

PENERAPAN ALGORITMA DECISION TREE C4.5 DALAM KLASIFIKASI IDENTIFIKASI PASIEN PENYAKIT TUBERKULOSIS (TB) DI PUSKESMAS SUKORAMBI JEMBER

¹Muhammad Bakhtiar Lazuardi, ²Hardian Octavianto, ³Henny Wahyu Sulisty

Jurusan Teknik Informatika Fakultas Teknik Universitas Muhammadiyah Jember

¹tyarlazuardy22@gmail.com, ²hardian@unmuhjember.ac.id, ³hennywahyu@unmuhjember.ac.id

ABSTRAK

Tuberkulosis (TB) adalah penyakit yang mempengaruhi paru-paru yang tergolong sebagai penyakit menular yang menyebabkan masalah kesehatan terbesar kedua di dunia setelah HIV. Penelitian ini bertujuan mengetahui berapa tingkat akurasi dan presisi hasil klasifikasi rekomendasi pendonor darah menggunakan algoritma decision tree c4.5 berdasarkan atribut batuk ≥ 2 Minggu, penurunan berat badan, nafsu makan, demam, batuk darah, nyeri dada dengan output positif terjangkit TB dan negatif tidak terjangkit TB. Setelah dilakukan pengujian data sebanyak 5 kali pengujian terhadap data pasien penyakit Tuberkulosis (TB). Dari penelitian ini diperoleh akurasi rata-rata dari 5 kali uji yaitu sebesar 91.47%. Dari penelitian ini diperoleh presisi rata-rata dari 5 kali uji yaitu sebesar 76.12%. Tingkat akurasi tertinggi yang diperoleh pada penelitian ini yaitu sebesar 96.99% yaitu diperoleh pada uji fold ke 2. Tingkat presisi tertinggi yang diperoleh pada penelitian ini yaitu sebesar 95.00% yaitu diperoleh pada uji fold ke 2. Maka dapat disimpulkan bahwa metode klasifikasi algoritma decision tree c4.5 dapat digunakan dalam mengidentifikasi pasien penyakit Tuberkulosis.

Kata kunci: Tuberkulosis, Algoritma Decision Tree C4.5

I Pendahuluan

1. Latar Belakang

Di Puskesmas Sukorambi Jember dalam periode tahun 2019 lalu, telah menerima 665 pasien suspect penyakit Tuberkulosis dengan berbagai macam gejala yang berbeda outputnya sesuai panduan yang telah diterbitkan oleh Kementerian Kesehatan. Karena itu Puskesmas adalah pelayanan masyarakat yang dapat diartikan bahwa ujung tombak sistem pelayanan kesehatan di Indonesia karena menyentuh masyarakat sebelum Rumah Sakit sebelum mendapat rujukan. Oleh karena itu data rekam medik pasien yang tersimpan di puskesmas dapat digunakan untuk pengembangan data menggunakan klasifikasi sehingga dapat menciptakan informasi dan menemukan pola dari data sehingga memudahkan petugas mengambil keputusan terkait penyakit Tuberkulosis. Penelitian ini akan menggunakan algoritma C4.5 dalam prediksi TBC. Lalu dari pengimplementasian algoritma C4.5 akan didapatkan pohon keputusan sebuah informasi untuk acuan dalam melakukan setiap prediksi terhadap TBC terutama di daerah sekitar Puskesmas Sukorambi Jember.

2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan sebelumnya, permasalahan dalam tugas akhir ini adalah sebagai berikut :

- Berapa tingkat akurasi terhadap klasifikasi Tuberkulosis (TB) berdasarkan gejala menggunakan algoritma C4.5?
- Berapa tingkat presisi algoritma C4.5 untuk klasifikasi Tuberkulosis (TB) berdasarkan gejala?

3. Tujuan

Tujuan yang akan dicapai dari tugas akhir ini adalah:

- Mengetahui tingkat presisi hasil klasifikasi pasien Tuberkulosis (TB) terhadap gejala

dilihat dari riwayat pasien menggunakan algoritma C4.5.

- Mengukur tingkat akurasi dari hasil perhitungan algoritma C4.5

4. Manfaat

Adapun manfaat dari tugas akhir ini adalah :

- Penelitian ini dapat memberikan informasi untuk petugas puskesmas sebagai bahan acuan terhadap pasien yang mengidap Tuberkulosis melalui gejala yang telah ada melalui algoritma.
- Mengetahui tingkat akurasi algoritma C4.5 terhadap klasifikasi data pasien Tuberkulosis.

5. Batasan Masalah

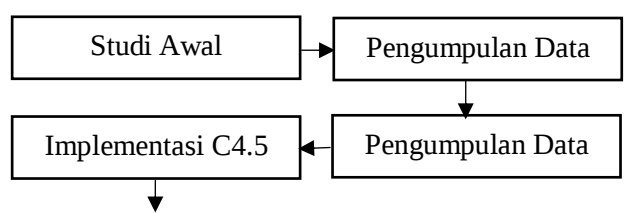
Agar pembahasan ini dapat dilakukan secara terarah, maka perlu dilakukan batasan masalah meliputi :

- Klasifikasi pasien Tuberkulosis (TB) menggunakan algoritma C4.5 dengan parameter antara lain: batuk ≥ 2 Minggu, penurunan berat badan, nafsu makan, demam, batuk darah, nyeri dada,
- Data yang digunakan adalah data pasien Tuberkulosis secara acak diperoleh langsung di Puskesmas Sukorambi Jember pada tahun 2019 bulan Januari sampai Desember.

II Metode Penelitian

1. Kerangka Penelitian

Metodologi penelitian yang dilakukan ini melalui beberapa tahapan yaitu studi awal, pengumpulan data, pra proses data, implementasi metode algoritma C4.5, hasil dan pembahasan, kesimpulan dan saran.



suspect pasien negatif atau tidak terjangkit penyakit Tuberkulosis (TB). Jumlah pasien yang terdaftar dalam dataset berjumlah 665 suspect. Untuk keterangan atribut di jelaskan pada tabel di bawah ini.

Gambar 1. Tahapan Penelitian

2. Studi Awal

Tahapan langkah awal yaitu dengan mempelajari tentang algoritma C4.5 dan mencari informasi informasi tentang penyakit Tuberkulosis terutama berhubungan dengan gejala penyakit tersebut, kemudian mempelajari beberapa literatur, jurnal dan paper yang berkaitan dengan permasalahan dengan latar belakang yang telah dijelaskan dan menemukan informasi yang dicari sesuai penelitian.

3. Pengumpulan Data

Penelitian ini menggunakan dataset yang didapatkan secara langsung dari Puskesmas Sukorambi Jember pada tahun 2019 dengan total jumlah data keseluruhan adalah 665 pasien. Data ini mencakup 6 atribut, yaitu : batuk ≥ 2 minggu, penurunan berat badan, nafsu makan, demam, batuk darah, dan sesak yang akan diolah dengan menggunakan algoritma C4.5.

4. Praproses Data

Dari dataset yang telah didapat lalu dipelajari secara mendalam akan dipilih atribut yang dipakai untuk penelitian. Pada penelitian ini atribut yang dipakai adalah batuk ≥ 2 minggu, penurunan berat badan, nafsu makan, demam, batuk darah, dan sesak. Dengan kesimpulan semua atribut dapat dipakai dengan sesuai pedoman dan data yang telah didapat.

5. Implementasi Metode C4.5

Pada implementasi metode C4.5 dilakukan juga pengkategorian menggunakan matrix confusion, pengukuran akurasi dan presisi serta scenario uji menggunakan cross fold validation.

III Implementasi

1. Implementasi Sistem

Tahap implementasi sistem ini merupakan suatu tahapan menerapkan semua desain sistem yang telah dibuat kedalam tool sehingga dapat digunakan dengan mudah. Maka dilakukan implementasi sistem “Penerapan Algoritma Decision Tree C4.5 Dalam Klasifikasi Identifikasi Pasien Penyakit Tuberkulosis (TB) di Puskesmas Sukorambi Jember”. Sistem ini menggunakan K-Fold Validation menggunakan tool Rapid Miner.

2. Gambaran dataset

Data penelitian yang digunakan diambil dari Puskesmas Sukorambi Jember. Dataset tersebut didapat melalui petugas puskesmas yang menangani tentang Penyakit Tuberkulosis (TB), pada dataset ini memiliki 6 (enam) atribut dan satu output yang terdiri dari dua kelas, yaitu “POSITIF” yang berarti suspect pasien positif terjangkit penyakit Tuberkulosis (TB) dan “NEGATIF” yang berarti

Table 1 Keterangan Atribut

NO	Nama Atribut	Parameter	Keterangan
1	Batuk ≥ 2 minggu	Ya Tidak	Apakah batuk telah lebih dari 2 minggu atau kurang
2	Penurunan Berat Badan	Turun Tidak	Apakah berat badan pasien turun atau tidak
3	Nafsu Makan	Turun Normal	Menurunnya nafsu makan atau tetap normal
4	Demam	Ya Tidak	Pasien mengalami demam atau tidak
5	Batuk Darah	Ya Tidak	Apakah terjadi batuk berdarah atau tidak
6	Sesak	Ya Tidak	Pasien mengalami sesak atau tidak
7	Output	Positif Negatif	Pasien terdiagnosa positif Tuberkulosis (TB) Pasien negatif tidak terdiagnosa Tuberkulosis (TB)

3. Praproses Data

Tahap ini akan mengubah dan menghapus parameter untuk menyesuaikan kebutuhan untuk klasifikasi yang biasa disebut normalisasi data. Dalam penelitian ini atribut yang mendapat preprocessing adalah atribut nama dan umur, karena nama dan umur tidak dipakai dalam klasifikasi.

4. Pemanfaatan Algoritma C4.5

Dalam penerapan algoritma C4.5, penelitian ini menggunakan tool Rapid Miner Studio untuk pengaplikasian. Sebelum data dimasukkan kedalam Rapid Miner, data akan disimpan dalam format CSV supaya memudahkan Rapid Miner untuk membaca data. Proses klasifikasi akan dilakukan pembagian pengujian dengan data testing sesuai

data yang tersaji dimana akan dibagi menjadi 5 bagian yang disetiap bagian terhitung menjadi 133 data. Data testing dapat dilihat pada tabel berikut.

Table 2 skenario uji

P1	P2	P3	P4	P5
1-133	1-133	1-133	1-133	1-133
134-266	134-266	134-266	134-266	134-266
267-399	267-399	267-399	267-399	267-399
400-532	400-532	400-532	400-532	400-532
533-665	533-665	533-665	533-665	533-665

Dapat dilihat bahwa pada tabel diatas sudah terbagi menjadi beberapa bagian. Dalam tahap ini akan dilakukan pengujian dengan data testing yang berbeda seperti P1 dan terletak pada kolom pertama dan pada P2 data testing terletak di kolom 2 dan begitu seterusnya hingga P5.

5. Hasil pengujian

Pada penelitian ini dilakukan scenario pengujian menggunakan cross fold validation dengan nilai k = 5 atau lima kali pengujian. Berikut hasil yang diperoleh.

a. Pengujian 1

Berdasarkan hasil pengujian P1 dengan memasukkan data sebanyak 133 data maka langkah selanjutnya adalah pengujian menggunakan Confusion Matrix untuk mendapatkan hasil akurasi dan presisi.

Aktual	klasifikasi	
	Positif	Negatif
	Positif	Negatif
	17	4
	6	106

Gambar 2. Hasil confusion Matrix P1

Dari gambar 2 dapat dilihat bahwa dari 133 kasus P1 nilai akurasi algoritma C4.5 sebagai berikut:

$$Akurasi = \frac{17+106}{17+106+6+4} = 92.48 \%$$

$$Precision = \frac{17}{17+6} = 73.91 \%$$

b. Pengujian 2

Berdasarkan hasil pengujian P2 dengan memasukkan data sebanyak 133 data maka langkah selanjutnya adalah pengujian menggunakan Confusion Matrix untuk mendapatkan hasil akurasi dan presisi.

Aktual	klasifikasi	
	Positif	Negatif
	Positif	Negatif
	19	3
	1	110

Gambar 3. Hasil confusion Matrix P2

Dari gambar 3 dapat dilihat bahwa dari 133 kasus P2 nilai akurasi algoritma C4.5 sebagai berikut:

$$Akurasi = \frac{19+110}{19+110+1+3} = 96.99 \%$$

$$Precision = \frac{19}{19+1} = 95.00 \%$$

c. Pengujian 3

Berdasarkan hasil pengujian P3 dengan memasukkan data sebanyak 133 data maka langkah selanjutnya adalah pengujian menggunakan Confusion Matrix untuk mendapatkan hasil akurasi dan presisi.

Aktual	klasifikasi	
	Positif	Negatif
	Positif	Negatif
	16	5
	7	105

Gambar 4. Hasil confusion Matrix P3

Dari gambar 4 dapat dilihat bahwa dari 133 kasus P3 nilai akurasi algoritma C4.5 sebagai berikut:

$$Akurasi = \frac{16+105}{16+105+7+5} = 90.98 \%$$

$$Precision = \frac{16}{16+7} = 69.57 \%$$

d. Pengujian 4

Berdasarkan hasil pengujian P4 dengan memasukkan data sebanyak 133 data maka langkah selanjutnya adalah pengujian menggunakan Confusion Matrix untuk mendapatkan hasil akurasi dan presisi.

Aktual	klasifikasi	
	Positif	Negatif
	Positif	Negatif
	24	6
	13	90

Gambar 5. Hasil confusion Matrix P4

Dari gambar 5 dapat dilihat bahwa dari 133 kasus P4 nilai akurasi algoritma C4.5 sebagai berikut:

$$Akurasi = \frac{24+90}{24+90+13+6} = 85.71 \%$$

$$Precision = \frac{24}{24+13} = 64.86 \%$$

e. Pengujian 5

Berdasarkan hasil pengujian P5 dengan memasukkan data sebanyak 133 data maka langkah selanjutnya adalah pengujian menggunakan Confusion Matrix untuk mendapatkan hasil akurasi dan presisi.

Aktual	klasifikasi	
	Positif	Negatif
	Positif	Negatif
	17	6
	5	97

Gambar 6. Hasil confusion Matrix P5

Dari gambar 6 dapat dilihat bahwa dari 133 kasus P4 nilai akurasi algoritma C4.5 sebagai berikut:

$$Akurasi = \frac{17+97}{17+97+5+6} = 91.20 \%$$

$$\text{Precision} = \frac{17}{17+5} = 77.27\%$$

6. Hasil implementasi

Melalui perhitungan diatas dengan menggunakan K-Fold Validation sebanyak 5 Fold dan dengan cara dari total data sebanyak 665 data dibagi 5 yang sehingga selisih antara P1, P2 dan seterusnya sampai P5 sebanyak 133 data. Maka didapatkan hasil presisi dan akurasi untuk P1 sampai dengan P5 sebagai berikut:

- Pengujian pertama untuk P1 dengan data mulai dari data nomor 1 sampai 133 mendapatkan akurasi sebesar 92.48% dan presisi 73.91%.
- Pengujian kedua untuk P2 dengan data mulai dari data nomor 134 sampai 266 mendapatkan akurasi sebesar 96.99% dan presisi 95.00%
- Pengujian ketiga untuk P3 dengan data mulai dari data nomor 267 sampai 399 mendapatkan akurasi sebesar 90.98% dan presisi 69.57%
- Pengujian keempat untuk P4 dengan data mulai dari data nomor 400 sampai 532 mendapatkan akurasi sebesar 85.71% dan presisi 64.86%
- Pengujian kelima untuk P5 dengan data mulai dari data nomor 533 sampai 665 mendapatkan akurasi sebesar 91.20% dan presisi 77.27%.

Dari 5 kali pengujian menggunakan k fold cross valiation diperoleh hasil bahwa atribut “penurunan berat badan” selalu menjadi akar pada kelima pohon kuputusan, hal ini sejalan dengan hasil uji korelasi menggunakan correlation attribute evaluation dengan metode yang digunakan adalah ranker dan software bantu Weka. Berikut hasil pengukuran tingkat korelasi yang diperoleh.

Table 3 Hasil pengukuran korelasi atribut

Ranked	Attributes
0.5761	Penurunan Berat Badan
0.5552	Nafsu Makan
0.4516	Batuk darah
0.0922	Batuk > dari 2 Minggu
0.0474	Demam
0.0439	Sesak

IV Hasil

1. Kesimpulan

Setelah dilakukan analisis dan implementasi dan pengujian pemanfaatan algoritma C4.5 untuk klasifikasi penyakit Tuberkulosis (TB) dengan total data 655 pasien dengan menggunakan 5 kali cross validation uji coba, maka dapat disimpulkan bahwa:

- Tingkat akurasi tertinggi yang diperoleh pada penelitian ini yaitu sebesar 96.99% yaitu diperoleh pada uji fold ke 2.
- Tingkat presisi tertinggi yang diperoleh pada penelitian ini yaitu sebesar 95.00% yaitu diperoleh pada uji fold ke 2.

- Dari penelitian ini diperoleh akurasi rata-rata dari 5 kali uji yaitu sebesar 91.47%.
- Dari penelitian ini diperoleh presisi rata-rata dari 5 kali uji yaitu sebesar 76.12%.

2. Saran

Berdasarkan penelitian yang telah dilakukan terhadap klasifikasi penyakit Tuberkulosis (TB) maka didapatkan saran sebagai berikut:

- Dalam penelitian selanjutnya dapat dikembangkan dengan menggunakan algoritma klasifikasi yang lain untuk dijadikan perbandingan dalam mencari algoritma terbaik.
- Penambahan atribut dan parameter sesuai ketentuan yang telah ditetapkan oleh dinas kesehatan untuk penunjang penelitian.
- Pengembangan membuat program atau aplikasi pendeteksi penyakit Tuberkulosis yang digabungkan dengan metode pembangunan perangkat lunak sesuai algoritma.

Daftar Pustaka

- Ente, D. R. (2020). Klasifikasi Faktor-Faktor Penyebab Penyakit Diabetes Melitus di Rumah Sakit UNHAS Menggunakan Algoritma C4.5.
- Septiani, W. D. (2014). Penerapan Algoritma C4.5 Untuk Prediksi Penyakit Hepatitis.
- Setyo, J. S. (2017). Penerapan Algoritma C4.5 Terhadap Penyakit Tuberkulosis Paru.
- Kusumawardani, Y. M. (2019). Klasifikasi Persalinan Normal atau Caesar Menggunakan Algoritma C4.5.
- Fernitha, Ventya.. (2019). Penerapan Algoritma C4.5 Untuk Deteksi Penyakit Kanker Serviks
- Harama, A. A. K.. (2017). Klasifikasi Penyakit Hipertensi Menggunakan Algoritma C4.5 Studi Kasus RSUD Provinsi NTB.
- Mashlahah, Susi.. (2013). Prediksi Kelulusan Mahasiswa Menggunakan Metode Decision Tree Dengan Penerapan Algoritma C4.5.
- Trihartati, Agustin. (2016). Identifikasi Penyakit Tuberkulosis (TB) Pada Manusia Menggunakan Metode Naïve Bayesian.
- Saiyar, Hafdiarsya. (2018). Aplikasi Diagnosa Penyakit Tuberkulosis Menggunakan Algoritma Data Mining.
- Santosa, Ivan. (2018). Implementasi Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Tuberkulosis (TB).

Santosa, Ivan. (2018). Implementasi Algoritma Decision Tree C4.5 Untuk Diagnosa Penyakit Tuberkulosis (TB).

Febriyanto, Deny. (2018). Prediksi Penyakit Tuberculosis (TBC) Menggunakan Algoritma C4.5.

