

MODEL PROGNOISIS MASA PENGOBATAN PASIEN TUBERKULOSIS DENGAN METODE C4.5

Rusdah^{*1}, Brian Agni Bregastantyo²

^{1,2}Universitas Budi Luhur, Jakarta Selatan
Email: ¹rusdah@budiluhur.ac.id, ²brian.agni7@gmail.com

*Penulis Korespondensi

(Naskah masuk: 23 Juni 2023, diterima untuk diterbitkan: 03 November 2023)

Abstrak

Pasien Tuberkulosis mempunyai jangka waktu pengobatan yang relatif beragam karena tingkat kepatuhan tiap pasien untuk meminum obat sampai dengan habis dan jangka waktu yang sudah ditentukan oleh Dokter Spesialis Paru. Apabila salah diagnosa terkait dosis obat maka akan meningkatkan faktor resiko kesehatan yaitu dimana proses pengobatan akan lebih memakan waktu dan lebih lama karena adanya kondisi *Multi-Drug Resistant*. Hal ini yang harus menjadi perhatian semua pihak agar tingkat kegagalan atas proses pengobatan pasien Tuberkulosis harus ditekan se minimal mungkin. Faktor kebiasaan pasien dan waktu minum obat pasien harus dijaga ketat agar masa pengobatan dapat lebih dipersingkat. Dokter Spesialis Paru berupaya untuk menekan tingkat *Drop Out* pasien Tuberkulosis dengan cara mengawasi jadwal mereka dengan pengelolaan yang baik. Oleh karena itu, dibutuhkan sistem untuk membantu proses prediksi masa pengobatan pasien dengan menerapkan *Cross-Industry Standard Process for Data Mining (CRISP-DM)* dan menggunakan pendekatan data mining dengan mengimplementasikan algoritma *C4.5* setelah dilakukan eksplorasi data menggunakan beberapa algoritma untuk klasifikasi dengan tujuan untuk hasil akurasi performa model untuk prognosis masa pengobatan pasien tuberkulosis. Melalui tahap *Data Understanding* dan *Data Preprocessing* menghasilkan atribut baru yaitu Lama Pengobatan. Dengan menggunakan 596 record mendapatkan hasil akurasi sebesar 74.33%.

Kata kunci: *Data Mining, Prognosis, Tuberkulosis, C4.5, CRISP-DM*

PROGNOSIS MODEL OF TREATMENT OF TUBERCULOSIS PATIENTS WITH C4.5 METHOD

Abstract

Tuberculosis patients have a relatively diverse treatment period because of the level of compliance of each patient to take the drug until it runs out and the time period has been determined by the Pulmonary Specialist. If a wrong diagnosis is related to drug dosage, it will increase health risk factors, namely where the treatment process will take more time and longer due to the Multi-Drug Resistant condition. This should be the concern of all parties so that the failure rate of the treatment process for tuberculosis patients must be kept to a minimum. The patient's habit factor and the patient's time to take medication must be closely monitored so that the treatment period can be shortened. Pulmonary Specialists try to reduce the Drop Out rate of Tuberculosis patients by monitoring their schedule with good management. Therefore, a system is needed to help predict the patient's treatment period by applying the Cross-Industry Standard Process for Data Mining (CRISP-DM) and using a data mining approach by implementing the C4.5 algorithm after exploring the data using several algorithms for classification with the aim of for the results of model performance accuracy for the prognosis of the treatment period of tuberculosis patients. Through the Data Understanding and Data Preprocessing stages, a new attribute is produced, namely the Length of Treatment. By using 596 records to get an accuracy of 74.33%.

Keywords: *Data Mining, Prognosis, Tuberkulosis, C4.5, CRISP-DM*

1. PENDAHULUAN

Tuberkulosis (TB) atau dikenal dengan nama Tuberkulosis adalah penyakit menular yang disebabkan oleh kuman *Mycobacterium Tuberculosis* yang menyerang paru – paru. Pada 2017, WHO melaporkan jumlah kasus TB di dunia mencapai 10

juta kasus dan Indonesia memiliki sejarah yang panjang, dan bisa ditelusuri dengan membagi beberapa periode. Setelah merdeka yaitu pada zaman Orde Lama (1945-1966) didirikan Lembaga Pemberantasan Penyakit Paru – paru (LP4) di Yogyakarta. Lembaga tersebut disebarluaskan hingga 53 lokasi (PDPI, 2002).

Multidrug-resistant tuberculosis (MDR), juga dikenal sebagai tuberkulosis yang resistan terhadap obat, berevolusi dari tuberkulosis pada umumnya, akhirnya menjadi kebal terhadap obat-obatan tertentu dan beberapa lainnya, tergantung kondisinya. Efek samping yang paling umum pada MDR-TB adalah mual dan muntah dan analgesia. Efek samping disfungsi ginjal berhubungan dengan terapi tuberkulosis sebelumnya. Dari perspektif psikiatri, ada berbagai macam penyakit yang memerlukan kerja sama erat antara tim klinis dan profesional kesehatan mental.

Rekam medis klinik JRC-PPTI menunjukkan bahwa 85% pasien tuberkulosis yang telah menjalani pengobatan selama lebih dari 6 bulan telah didiagnosis dengan resistensi multiobat, dan pada skala tersebut, adalah durasi pengobatan. Ini menunjukkan prognosis. Persentase pasien tuberkulosis yang penting untuk menentukan rencana pengobatan lebih lanjut. Pertimbangan Pengobatan Tuberkulosis. Diagnosis yang tepat diperlukan untuk hasil prognostik yang akurat. Penelitian ini menggunakan teknik klasifikasi data mining untuk membuat model prediksi durasi pengobatan tuberkulosis. Metode yang digunakan untuk memprediksi durasi pengobatan adalah *C4.5*. Variabel *input* disesuaikan dengan situasi Indonesia dengan segala kendala pada kebiasaan pengobatan pasien dan gaya hidup pasien. Tujuan penelitian ini adalah untuk mengidentifikasi pasien yang mengalami gejala berikut pada 6 bulan, 6–9 bulan, 9–12 bulan, dan >12 bulan setelah penerapan metode *C4.5*: pengobatan dapat dikategorikan Durasi pengobatan untuk pasien yang didiagnosis lebih dari enam bulan lalu perlu dipantau lebih dekat oleh Pengawas Minum Obat (PMO), dan penelitian ini akan membantu pemerintah Indonesia memperbaiki pengobatan kasus serupa di masa mendatang. diharapkan.

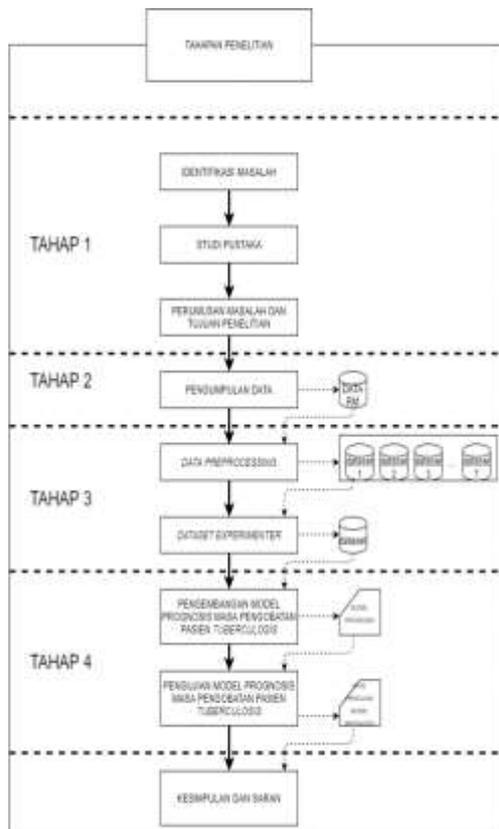
Sebuah penelitian yang bertujuan memprediksi kemungkinan kematian akibat tuberkulosis dilakukan oleh da Silva (Lino Ferreira da Silva Barros, 2021). Model terbaik dibangun dari model ansambel hutan acak, peningkatan gradien, dan perseptron berlapis. Sebuah studi untuk mengembangkan model prognostik jangka panjang untuk kasus TB paru yang baru diobati dilakukan dengan menggunakan single classifier dengan akurasi 74% dan sensitivitas 76%. Pemilihan fitur dan metode ansambel dapat digunakan untuk meningkatkan akurasi. Penelitian untuk menentukan determinan waktu penyembuhan penyakit tuberkulosis kategori I di Kabupaten Semarang menunjukkan bahwa variabel keteraturan berobat tidak berhubungan dengan lama waktu penyembuhan. Hal ini berbeda dengan Strategi Nasional Pengendalian TB Indonesia Tahun 2020-2024 (Azizah, 2019). Oleh karena itu, pada penelitian ini ditambahkan variabel 'kepatuhan minum obat' untuk mengetahui faktor penentu waktu penyembuhan atau akhir pengobatan pada kasus baru

pasien TB Paru. Setelah faktor penentu diketahui, model dikembangkan untuk memprediksi berapa lama pasien akan menerima pengobatan untuk swab positif atau negatif. Manfaat dari penelitian ini adalah memberikan peringatan dini dan membantu meningkatkan kepatuhan pasien dalam menyelesaikan proses pengobatan. Oleh karena itu, ini membantu mengurangi jumlah kasus putus sekolah, kegagalan, atau resistensi multi-obat (MDR).

Penelitian selanjutnya dengan judul “Pemanfaatan *Data Mining* Untuk Diagnosis Penyakit Tuberkulosis” dengan hasil pembahasan bahwa diagnosis penyakit Tuberkulosis dapat dilakukan secara konvensional dan teknik penambangan data. Diagnosis secara konvensional dapat dilakukan dengan Uji Tuberkulin dan pemeriksaan dahak dengan akurasi 70%. Diagnosis menggunakan teknik penambangan data memberikan hasil akurasi terbaik diperoleh dari penelitian Asha dkk (2013) dengan menggunakan SVM dengan akurasi tertinggi 98.7% diikuti oleh Bagging 98.4%, dan Random Forest dengan 98.3% dibandingkan dengan metode lainnya (Iswanto, 2015).

Di era di mana genomik diimplementasikan ke dalam pengiriman kesehatan dan data kesehatan menjadi semakin digital, kecerdasan buatan dan DL akan digunakan dalam pengembangan, validasi, dan implementasi alat pendukung keputusan untuk memfasilitasi onkologi presisi. Dalam ulasan ini, kami memamerkan sejumlah aplikasi DL yang menjanjikan di berbagai bidang onkologi, termasuk histopatologi digital, subtipe molekuler, diagnosis kanker, prognostikasi, inferensi histologis karakteristik genomik, lingkungan mikro tumor, dan batas yang muncul seperti transkriptomik spasial dan farmakogenomik. Saat penelitian matang, masa depan DL terapan dalam onkologi kemungkinan akan fokus pada integrasi gambar medis dan data omics menggunakan pembelajaran multimodal yang dapat mengidentifikasi biomarker yang bermakna secara biologis. Menariknya, kombinasi pembelajaran multimodal dan kemampuan menjelaskan dapat mengungkapkan wawasan baru. Prasyarat penting dari adopsi DL secara luas dalam pengaturan klinis adalah data yang kaya secara fenotip untuk model pelatihan dan validasi klinis dari relevansi biologis dari wawasan yang dihasilkan DL. Kami berharap karena teknologi baru seperti pengurutan sel tunggal, transkriptomik spasial, dan pencitraan multipleks menjadi lebih mudah diakses, lebih banyak upaya akan dilakukan untuk meningkatkan kuantitas dan kualitas pelabelan/anotasi data medis (Tran., 2021).

2. METODE PENELITIAN



Gambar 1 Tahapan Penelitian

Gambar 1 menunjukkan tahapan yang digunakan dalam penelitian ini, yaitu pada Tahap 4 untuk membuat pengembangan model prognosis masa pengobatan pasien tuberkulosis. Tahapan ini terdiri dari beberapa tahap yang akan dijelaskan sebagai berikut :

2.1 Identifikasi Masalah

Peneliti mengidentifikasi masalah terkait dengan topik yang bersangkutan berdasarkan informasi yang didapat pada studi literatur, wawancara, observasi langsung dan sumber-sumber yang mendukung penelitian ini.

2.2 Analisa Masalah

Berdasarkan penelitian sebelumnya belum ada parameter kepatuhan minum obat. Serta nilai perfroma dari penelitian sebelumnya masih kecil. Sehingga parameter kepatuhan minum obat diharapkan dapat meningkatkan tingkat kesembuhan pasien sesuai dengan jadwal nya. Dengan demikian akan mengurangi tingkat MDR ketika kepastian dari obat tersebut pasti di habiskan. Tahap pertama dalam penelitian ini untuk menganalisa permasalahan yang terkait dengan penelitian. Pada tahapan ini terdapat sub-tahapan dengan menerapkan metodologi *CRISP-DM*. Model proses *CRISP-DM* memiliki beberapa

fase diantaranya adalah sebagai berikut (Shearer, 2000) :

a. *Business Understanding*

Fase pemahaman bisnis adalah tahap pertama dari pengembangan data mining. Ini adalah fase memahami tujuan dan kebutuhan dalam hal rentang durasi perawatan pasien JRC PPTI, kemudian menerjemahkan pemahaman ini ke dalam definisi data mining dan kemudian ke dalam rencana dan strategi untuk mencapai tujuan tersebut.

b. *Data Understanding*

Tahap pemahaman data dimulai dengan pengumpulan data awal dan pengumpulan hasil yang dapat digunakan secara konsekuen. Kemudian dilanjutkan melalui proses mendapatkan pemahaman yang lebih dalam tentang data, mengidentifikasi masalah kualitas data, dan mengidentifikasi bagian menarik dari data yang dapat digunakan untuk informasi hipotetis tersembunyi.

c. *Data Preparation*

Tahap persiapan data (data preparation) adalah tahap yang mencakup semua kegiatan yang diperlukan untuk membuat data yang akan menjadi masukan ke dataset akhir atau alat pemodelan dari data mentah awal. Fase data awal dapat dieksekusi beberapa kali tanpa urutan tertentu. Fase ini meliputi pemilihan tabel, dataset, dan atribut data, termasuk proses pembersihan dan transformasi data selama fase pemodelan.

d. *Modeling*

Tahap pemodelan berjalan setelah tahap persiapan data selesai. Selama tahap pemodelan, teknik pemodelan dipilih dan diterapkan, dan beberapa parameter disesuaikan untuk mendapatkan nilai optimal. Ada beberapa model untuk memecahkan masalah yang sama dalam data mining. Beberapa model ini memiliki spesifikasi dan format data tertentu. Pada penelitian ini digunakan algoritma *C4.5*.

e. *Evaluation*

Pada fase ini, Anda harus mengevaluasi dan memvalidasi bahwa model yang dibangun dengan algoritma *C4.5* benar-benar mencapai tujuan bisnis yang ditetapkan pada fase pemahaman bisnis.

2.3 Kesimpulan dan Saran

Berisi tentang kesimpulan dan saran dari keseluruhan proses tersebut berdasarkan hasil analisa masalah menggunakan metodolgi *CRISP-DM*.

2.4 Teknik Pengumpulan Data

Masuk pada tahap ke dua, Metode pengumpulan data pada penelitian ini merupakan data sekunder yang penulis dapatkan melalui Dr. Rusdah selaku peneliti sebelumnya pada Klinik JRC - PPTI. Dalam analisis ini, yang diteliti adalah data rekam medis sebesar 636 data yang terdiri dari atribut Jenis Kelamin, Usia, Kelurahan, Kecamatan, Pekerjaan, Riwayat Lama Batuk, Dahak, Darah, Demam,

Penurunan BB, Nafsu Makan, Keringat, Sesak, Nyeri Dada, Berat Badan, Diagnosis Awak, RO Awal, Dahak Awal, Diagnosis Akhir, Tanggal Diagnosis, Bulan 1, OAT Bulan 2, Akhir Intensif, OAT Bulan 3, OAT Bulan 4, OAT Bulan 5, Hasil Dahak, OAT Akhir Pengobatan, RO Akhir Pengobatan, Hasil Akhir.

2.5 Data Preprocessing

Pada Tahap ke tiga ini, Data yang digunakan harus melalui data *preprocessing*. *Preprocessing* terdiri dari empat jenis: pembersihan data, integrasi data, transformasi data, dan reduksi data. Sebanyak 636 rekam medis diterima. Pada penelitian ini *preprocessing* data yang dilakukan terdiri dari pengujian data asli dan penambahan atribut baru untuk memenuhi persyaratan.

2.6 Penentuan Data Latih (*Data Training*) dan Data Uji (*Data Testing*)

Pada penelitian ini penentuan *data training* dan *data testing* dilakukan pembagian data sebesar 70:30 pada data rekam medis, 70% *data training* dan 30% *data testing* berdasarkan dari lama pengobatan sebagai *label class*.

2.7 Pengembangan Model

Masuk dalam tahap ke empat, Dalam pengembangan model, peneliti menerapkan algoritma *C4.5* untuk prognosis masa pengobatan pasien tuberkulosis. Algoritma *C4.5* merupakan metode klasifikasi yang dapat memprediksi probabilitas kelas dan mengambil keputusan berdasarkan data pelatihan. *C4.5* memiliki keunggulan karena mudah digunakan, sederhana, cepat, dan akurat. Keuntungan dari metode ini adalah hanya dibutuhkan sedikit data pelatihan untuk menentukan estimasi parameter yang dibutuhkan dalam proses klasifikasi.

2.8 Teknik Pengujian

Pada tahap ini dilakukan pengujian fungsional sistem dengan menggunakan metode multiclass kebingungan matriks, yang merupakan ringkasan dari hasil prediksi untuk masalah klasifikasi. Nilai yang dihitung meringkas jumlah prediksi yang benar dan salah dan mengelompokkannya berdasarkan kelas. Ini adalah kunci dari matriks kebingungan multikelas. Matriks kebingungan multikelas menunjukkan bagaimana model diklasifikasikan saat membuat prediksi. Ini akan memberi tahu Anda tidak hanya kesalahan apa yang menyebabkan hasil klasifikasi, tetapi yang lebih penting, jenis kesalahan apa yang terjadi. Detail ini mengatasi keterbatasan penggunaan akurasi klasifikasi saja. Di bawah ini adalah proses menghitung *Multi-class Confusion Matrix*.

- Anda memerlukan dataset pengujian atau dataset validasi dengan nilai hasil yang diharapkan.

- Buat prediksi untuk setiap baris dalam dataset uji Anda.

Dari hasil dan perkiraan prediksi yang diharapkan:

- Jumlah prediksi yang benar untuk setiap kelas.
- Jumlah prediksi yang salah untuk setiap kelas, yang diselenggarakan oleh kelas yang diprediksi. Angka-angka ini kemudian disusun dalam tabel, atau matriks sebagai berikut:
- Diharapkan di sisi: Setiap baris matriks sesuai dengan kelas prediksi.
- Diprediksi di bagian atas: Setiap kolom matriks sesuai dengan kelas aktual.
- Hitungan klasifikasi yang benar dan salah kemudian diisi ke dalam tabel.

Jumlah total prediksi yang benar untuk suatu kelas masuk ke baris yang diharapkan untuk nilai kelas itu dan kolom yang diprediksi untuk nilai kelas itu. Dengan cara yang sama, jumlah total prediksi yang salah untuk suatu kelas masuk ke baris yang diharapkan untuk nilai kelas itu dan kolom prediksi untuk nilai kelas itu.

3. HASIL DAN PEMBAHASAN

3.1. Deskripsi Umum Penelitian

Model prognosis masa pengobatan pasien Tuberkulosis tujuannya adalah untuk mengklasifikasikan masa pengobatan pasien Tuberkulosis yang memiliki 4 kelompok yaitu pada jangka waktu 0-6 bulan, 6-9 bulan, 9-12 bulan, dan lebih dari 12 bulan. Berdasarkan latar belakang dan rumusan masalah yang telah disampaikan pada bab 1 penelitian ini, maka pada bab ini akan disampaikan mengenai rancangan solusi yang diusulkan. Model prognosis masa pengobatan diusulkan untuk mendiagnosa kemungkinan masa pengobatan yang akan ditempuh oleh seorang pasien tuberkulosis tersebut, setelah pasien didiagnosis Tuberkulosis Paru BTA Positif (*Smear-Positive Pulmonary Tuberculosis*) atau Tuberkulosis Paru BTA Negatif (*Smear-Negative Pulmonary Tuberculosis*). Minimal masa pengobatan yang harus dijalani pasien adalah 6 bulan. Kemudian apabila diketahui sejak awal bahwa prognosis masa pengobatan seorang pasien lebih dari 6 bulan, maka pasien tersebut perlu pengawasan lebih ketat baik dari pihak keluarga, Pengawas Minum Obat (PMO) atau dari pihak DOTS di unit kesehatan pasien tersebut berobat. Meningkatkan kesadaran dan kepatuhan untuk meminum obat dengan rutin dan teratur akan menaikkan prosentase kesembuhan pasien, serta dapat menurunkan prosentase *Multi-Drug Resistance (MDR)* atau kekebalan pasien terhadap 2 jenis obat tuberkulosis karena ketidak teraturan dan ketidak patuhan pasien dalam meminum obat. Dengan naiknya prosentase kesembuhan maka akan mengurangi jumlah pasien dalam kasus Gagal atau disebut *Drop Out*.

3.2. Data Rekam Medis

Data pasien yang telah terkumpul sejumlah 636 record. Terdiri atas 503 pasien dengan diagnosa awal TB Paru, 93 pasien dengan diagnosa awal non TB, dan 40 pasien dengan diagnose awal TB Ekstra Paru. Dengan kategori seluruh pasien berusia diatas 15 tahun. Total data pasien yang digunakan adalah 636 pasien. Berdasarkan jenis kelamin, terdapat 267 pasien perempuan (42%) dan 369 pasien laki-laki (58%). Rentang usia antara 15 hingga 89 tahun.

No	Range	Kategori
1	Pengobatan 1-6 Bulan	0-6 Bulan
2	Pengobatan >1 hari dari 6 Bulan	6-9 Bulan
3	Pengobatan >1 hari dari 9 Bulan	9-12 Bulan
4	Pengobatan >1 hari dari 12 Bulan	>12 Bulan

Tabel 1 Transformasi Atribut

3.3. Data Preprocessing

3.3.1 Tahap 1

Kegiatan preprocessing dalam penelitian ini meliputi reduksi data dengan melakukan *data cleaning* untuk menghilangkan *missing value*, kemudian melakukan *attribute reduction* yaitu dengan mereduksi atribut kecamatan karena menurut para ahli atribut ini tidak diperlukan untuk proses klasifikasi. Kemudian proses reduksi selanjutnya adalah dengan *remove* nilai Ekstra TB karena pada proses klasifikasi hanya diperlukan data yang pasiennya terindikasi sebagai TB Paru dan Non TB. Yang di maksud dengan Ekstra TB adalah nilai dari TB Tulang, dan jenis TB lainnya. Kemudian pada atribut Diagnosa Akhir *replace* nilai dari pemeriksaan Dahak Awal apabila nilainya negatif maka hasil pada Diagnosa Akhirnya adalah TB Paru BTA Negatif. proses *feature selection* untuk menentukan data mana saja yang diperlukan untuk proses *modelling*.

3.3.2 Tahap 2

Transformasi data membentuk atribut baru yaitu lama pengobatan. Lama pengobatan adalah hasil perhitungan tanggal akhir pengobatan dikurangi dengan tanggal akhir diagnosis. Kemudian dilakukan reduksi data dengan menghilangkan atribut Tanggal Akhir Pengobatan dan Tanggal Akhir Diagnosis sebagai penilaian terhadap independensi data atau ke tidak tergantungan data dari hasil transformasi data dengan nilai kedua atribut sebelumnya yang di hilangkan Pada tahap ini akan menseleksi 28 atribut yang ada pada RM 4 yang terdiri dari JK, Usia,

Pekerjaan, Batuk(Hari), Dahak, Darah, Demam >1 Bulan, Penurunan Berat Badan, Nafsu Makan, Keringat Malam, Sesak, Nyeri Dada, Tambahan, Berat Badan, Diagnosa Awal, RO Awal, Dahak Awal, Diagnosa Akhir, Bulan ke-1, OAT Bulan 2, Akhir intensif, OAT Bulan 3, OAT Bulan 4, OAT Bulan 5. Hasil Dahak Bulan ke-5, OAT AP, RO AP, Hasil Akhir Pengobatan.

Kegiatan ini dilakukan untuk menentukan atribut mana yang mendukung untuk proses penentuan model prognosis masa pengobatan pasien tersebut dengan nilai akurasi dengan algoritma yang sesuai dengan tingkat ke akuratan data tersebut.

Kegiatan ini dilakukan untuk menentukan atribut mana yang mendukung untuk proses penentuan model prognosis masa pengobatan pasien tersebut dengan nilai akurasi dengan algoritma yang sesuai dengan tingkat ke akuratan data tersebut. Lebih jelasnya bisa lihat pada tabel 1

3.3.3 Tahap 3

Pada tahap ini dilakukan proses eksplorasi *feature selection* dengan menggunakan *wrapper method* dengan menggunakan 9 algoritma *classifier* yaitu *Naive Bayes*, *SVM Poly Kernel*, *SVM Normalized PolyK*, *SVM PUK*, *SVM RBFK*, *K-Nearest Neighbour* dengan masing masing percobaan menggunakan nilai k yang berbeda yaitu dengan k1, k3, k5, dan C4.5. Setiap *classifier* mengeluarkan hasil yang berbeda. Kemudian nilai hasil keluar dengan menggunakan 9 *classifier* tersebut dengan ada nilai yang berhasil keluar dan gagal. Nilai-nilai tersebut terbagi menjadi 7 dataset, yaitu:

Tabel 2 Hasil Eksplorasi Feature Selection dengan Wrapper Method

Classifier	Atribut Terpilih
Naive Bayes	Batuk (Hari), Darah, Penurunan BB, Tanggal Diagnosa Akhir
	Jenis Kelamin,
SVM PUK	Pekerjaan, Nyeri Dada, Tanggal Diagnosa Akhir
	Pekerjaan, Dahak,
KNN K1	Diagnosa Awal, dan Tanggal Diagnosa Akhir
	Jenis Kelamin,
KNN K3	Penurunan Berat Badan, dan RO Awal
	Pekerjaan, Dahak,
KNN K5	Diagnosa Awal, dan Tanggal Diagnosa Akhir
	Pekerjaan dan Dahak
C4.5	Jenis Kelamin, Kelurahan, Pekerjaan, Batuk, Dahak, Darah, Penurunan Berat Badan,
	Sesak, Nyeri Dada, Diagnosa Awal, RO Awal, Tanggal Diagnosa Akhir, Akhir Intensif, dan RO AP

3.3.4 Tahap 4

Pada tahap ini akan dilakukan perbandingan dataset dengan 9 algoritma yang akan dibandingkan lagi dengan 6 baseline. Dimana tahap ini akan membandingkan hasil dataset RM 4 sebelumnya yang telah dilakukan *feature selection* yang menghasilkan 6 dataset. Yaitu dataset RM 4 dengan *Naive Bayes*, dataset RM 4 dengan *SVM PUK*, dataset RM 4 dengan *K-Nearest Neighbour* dengan nilai K, K2, dan K3, dan dataset RM 4 dengan *C4.5*. Keenam dataset ini akan diseleksi kembali menjadi 6 *baseline*. Dari hasil pengujian berikut maka diputuskan untuk menggunakan hasil dataset RM 4 yang telah dilakukan *feature selection* dengan menggunakan algoritma *K-Nearest Neighbour* dengan nilai K3.

3.4. Modeling

3.4.1 Hasil Pengujian dengan WEKA

Pada tahap penelitian ini, tingkat akurasi terbaik diuji dengan membandingkan tingkat akurasi yang diperoleh dari beberapa metode data mining. Metode yang digunakan sebagai pembanding adalah Algoritma *Naive Bayes*, *Support Vector Machines*, *K-Nearest Neighbors*, dan *C4.5*. Pengujian dilakukan dengan menggunakan alat WEKA untuk lebih memastikan akurasi yang dicapai.

Fase ini menggunakan validasi silang operator. Ini membantu dalam memperkirakan tingkat akurasi kinerja berdasarkan model pelatihan dari berbagai algoritme yang disertakan dalam WEKA.

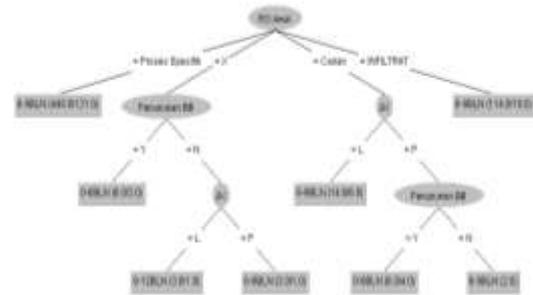
Tabel 3 Perbandingan Hasil Klasifikasi

No.	Metode	Akurasi	Presisi	Recall
1	<i>C4.5</i>	74.33%	76%	98%
2	<i>K-Nearest Neighbour K1</i>	74.32%	76%	98%
3	<i>K-Nearest Neighbour K3</i>	73.45%	74%	97%
4	<i>K-Nearest Neighbour K5</i>	72.45%	74%	99%
5	<i>Naive Bayes</i>	73.82%	74%	98%
6	<i>Support Vector Machine</i>	73.45%	75%	99%

Dari hasil percobaan dengan WEKA maka diperoleh informasi nilai akurasi *C4.5* merupakan nilai terbaik disbanding algoritma lain terhitung dengan pertimbangan nilai *precision* dan *recall* sehingga bisa disimpulkan metode *C4.5* adalah metode yang paling tepat untuk prognosis masa pengobatan pasien *Tuberculosis* pada penelitian ini.

3.4.2 Pembuatan Rule Hasil Decision Tree

Dari pohon keputusan akan terbentuk aturan – aturan dari penurunan data. Maka dari hasil *visualize tree* terbentuklah *rule* atau aturan.



Gambar 2 Hasil Visualize Tree

Hasil *Rule* diatas menyebutkan bahwa atribut RO Awal menjadi *node* awal berdasarkan hasil eksplorasi yang dilakukan oleh algoritma *C4.5*. Sedangkan dari atribut Penurunan Berat Badan yang secara fisik dapat kita lihat menjadi ciri utama pasien Tuberculosis berada menjadi cabang dari *node* awal dari atribut RO Awal karena pada proses diagnosa dan daftar rekam medis harus dilakukan pengecekan *rontgen* untuk mendiagnosa apakah pasien dengan ciri tersebut menjadi pasien terduga Tuberculosis atau tidak. Daftar aturan yang terbentuk adalah:

- Rule 1 Jika RO Awal bernilai Proses Spesifik maka masa pengobatan pasien maka masa pengobatan pasien dalam jangka waktu 6-9 Bulan.
- Rule 2 Jika RO Awal bernilai Cairan dan Jenis Kelamin laki-laki maka masa pengobatan pasien dalam jangka waktu 6-9 Bulan.
- Rule 3 Jika RO Awal bernilai Cairan dan Jenis Kelamin perempuan serta tidak mengalami Penurunan Berat Badan maka masa pengobatan pasien dalam jangka waktu 6-9 Bulan.
- Rule 4 Jika RO Awal bernilai Infiltrat maka masa pengobatan pasien dalam jangka waktu 6-9 Bulan.
- Rule 5 Jika tidak dilakukan RO Awal dan tidak terjadi Penurunan Berat Badan serta Jenis Kelamin perempuan maka masa pengobatan pasien dalam jangka waktu 6-9 Bulan.
- Rule 6 Jika RO Awal bernilai X (tidak dilakukan RO) dan tidak terjadi Penurunan Berat Badan serta Jenis Kelamin laki-laki maka masa pengobatan pasien dalam jangka waktu 9-12 Bulan.
- Rule 7 Jika RO Awal bernilai X (tidak dilakukan RO) dan terjadi Penurunan Berat Badan maka masa pengobatan pasien dalam jangka waktu 0-6 Bulan.
- Rule 8 Jika RO Awal bernilai Cairan dan Jenis Kelamin perempuan serta mengalami penurunan berat badan maka masa pengobatan pasien dalam jangka waktu 0-6 Bulan.

Berdasarkan hasil aturan atau pengetahuan yang dihasilkan, ada beberapa aturan yang sangat cocok dengan apa yang terjadi dengan prognosis lama pengobatan pasien tuberkulosis. Nilai pasien yang ditentukan dalam aturan termasuk dalam empat nilai kelas lama pengobatan untuk pasien tuberkulosis.

3.5. Evaluation

Pada tahap ini, hasil yang diperoleh diukur. Berdasarkan pengujian pembacaan pohon keputusan dengan sinkronisasi dataset, tingkat akurasi algoritma C4.5 dalam hal ini kurang dari 100%. Akurasi klasifikasi berdasarkan tabel matriks kebingungan multikelas. Matriks konfusi multikelas untuk uji data 0-6 bulan, 6-9 bulan, 9-12 bulan, dan >12 bulan. Berdasarkan matriks konfusi multikelas di atas, tingkat akurasi klasifikasi algoritma C4.5 adalah 74,33%. Akurasi adalah persentase keseluruhan data percobaan yang dinyatakan benar saat memproses query (Astuti, 2017). Perhitungan akurasi, presisi, dan sensitivitas/recall diberikan oleh rumus berikut.:

$$\begin{aligned} \text{Accuracy} &= \frac{TP+TN}{TP+TN+FP+FN} \\ &= \frac{(0+437+5+1)/(0+25+2+1+0+437+5+2+0+91+5+1+1+23+2+1)}{443 / 596} \\ &= 0.74328859060402684563 \\ &= 0.74 \\ &= 74\% \end{aligned}$$

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP+FP} \\ &= \frac{(437)/(437+25+91+23)}{437/576} \\ &= 0.75868055555555 \\ &= 0.76 \\ &= 76\% \end{aligned}$$

$$\begin{aligned} \text{Sensitivity} &= \frac{TP}{TP+FN} \\ &= \frac{(437)/(437+0+5+2)}{437/444} \\ &= 0.98423423423423423423 \\ &= 0.98 \\ &= 98\% \end{aligned}$$

4. KESIMPULAN

Berdasarkan hasil dari pengujian model prognosis masa pengobatan pasien Tuberkulosis pada Klinik JRC dapat kesimpulan bahwa algoritma C4.5 dapat digunakan untuk model prognosis masa pengobatan pasien Tuberkulosis karena bentuk *output* nya berbentuk *tree* dan *rule* sehingga dapat mempermudah Dokter Spesialis Paru atau tenaga kesehatan yang berwenang untuk menentukan rencana terapi selanjutnya dengan pertimbangan hasil *visualize tree* dan *rule* tersebut. Sehingga setelah diberikan panduan akan rencana terapi selanjutnya dapat meningkatkan tingkat kepatuhan minum obat dan mengurangi tingkat *Drop Out* dari kasus *Multi-Drug Resistant*. Saran untuk selanjutnya bahwa penelitian ini :

1. Komposisi dari data yang digunakan lebih dominan kelas b sehingga penelitian selanjutnya bisa dilanjutkan dengan *classifier* yang lebih baik dalam mengerjakan *imbalance dataset*.

2. Perlu dilakukan penelitian lanjutan dengan menggunakan data kepatuhan minum obat dengan tujuan tingkat akurasi dari model prognosis masa pengobatan pasien dapat lebih tinggi.

Penelitian ini dapat dikembangkan lagi dengan *label class* numerik dan dengan menggunakan algoritma regresi agar bentuk *output* nya dapat langsung diketahui berapa lama masa pengobatan pasien Tuberkulosis tersebut.

DAFTAR PUSTAKA

- AMRAN, R., ABDULKADIR, W. AND MADANIA, M., 2021. 'Tingkat Kepatuhan Penggunaan Obat Anti Tuberkulosis Pada Pasien Di Puskesmas Tombulilato Kabupaten Bone Bolango', *Indonesian Journal of Pharmaceutical Education*, 1(1), pp. 57–66. doi: 10.37311/ijpe.v1i1.10123.
- AYAZ, M., SHAUKAT, F. AND RAJA, G., 2021. 'Ensemble learning based automatic detection of tuberculosis in chest X-ray images using hybrid feature descriptors', *Physical and Engineering Sciences in Medicine*, 44(1), pp. 183–194. doi: 10.1007/s13246-020-00966-0.
- AZIZAH, I., 2019. 'Determinan Lama Waktu Kesembuhan Pada Kategori I Di Rsd Ungaran Kabupaten Semarang', p. 167.
- ISWANTO, PERMANASARI M. H. NUGROHO A.E., GRAFIKA H. A., 2015 'Pemanfaatan Teknik Data Mining Untuk Diagnosis Tuberculosis (TBC)', *Seminar Nasional Teknologi Informasi dan Multimedia (STMIK AMIKOM), Yogyakarta*, pp. 121–126.
- LAKSHMI, K. R., 2013. 'Utilization of Data Techniques for Prediction and Diagnosis of Tuberculosis Disease Survivability'. *International Journal of Modern Education and Computer Science*, 5(8), pp. 8–17. doi: 10.5815/ijmecs.2013.08.02.
- LINO FERREIRA DA SILVA BARROS, M. H., SOUZA G. O., ROCHA L., LYNN J. F., SAMPAIO T., ENDO V., VANDERSON., 2021. 'Benchmarking machine learning models to assist in the prognosis of *Informatics*, 8(2), pp. 1–17: 10.3390/informatics8020027.
- NESREDIN, A., 2012. *Mining Patients' Data for Effective Tuberculosis Diagnosis: Case of Menelik II Hospital*. Addis Ababa University. doi: 10.1002/ana.23645..
- PDPI, 2002. *Tuberkulosis Pedoman Diagnosis dan Penatalaksanaan di Indonesia*.
- RUSDH, WINARKO, E. AND WARDOYO, R., 2017. 'Predicting The Suspect of New Pulmonary Tuberculosis Case using SVM, C5.0 and Modified Moran's I', *International Journal of Computer Science and Network*

- Security*, 17(12), pp.164–171.
- SRIRAM, A., MUCKLEY A., SINHA M., SHAMOUT K., PINEOU F., AZOUR K. J., YAKUBOVA V., MOORE N., 2021. ‘COVID-19 Deterioration Prediction via Self-Supervised Representation Learning and Multi-Image Prediction.’, *ArXiv*, pp.–17.
- TRAN, KONDRASHOVA K. A., WILLIAMS A., PEARSON E. D., WADELL J. V., NICOLA., 2021. ‘Deep learning in cancer diagnosis, prognosis and treatment selection’, *Genome Medicine*, 13(1), pp. 1–17. doi:1186/s13073-021-00968-x.
- Tuberculosis Coalition for Technical Assistance (2006) *International Standards for Tuberculosis Care (ISTC)*. San Francisco, Canada.
- TUFAIL, A., KAABAR Y. K., MARTINEZ M. K., JUNEJO F., KHAN I., RAHIM., 2021. ‘Deep Learning in Cancer Diagnosis and Prediction: A Minireview on Challenges, Recent Trends, and Future Directions’, *Computational and Mathematical Methods in Medicine*. doi: 10.1155/2021/9025470.
- VALVERDE-ALBACETE, F. J. AND PELÁEZ-MORENO, C., 2010. ‘Two information-theoretic tools to assess the performance of multi-class classifiers’, *Pattern Recognition Letters*, 31(12), pp. 1665–1671. doi: 10.1016/j.patrec.2010.05.017.
- WARJIMAN, W., BERNIATI, B. AND ER UNJA, E., 2022. ‘Hubungan Dukungan Keluarga Terhadap Kepatuhan Minum Obat Pasien Paru Di Puskesmas Bilu’, *Jurnal Keperawatan Suaka Insan (Jksi)*, 7(2), pp. 163–168. doi: 10.51143/jksi.v7i2.366.
- WHO, 2011. *Global Tuberculosis Control: WHO Report 2011*.
- WHO, 2022. *Global Tuberculosis Control: WHO Report 2022*.