMA C5.0

Dinyatakan lulus setelah dipertahankan di depan Tim Penguji Skripsi
Departemen Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Negeri Padang

Padang, 3 Februari 2026

Tim Penguji	Nama
Ketua	: Zilrahmi, S.Pd., M.Si.
Anggota	: Dr. Dony Permana, M.Si.
Anggota	: Dwi Sulistiowati, S.Si., M.Stat.

Tanda Tangan



SURAT PERNYATAAN TIDAK PLAGIAT

Saya yang bertanda tangan di bawah ini :

Nama : Meliani Maya Sari
NIM : 21337080
Program Studi : S1 Statistika
Jurusan : Statistika
Fakultas : Matematika dan Ilmu Pengetahuan Alam

Dengan ini menyatakan bahwa, skripsi saya dengan judul “**Klasifikasi Penyakit Tuberculosis di Rumah Sakit Paru Sumatera Barat Menggunakan Algoritma C5.0**” adalah benar merupakan hasil karya saya dan bukan merupakan plagiat dari karya orang lain atau pengutipan cara-cara yang tidak sesuai dengan etika yang berlaku dalam tradisi keilmuan. Apabila suatu saat terbukti saya melakukan plagiat maka saya bersedia diproses dan menerima sanksi akademis maupun hukum sesuai dengan hukum dan ketentuan yang berlaku, baik di institusi UNP maupun di Masyarakat dan negara.

Demikian pernyataan ini saya buat dengan penuh kesadaran dan rasa tanggung jawab sebagai anggota masyarakat ilmiah

Diketahui Oleh,
Ketua Departemen Statistika



Dr. Yenni Kurniawati, S.Si., M.Si.
NIP. 198402232010122005

Saya yang menyatakan,



Meliani Maya Sari
NIM. 21337080

Klasifikasi Penyakit *Tuberculosis* di Rumah Sakit Paru Sumatera Barat Menggunakan Algoritma C5.0

Meliani Maya Sari

ABSTRAK

Tuberculosis (TBC) merupakan penyakit menular yang masih menjadi permasalahan kesehatan utama di Indonesia, termasuk di Provinsi Sumatera Barat. Proses diagnosis TBC umumnya dilakukan melalui pemeriksaan Basil Tahan Asam (BTA) yang memerlukan waktu dan ketelitian tinggi, sehingga berpotensi memperlambat penanganan pasien. Oleh karena itu, diperlukan metode alternatif yang dapat membantu proses identifikasi awal penyakit TBC secara lebih cepat dan akurat berdasarkan data klinis pasien. Penelitian ini bertujuan untuk mengklasifikasikan penyakit TBC di Rumah Sakit Paru Sumatera Barat serta mengevaluasi kinerja algoritma C5.0 dalam melakukan klasifikasi tersebut.

Metode penelitian yang digunakan adalah metode *data mining* dengan teknik klasifikasi menggunakan algoritma C5.0. Data yang digunakan berupa data sekunder rekam medis pasien di Rumah Sakit Paru Sumatera Barat sebanyak 150 data pada periode Oktober hingga Desember 2024. Variabel penelitian terdiri dari delapan variabel prediktor, yaitu jenis batuk, lama batuk, nafsu makan berkurang, penurunan berat badan, nyeri dada, sesak napas, keringat malam, dan demam, serta satu variabel target yaitu hasil pemeriksaan BTA. Data melalui tahap *preprocessing*, kemudian dibagi menjadi data *training* dan data *testing*, selanjutnya dilakukan pembentukan model klasifikasi dan evaluasi menggunakan confusion matrix.

Hasil penelitian menunjukkan bahwa algoritma C5.0 mampu mengklasifikasikan penyakit TBC dengan kinerja yang baik. Model yang dihasilkan memperoleh nilai akurasi sebesar 91,11%, presisi sebesar 95,83%, sensitivitas sebesar 88,46%, dan spesifisitas sebesar 94,74%. Pohon keputusan yang dihasilkan bersifat sederhana dan mudah diinterpretasikan, sehingga dapat membantu tenaga medis dalam proses pengambilan keputusan. Berdasarkan hasil tersebut, algoritma C5.0 dapat digunakan sebagai metode pendukung dalam identifikasi awal penyakit TBC di Rumah Sakit Paru Sumatera Barat.

Kata kunci: *Tuberculosis*, Klasifikasi, Pohon Keputusan, Algoritma C5.0.

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

Meliani Maya Sari

ABSTRACT

Tuberculosis (TB) is an infectious disease that remains a major health problem in Indonesia, including in West Sumatra Province. The TB diagnosis process is generally carried out through acid-fast bacilli (AFB) testing, which requires time and high accuracy, thereby potentially delaying patient treatment. Therefore, an alternative method is needed to help identify TB more quickly and accurately based on patient clinical data. This study aims to classify TB cases at the Rumah Sakit Paru Sumatera Barat and evaluate the performance of the C5.0 algorithm in performing this classification.

The research method used is data mining with classification techniques using the C5.0 algorithm. The data used is secondary data from patient medical records at the West Sumatra Lung Hospital, consisting of 150 data points from the period October to December 2024. The research variables consisted of eight predictor variables, namely type of cough, duration of cough, decreased appetite, weight loss, chest pain, shortness of breath, night sweats, and fever, as well as one target variable, namely the BTA examination results. The data went through a preprocessing stage, then was divided into training data and testing data, followed by the formation of a classification model and evaluation using a confusion matrix.

The results showed that the C5.0 algorithm was able to classify TB with good performance. The resulting model achieved an accuracy of 91.11%, precision of 95.83%, sensitivity of 88.46%, and specificity of 94.74%. The resulting decision tree is simple and easy to interpret, thus assisting medical personnel in the decision-making process. Based on these results, the C5.0 algorithm can be used as a supporting method in the early identification of tuberculosis at the West Sumatra Lung Hospital.

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

KATA PENGANTAR

Alhamdulillah puji dan syukur penulis panjatkan ke hadirat Allah SWT. karena atas rahmat dan karunia-Nya penulis dapat menyelesaikan penyusunan skripsi yang berjudul “Klasifikasi Penyakit *Tuberculosis* di Rumah Sakit Paru Sumatera Barat Menggunakan Algoritma C5.0”. Skripsi ini disusun sebagai salah satu syarat untuk menyelesaikan pendidikan pada Program Studi Sarjana Statistika, Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Negeri Padang, serta sebagai bentuk penerapan ilmu yang telah diperoleh selama masa perkuliahan.

Selama proses menyelesaikan skripsi ini tidak lepas dari dukungan, bimbingan, dan bantuan berbagai pihak. Oleh karena itu, penulis ingin menyampaikan terima kasih kepada:

1. Ibu Zilrahmi, S.Pd., M.Si., sebagai Dosen Pembimbing Skripsi dan Penasehat Akademik yang telah membimbing dan memberikan arahan selama masa perkuliahan dan penyusunan skripsi ini
2. Bapak Dr. Dony Permana, M.Si., dan Ibu Dwi Sulistiowati, S.Si., M.Stat., selaku dosen pembahas skripsi yang telah memberikan arahan dan masukan dalam penyusunan skripsi ini.
3. Ibu Dr. Yenni Kurniawati, S.Si., M.Si., sebagai Kepala Departemen Statistika sekaligus Koordinator Program Studi S1 Statistika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Negeri Padang.
4. Bapak/Ibu Dosen dan Tenaga Kependidikan Departemen Statistika yang telah memberikan ilmu dan motivasi selama perkuliahan.
5. Kedua orang tua tercinta, Bapak Ujang dan Ibu Arlimah, serta Kakak laki-laki penulis, Edo Deskia Putra yang dengan penuh kesabaran senantiasa memberikan

do'a, kasih sayang, perhatian, dan dukungan yang tak ternilai selama penulis menempuh pendidikan hingga penyusunan skripsi ini dapat diselesaikan.

6. Teman-teman seperkuliah Departemen Statistika, khususnya Angkatan 2021 yang telah memberikan semangat dan dukungan selama masa perkuliahan.

Semoga semua bimbingan, arahan, dukungan, bantuan dan do'a yang telah diberikan dibalas oleh Allah SWT dengan pahala yang berlipat. Penulis menyadari bahwa skripsi ini masih memiliki keterbatasan, sehingga kritik dan saran yang membangun sangat diharapkan demi penyempurnaan di masa mendatang. Penulis berharap skripsi ini dapat memberikan manfaat bagi pengembangan ilmu pengetahuan serta menjadi referensi bagi pihak-pihak yang berkepentingan, khususnya dalam bidang yang berkaitan dengan klasifikasi penyakit TBC.

Padang, 3 Januari 2026
Penulis,

Meliani Maya Sari

DAFTAR ISI

ABSTRAK	i
ABSTRACT	ii
KATA PENGANTAR.....	iii
DAFTAR ISI.....	v
DAFTAR TABEL	vii
DAFTAR GAMBAR	viii
DAFTAR LAMPIRAN	ix
BAB I PENDAHULUAN.....	1
A. Latar Belakang Masalah.....	1
B. Batasan Masalah.....	8
C. Rumusan Masalah	8
D. Tujuan Penelitian.....	8
E. Manfaat Penelitian	9
BAB II KERANGKA TEORITIS.....	10
A. Kajian Teori.....	10
1. Tuberculosis	10
2. Data Mining	15
3. Klasifikasi	15
4. Decision Tree	17
5. Algoritma C5.0.....	19
B. Penelitian Relevan.....	23
BAB III METODOLOGI PENELITIAN	26
A. Jenis Penelitian.....	26
B. Jenis dan Sumber Data	26
C. Variabel Penelitian	27
D. Struktur Data	27
E. Teknik Analisis Data	28
BAB IV HASIL DAN PEMBAHASAN.....	30
A. Eksplorasi Data	30
B. Hasil Analisis	37
C. Pembahasan.....	44
BAB V PENUTUP	47

A. Kesimpulan	47
B. Saran.....	47
DAFTAR PUSTAKA.....	49
LAMPIRAN.....	52

DAFTAR TABEL

Tabel	Halaman
1 <i>Confusion Matrix</i>	21
2. Standar Evaluasi Kriteria <i>Confusion Matrix</i>	23
3. Variabel Penelitian	27
4. Struktur Data	27
5. Proporsi Data Latih dan Uji	38
6. <i>Gain Ratio</i> dengan Proporsi 70:30	39
7. <i>Confusion Matrix</i> Algoritma C5.0 dengan Proporsi 70:30	41
8. Hasil Evaluasi Model Pada setiap Proporsi Data	42

DAFTAR GAMBAR

Gambar	Halaman
1. Jumlah Kasus TBC di Provinsi Sumatera Barat (%)	3
2. Bakteri <i>Mycobacterium Tuberculosis</i>	11
3. Pohon Keputusan	18
4. Diagram Alir Penelitian.....	29
5. Pasien TBC di Rumah Sakit Paru Sumatera Barat.....	30
6. Pasien TBC Berdasarkan Jenis Batuk	31
7. Pasien TBC Berdasarkan Lama Batuk	32
8. Pasien TBC Berdasarkan Penurunan Berat Badan.....	33
9. Pasien TBC Berdasarkan Nafsu Makan Berkurang	34
10. Pasien TBC Berdasarkan Nyeri Dada	34
11. Pasien TBC Berdasarkan Sesak Napas	35
12. Pasien TBC Berdasarkan Keringat Malam	36
13. Pasien TBC Berdasarkan Demam	37
14. Pohon Keputusan Algoritma C5.0	43

DAFTAR LAMPIRAN

Lampiran	Halaman
1. Surat Izin Penelitian	52
2. Istilah dan Pengertiannya	53
3. <i>Syntax</i> Algoritma C5.0	54
4. Nilai <i>Gain Ratio</i> dengan Proporsi 70:30	55
5. Nilai <i>Gain Ratio</i> dengan Proporsi 75:25	56
6. Nilai <i>Gain Ratio</i> dengan Proporsi 80:20	57
7. Nilai <i>Gain Ratio</i> dengan Proporsi 85:15	58
8. Nilai <i>Gain Ratio</i> Untuk Cabang Selanjutnya dengan Proporsi 70:30	59
9. Pohon Keputusan Algoritma C5.0 dengan Proporsi 75:25	60
10. Pohon Keputusan Algoritma C5.0 dengan Proporsi 80:20	60
11. Pohon Keputusan Algoritma C5.0 dengan Proporsi 85:15	60

BAB I

PENDAHULUAN

A. Latar Belakang Masalah

Tuberculosis (TBC) merupakan penyakit yang disebabkan oleh bakteri *mycobacterium tuberculosis*. Penyakit ini bukanlah penyakit keturunan melainkan penyakit menular yang dapat menyebar dari satu orang ke orang lain melalui udara (Nadesul, H. 1995). Bakteri *mycobacterium tuberculosis* berbentuk batang dan mempunyai sifat khusus tahan terhadap asam. Oleh karena itu disebut sebagai Basil Tahan Asam (BTA). Sumber penularan penyakit TBC adalah penderita TBC yang mengalami batuk dan bersin dengan BTA positif. Saat batuk dan bersin, penderita TBC menyebarkan bakteri ke udara dalam bentuk *droplet* (percikan dahak). Infeksi bakteri TBC akan menimbulkan gejala umum yang dialami bagi penderita TBC seperti batuk terus menerus yang berlangsung selama 3 minggu atau lebih, nafsu makan berkurang, berat badan menurun, demam, sesak napas, nyeri dada dan berkeringat malam (Dewi, 2020).

Menurut WHO (2025), TBC merupakan penyebab utama kematian akibat satu agen infeksi tunggal di dunia, dengan jumlah kematian yang lebih tinggi dibandingkan penyakit infeksi utama lainnya seperti HIV/AIDS dan malaria berdasarkan data mortalitas global. Setelah periode pandemi COVID-19, TBC kembali menempati posisi teratas seiring dengan menurunnya dampak pandemi tersebut. Selain itu, TBC menjadi salah satu dari sepuluh penyebab utama kematian di dunia, yang menunjukkan bahwa beban penyakit ini signifikan tidak hanya dalam kelompok penyakit menular, tetapi juga dalam konteks kematian global secara keseluruhan. Pada tahun 2024 sebanyak 1,23 juta orang meninggal akibat TBC,

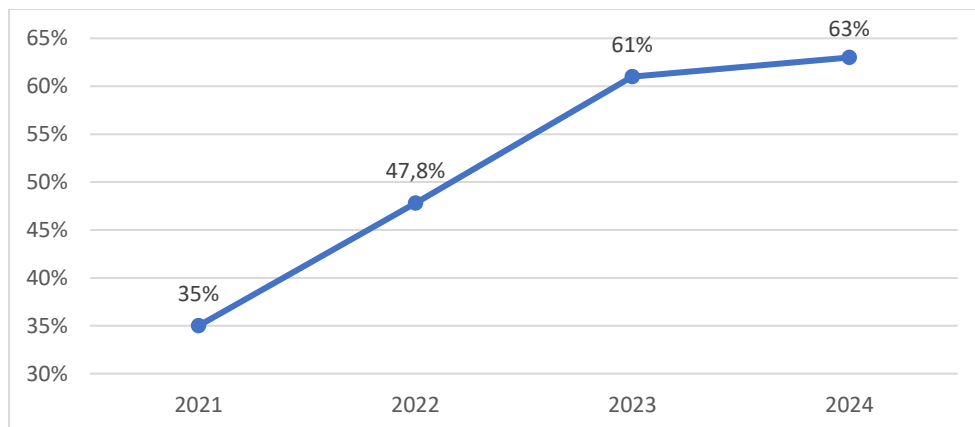
sekitar 10,7 juta orang di seluruh dunia jatuh sakit karena penyakit TBC dengan kasus 5,8 juta pria, 3,6 juta wanita dan 1,2 juta anak-anak.

Penularan penyakit TBC dapat terjadi dengan mudah, terutama di daerah dengan mobilitas penduduk yang tinggi serta kualitas kesehatan hidup yang rendah. Persebaran penduduk yang tidak merata juga menjadi faktor penting dalam tingginya angka keterpaparan terhadap penyakit TBC. Wilayah dengan kepadatan penduduk yang tinggi umumnya memiliki lingkungan tempat tinggal yang kurang layak, sanitasi yang buruk, serta asupan gizi yang rendah. Kondisi tersebut mempercepat penyebaran TBC apabila terdapat individu yang terinfeksi di lingkungan tersebut (Girsang dan Tobing, 2010).

Menurut Nuraini dkk. (2022), selain aspek kepadatan dan persebaran penduduk, faktor lingkungan memegang peran penting dalam penularan penyakit TBC, terutama lingkungan rumah yang tidak memenuhi standar kesehatan. Rumah yang tidak memenuhi syarat kesehatan, seperti kurangnya pencahayaan alami dari sinar matahari dapat menjadi media berkembangbiaknya bakteri *mycobacterium tuberculosis*. Cahaya matahari pagi sangat dibutuhkan untuk membunuh bakteri *patogen* di dalam rumah. Selain itu, kelembapan, suhu, pencahayaan, dan kepadatan hunian merupakan beberapa faktor risiko yang dapat meningkatkan kemungkinan penularan TBC di lingkungan tempat tinggal.

Berdasarkan laporan BPS (2024a), Provinsi Sumatera Barat masuk ke dalam lima besar provinsi dengan tingkat kepadatan penduduk tertinggi di Pulau Sumatera. Kondisi ini tentunya berpotensi meningkatkan intensitas interaksi antar masyarakat, sehingga dapat mempercepat penularan dan penyebaran penyakit TBC di Provinsi Sumatera Barat. Dampak dari kondisi tersebut tercermin dari tingginya

angka kasus TBC di Provinsi Sumatera Barat yang terus mengalami peningkatan setiap tahunnya, hal ini dapat dilihat pada Gambar 1.



Sumber: (Badan Pusat Statistik, 2024b)

Gambar 1. Persentase Kasus TBC di Provinsi Sumatera Barat

Gambar 1 menampilkan persentase kasus TBC di Provinsi Sumatera yang terus mengalami peningkatan di setiap tahunnya. Persentase ini merupakan perbandingan antara jumlah kasus TBC yang berhasil didiagnosis, diobati, dan dilaporkan melalui Program Penanggulangan TBC Nasional dengan estimasi jumlah kasus insiden TBC yang diperkirakan oleh WHO pada periode tertentu. Dari Gambar 1 terlihat terjadi peningkatan yang signifikan pada tahun 2021 hingga tahun 2023. Peningkatan yang signifikan ini menunjukkan bahwa permasalahan tersebut memerlukan perhatian dan penanganan serius dengan melakukan pemeriksaan dan pengelolaan kasus yang lebih efektif.

Rumah Sakit Paru Sumatera Barat hadir sebagai solusi layanan kesehatan terpercaya yang memberikan perawatan kesehatan terbaik khususnya penyakit paru-paru, terutama TBC di Provinsi Sumatera Barat. Rumah Sakit ini merupakan rumah sakit rujukan utama penyakit paru di Provinsi Sumatera Barat yang terletak di Lubuk Alung Kabupaten Padang Pariaman. Sebagai rumah sakit rujukan, pasien yang datang sebagian besar adalah pasien dengan kecurigaan terinfeksi TBC yang

sudah melewati pemeriksaan awal. Selain itu, ada juga pasien yang datang hanya dengan gejala umum (*suspect*). Keunggulan utama rumah sakit ini adalah memiliki fasilitas layanan yang lebih lengkap seperti pengecekan BTA, TCM, Kultur, dan Radiografi *Thorax*. Sebagai rumah sakit rujukan, Rumah Sakit Paru Sumatera Barat menerima pasien dari berbagai daerah. Dengan demikian, data pasien di Rumah Sakit Paru Sumatera Barat dapat mencerminkan kondisi TBC di Provinsi Sumatera Barat.

Untuk menegakkan diagnosis TBC paru diperlukan pemeriksaan dahak secara mikroskopis atau yang dikenal dengan pemeriksaan Basil Tahan Asam (BTA). Pemeriksaan ini merupakan metode standar yang umum digunakan karena prosedurnya sederhana, biaya relatif murah, dan hasilnya cukup akurat dalam mendeteksi keberadaan bakteri penyebab TBC. Namun demikian, dalam praktiknya proses pemeriksaan BTA memerlukan waktu dua sampai tiga hari kerja. Tahapan laboratorium yang harus dilalui adalah seperti pengumpulan dahak, pewarnaan dahak, dan pembacaan hasil spesimen. Kondisi ini dapat memperlambat penentuan diagnosis awal sehingga berpotensi meningkatkan risiko penularan TBC kepada orang lain sebelum pasien mendapatkan pengobatan yang tepat.

Berdasarkan kondisi tersebut, diperlukan suatu pendekatan alternatif yang dapat membantu mempercepat proses diagnosis awal penyakit TBC. Salah satu pendekatan yang dapat diterapkan adalah analisis berbasis data klinis dan riwayat pasien untuk mengenali pola-pola gejala yang mengindikasikan adanya infeksi TBC. Hasil dari analisis tersebut diharapkan dapat digunakan sebagai pendukung dalam proses pengambilan keputusan medis secara cepat dan tepat, sehingga pasien dapat segera memperoleh tindakan penanganan yang sesuai. Dengan demikian,

sistem analisis ini dapat menjadi alat bantu bagi tenaga medis dalam pengambilan keputusan klinis serta berkontribusi pada upaya menekan angka penularan TBC agar tidak terjadi penularan yang lebih luas.

Dalam konteks penanganan penyakit TBC, *data mining* hadir sebagai salah satu pendekatan yang dapat dimanfaatkan untuk mengolah dan menganalisis data klinis pasien. Menurut Gorunescu (2011), *data mining* merupakan suatu proses menelusuri pengetahuan terbaru, pola, dan tren yang dipilih dari jumlah data yang besar dan disimpan dalam repositori atau tempat penyimpanan menggunakan teknik pengenalan pola statistik dan matematika. *Data mining* sering dikenal dengan nama *Knowledge Discovery in Databases* (KDD) yang merupakan kegiatan meliputi pengumpulan, pemakaian data historis untuk memecahkan pola atau hubungan keteraturan dalam dataset. Dengan menerapkan *data mining*, informasi penting seperti gejala dominan, faktor risiko, dan karakteristik pasien TBC dapat diidentifikasi secara lebih efisien, sehingga mendukung proses pengambilan keputusan medis. Banyak metode yang dapat dilakukan menggunakan *data mining* salah satunya adalah metode klasifikasi.

Klasifikasi adalah suatu proses menemukan sekumpulan pola atau fungsi-fungsi yang mendeskripsikan dan memisahkan kelas data satu dengan yang lainnya serta digunakan untuk memprediksi data yang belum memiliki kelas data tertentu (Setio dkk., 2020). Penerapan klasifikasi dalam bidang kesehatan menjadi sangat penting terutama pada penyakit menular seperti TBC. Melalui proses klasifikasi, sistem dapat membantu membedakan antara pasien yang terduga TBC dan pasien yang bukan TBC. Hal ini tentunya tidak hanya membantu tenaga medis, tetapi juga

menjadi solusi bagi pasien yang masih menunggu hasil pemeriksaan laboratorium untuk melakukan isolasi mandiri.

Salah satu metode klasifikasi adalah pohon keputusan yang merupakan sebuah diagram alur mirip dengan struktur pohon, dimana setiap simpul menotasikan atribut yang diuji, cabangnya menunjukkan hasil dari atribut dan simpul daun mempresentasikan kelas-kelas tertentu (Hand dkk., 2001). Konsep dari pohon keputusan atau yang biasa juga disebut *decision tree* adalah mengubah data menjadi pohon keputusan dan aturan-aturan keputusan. *Decision tree* juga dapat disebut sebagai pembelajaran induktif dimana kumpulan data tertentu digunakan untuk membangun model yang menangkap mekanisme dengan memunculkan data. Proses penyusunan model berjenis “*supervised learning*” karena pelatihan diawasi oleh variabel *outcome* yang disebut target (Kahfi dkk., 2023).

Banyak algoritma yang dapat digunakan dalam pembentukan pohon keputusan salah satunya adalah Algoritma C5.0. Metode algoritma C5.0 merupakan pengembangan dari algoritma C4.5 yang juga perpanjangan dari ID3. Algoritma C5.0 lebih baik daripada C4.5 pada kecepatan, memori, dan efisiensi. Langkah kerja pembuatan tree pada algoritma C5.0 mirip dengan pembuatan tree pada algoritma C4.5 (Pratiwi dkk., 2020). Kelebihan pada algoritma C5.0 adalah dapat mengatasi data dalam jumlah besar serta dapat digunakan untuk data kategorikal dan juga data numerik atau kontinu (Kusrin dan Luthfi, 2009).

Penelitian yang dilakukan oleh Rusdah dan Bregastantyo (2023), mengimplementasikan algoritma C4.5 dalam membangun model prognosis masa pengobatan pasien TBC. Dari hasil penelitian tersebut didapatkan hasil akurasi sebesar 74,33%. Penelitian ini menunjukkan bahwa algoritma *decision tree* efektif

dalam menganalisis data medis pasien TBC. Berdasarkan penelitian tersebut, algoritma C5.0 dikembangkan sebagai penyempurnaan dari C4.5 untuk mengklasifikasikan penyakit TBC dengan tingkat akurasi yang lebih tinggi.

Algoritma C5.0 telah banyak digunakan di beberapa penelitian terdahulu karena kemampuannya dalam menghasilkan model klasifikasi. Yogi (2007), membandingkan tiga metode klasifikasi yaitu algoritma C5.0, CART dan CHAID. Hasil penelitian menunjukkan bahwa algoritma C5.0 menghasilkan akurasi tertinggi yaitu sebesar 87,72%, dibandingkan dengan CART sebesar 87,28% dan CHAID sebesar 87,15%. Penelitian lainnya yang dilakukan oleh Sofyan dkk. (2023), terkait penerapan *data mining* dengan algoritma C5.0 untuk prediksi penyakit stroke, hasil yang didapatkan menunjukkan bahwa algoritma C5.0 menghasilkan akurasi sebesar 95%.

Algoritma C5.0 memiliki kemampuan yang baik dalam menangani permasalahan klasifikasi pada dataset berukuran kecil. Hal ini dibuktikan melalui penelitian *proof-of-concept* oleh Hannon dkk. (2024), yang mengembangkan model prediksi mortalitas 7-hari menggunakan C5.0 dengan jumlah data terbatas, namun tetap memperoleh hasil yang akurat. Keunggulan C5.0 terletak pada kemampuannya melakukan pemangkasan pohon keputusan (*pruning*) untuk menghindari *overfitting* serta menghasilkan model yang sederhana, efisien, dan mudah diinterpretasikan.

Dari hasil penelitian sebelumnya, algoritma C5.0 terbukti mampu dalam mengklasifikasikan data di bidang medis dengan tingkat akurasi yang tinggi. Keunggulan ini menjadikan algoritma C5.0 sebagai salah satu metode yang tepat

dan relevan untuk digunakan dalam penelitian klasifikasi penyakit TBC karena dapat membantu memprediksi atau mengidentifikasi pasien secara lebih akurat.

Berdasarkan paparan diatas, maka dilakukan penelitian dengan judul **“Klasifikasi Penyakit *Tuberculosis* di Rumah Sakit Paru Sumatera Barat Menggunakan Algoritma C5.0”**.

B. Batasan Masalah

Berdasarkan latar belakang yang dipaparkan, batasan masalah dalam penelitian ini adalah data yang digunakan yaitu data Penyakit TBC di Rumah Sakit Paru Sumatera Barat pada bulan Oktober hingga bulan Desember tahun 2024.

C. Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah:

1. Bagaimana hasil klasifikasi penyakit Tuberculosis di Rumah Sakit Paru Sumatera Barat menggunakan algoritma C5.0 berdasarkan data gejala klinis pasien?
2. Bagaimana hasil evaluasi model algoritma C5.0 dalam mengklasifikasikan penyakit TBC di Rumah Sakit Paru Sumatera Barat?

D. Tujuan Penelitian

Adapun tujuan dalam penelitian ini adalah:

1. Untuk mengetahui hasil klasifikasi penyakit TBC di Rumah Sakit Paru Sumatera Barat menggunakan algoritma C5.0 berdasarkan data gejala klinis pasien.
2. Untuk mengetahui akurasi algoritma C5.0 dalam mengklasifikasikan penyakit TBC di Rumah Sakit Paru Sumatera Barat.

E. Manfaat Penelitian

Adapun manfaat penelitian ini adalah sebagai berikut:

1. Bagi penulis, hasil dari penelitian ini berguna untuk menambah wawasan pengetahuan dan pemahaman tentang algoritma C5.0.
2. Bagi pembaca, penelitian ini dapat dijadikan referensi dan bahan acuan yang dapat dikembangkan lagi bagi peneliti selanjutnya dalam menerapkan metode pohon keputusan dengan algoritma C5.0.
3. Bagi Rumah Sakit Paru Sumatera Barat, sebagai hasil akhir dari penelitian ini semoga bisa membantu dan dijadikan bahan pendukung pihak Rumah Sakit Paru Sumatera Barat untuk melakukan identifikasi awal pasien TBC.

BAB V

PENUTUP

A. Kesimpulan

Berdasarkan hasil penelitian dan pembahasan yang telah dilakukan mengenai klasifikasi penyakit TBC di Rumah Sakit Paru Sumatera Barat menggunakan algoritma C5.0, maka dapat ditarik beberapa kesimpulan sebagai berikut:

1. Berdasarkan hasil penelitian, dapat disimpulkan bahwa algoritma C5.0 mampu menghasilkan klasifikasi penyakit Tuberculosis di Rumah Sakit Paru Sumatera Barat ke dalam kelas TBC positif dan TBC negatif berdasarkan data gejala klinis pasien. Hasil klasifikasi tersebut direpresentasikan dalam bentuk pohon keputusan, di mana keringat malam menjadi atribut utama, diikuti oleh nafsu makan berkurang, sesak napas, dan lama batuk sebagai atribut pemisah lanjutan.
2. Berdasarkan hasil evaluasi model pada beberapa proporsi pembagian data, dapat disimpulkan bahwa algoritma C5.0 memiliki tingkat *accuracy* yang tinggi. Hal ini menunjukkan bahwa model memiliki kemampuan yang sangat baik dalam mengklasifikasikan data pasien TBC di Rumah Sakit Paru Sumatera Barat.

B. Saran

Berdasarkan hasil penelitian yang telah diperoleh, beberapa saran yang dapat diajukan untuk pengembangan penelitian selanjutnya adalah sebagai berikut:

1. Penelitian selanjutnya disarankan untuk menggunakan jumlah data yang lebih besar dan periode waktu yang lebih panjang agar model yang dihasilkan memiliki kemampuan generalisasi yang lebih baik.

2. Penambahan variabel klinis atau hasil pemeriksaan penunjang lainnya diharapkan dapat meningkatkan ketepatan dan keandalan model dalam proses klasifikasi penyakit TBC.

DAFTAR PUSTAKA

- Azmi, B. N., Hermawan, A., & Avianto, D. (2023). Analisis Pengaruh Komposisi Data Training dan Data Testing pada Penggunaan PCA dan Algoritma Decision Tree untuk Klasifikasi Penderita Penyakit Liver. *JTIM: Jurnal Teknologi Informasi Dan Multimedia*, 4(4), 281–290.
- Badan Pusat Statistik. (2024a). *Jumlah Penduduk Menurut Provinsi di Indonesia (Ribuan Jiwa)*, 2024. Badan Pusat Statistik. <https://sulut.bps.go.id/id/statistics-table/2/OTU4IzI=/jumlah-penduduk-menurut-provinsi-di-indonesia.html>
- Badan Pusat Statistik. (2024b). *Kasus Penyakit Menurut Kabupaten/Kota dan Jenis Penyakit di Provinsi Sumatera Barat*, 2024. Badan Pusat Statistik. <https://sumbar.bps.go.id/id/statistics-table/3/YTA1Q1ptRmhUMEpXWTBsQmQyZzBjVzgwUzB4aVp6MDkzMw==/disease-by-regency-municipality-and-type-of-disease-in-sumatera-barat-province--2017.html?year=2024>
- Barberis, I., Bragazzi, N. L., Galluzzo, L., & Martini, M. (2017). The History of Tuberculosis : From the First Historical Records to the Isolation of Koch ' s Bacillus. *Journal of Preventive Medicine and Hygiene*, 58, 9–12.
- Bujlow, T., Riaz, T., & Pedersen, J. M. (2012). A method for classification of network traffic based on C5.0 machine learning algorithm. *Proceedings of the 2012 International Conference on Computing, Networking and Communications, ICNC'12*, 237–241. <https://doi.org/10.1109/ICCNC.2012.6167418>
- Dalvin, L. A., & Smith, W. M. (2017). Intraocular Manifestations of Mycobacterium Tuberculosis: A review of the literature. *Journal of Clinical Tuberculosis and Other Mycobacterial Diseases*, 7, 13–21. <https://doi.org/10.1016/j.jctube.2017.01.003>
- 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.
- 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
- Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques*. Morgan Kaufmann.
- Hand, D., Mannila, H. dan Smyth, P. (2001). *Principles of Data Minig*. MIT Press.

- Hannon, D. M., Syed, J. D. A., McNicholas, B., Madden, M., & Laffey, J. G. (2024). The Development of a C5.0 Machine Learning Model in a Limited Data Set to Predict Early Mortality in Patients With ARDS Undergoing an Initial Session of Prone Positioning. *Intensive Care Medicine Experimental*, 12(1). <https://doi.org/10.1186/s40635-024-00682-z>
- Harelvi, D. A., Salma, A., Kurniawati, Y., & Fitri, F. (2024). The Comparison of C4 . 5 and C5 . 0 Algorithms in Classifying the Nutritional Status of Stunted Toddlers. *UNP Journal of Statistics and Data Science*, 2(2), 213–218. <https://doi.org/https://doi.org/10.24036/ujsds/vol2-iss2/172>
- Isnaeni, P. N., Rakhmawati, H., & Tyas, F. A. (2024). Perbandingan Algoritma Naive Bayes dan C4.5 Pada Klasifikasi Penyakit Tuberculosis. *Journal of Informatics and Computing (RANDOM)*, 3(1), 41–48.
- Kahfi, M. D., Umbara, F. R., & Ashaury, H. (2022). Prediksi Pengagguran Menggunakan Decision Tree Dengan Algoritma C5.0 Pada Data Penduduk Kecamatan Caringin Kabupaten Bogor. *Informatics and Digital Expert (INDEX)*, 4(2), 75–80. <https://doi.org/10.36423/index.v4i2.913>
- Kementerian Kesehatan Republik Indonesia. (2019). *Pedoman Nasional Pelayanan Kedokteran Tata Laksana Tuberkulosis*. Kemenkes.
- Kementerian Kesehatan Republik Indonesia. (2024). *Ayo Kenali Gejala Tuberculosis yang Perlu Diwaspadai, Mulai dari Batuk hingga Berat Badan Menurun*. Kemenkes. <https://ayosehat.kemkes.go.id/ayokenali-gejala-tuberculosis-yang-perlu-diwaspadai-mulai-dari-batuk-hingga-berat-badan-menurun>
- Kementerian Kesehatan Republik Indonesia. (2025). *10 Provinsi dengan Kasus TBC Terbanyak di Indonesia Pada Tahun 2024*. Kemenkes. https://berinti.id/detail_berita/10-provinsi-dengan-kasus-tbc-terbanyak-di-indonesia-tahun-2024-ada-gorontalo?
- Nadesul. Handrawan, dr. (1995). *Penyebab, Pencegahan, dan Pengobatan TBC*. Jakarta: Puspa swara
- 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>
- Nusantara, B. S., & Akbar, M. (2024). Klasifikasi Penyakit Tuberculosis Berdasarkan Citra Rontgen Thorax Menggunakan Multi-Scale Convolutional Neural Network. *Sudo Jurnal Teknik Informatika*, 3(1), 1–8. <https://doi.org/10.56211/sudo.v3i1.483>
- Pardede, D., Hayadi, B. H., Komputer, I., Utama, U. P., Pembangunan, U., & Budi, P. (2022). *Kajian literatur multi layer perceptron: seberapa baik performa algoritma ini*. 1, 23–34.
- 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>
- Radji, M. 2010. *Buku Ajar Mikrobiologi: Panduan Mahasiswa Farmasi & Kedokteran*. Jakarta: EGC.
- Rusdah, & Bregastantyo, B. A. (2023). Model Prognosis Masa Pengobatan Pasien Tuberkulosis Dengan Metode C4.5. *Jurnal Teknologi Informasi Dan Ilmu Komputer*, 10(6), 1197–1204. <https://doi.org/10.25126/jtiik.2023107393>
- 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>
- Sebayang, V., Bangun, S. R., & Tarigan, R. V. B. (2024). Identifikasi Mycobacterium Tuberculosis Pada Pasien TB Paru dengan Metode Pewarnaan Ziehl Neelsen di UPT Puskesmas Medan Johor. *Jurnal Ilmu Kedokteran Dan Kesehatan*, 11(4), 809–817.
- 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.
- Wahdi, A., & Puspitosari, D. R. (2021). *Mengenal Tuberkulosis*. Pena Persada.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Partical Machine Learning Tools and Techniques* (3rd ed.). Morgan Kaufman.
- 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
- Zuriati, Suriya, M., & Ananda, Y. (2017). *Buku Ajar Asuhan Keperawatan Medikal Bedah Gangguan Pada Sistem Respirasi Aplikasi Nanda NIC & NOC*. Sinar Ultima Indah.