

## **PREDIKSI PENYAKIT *TUBERCULOSIS* (TBC) MENGGUNAKAN ALGORITMA C4.5**

Deny Febriyanto<sup>1)</sup>, Yogie Indra Kurniawan<sup>2)</sup>

<sup>1,2)</sup> Program Studi Informatika, Universitas Muhammadiyah Surakarta

<sup>1)</sup> denyfebriant@hotmail.com, <sup>2)</sup> yogie@ums.ac.id

### **ABSTRACT**

*Tuberculosis (TB) is an infectious disease caused by mycobacteria especially Mycobacterium tuberculosis. TB disease remains one of the threats, especially in a country with low and middle economic level. According to World Health Organization (WHO), TB disease became one of the biggest causes of death in the world. Indonesia ranks second with the largest number of TB cases in the world. In addition, there are some symptoms and factors that can cause a person affected by TB disease. The treatment of TB disease can be done intensively but takes a long time. Research in the diagnosis of disease with classification techniques using C4.5 algorithm has been done by several previous researchers and get good results. Therefore, this research will predict TB disease using C4.5 algorithm. with high precision accuracy. The C4.5 algorithm is chosen because it is very easy, fast and has high precision accuracy. The result of the study is an application that can help people to make the diagnosis of TB disease from early age.*

*Keywords : Data Mining, C4.5, Prediction of Disease, Tuberculosis*

### **I. PENDAHULUAN**

Penyakit *Tuberculosis* (TBC) adalah penyakit menular yang disebabkan oleh *mycobacteria*, terutama *Mycobacterium tuberculosis* (Ekata et al, 2016). Penyakit ini dapat menyebar dari orang ke orang melalui udara (Zulfia et al, 2017). Penyakit TBC saat ini masih menjadi salah satu ancaman penyakit yang perlu diwaspadai, terutama di suatu negara dengan perekonomian rendah dan menengah (Cheon et al, 2016).

Menurut *World Health Organization* (WHO) penyakit TBC menjadi salah satu penyebab kematian terbesar di dunia, pada tahun 2015 sebanyak 1.57 juta meninggal karena penyakit TBC (WHO, 2016). Di Indonesia, ditemukan pengidap *Tuberculosis* (TBC) pada tahun 2015 sebanyak 330.910 kasus. Meningkat 1,96 persen dari tahun sebelumnya, sebanyak 324.539 kasus (Kemenkes, 2016). Selain itu, WHO *Tuberculosis Report 2016* menyatakan Indonesia menempati urutan kedua dengan jumlah kasus TBC terbanyak di dunia (WHO, 2016).

Ada beberapa gejala yang dapat menyebabkan timbulnya penyakit TBC diantaranya demam, batuk, nyeri dada, sesak nafas, dahak disertai darah, nafsu makan menurun, berat badan menurun, dan berkeringat di malam hari (Padila, 2013). Selain itu, ada 2 faktor yang mempengaruhi yaitu faktor resiko dan faktor lingkungan. Faktor resiko antara lain HIV, riwayat penyakit diabetes melitus, dan merokok (WHO, 2015). Untuk faktor lingkungan meliputi kelembaban udara dan keadaan rumah (Kemenkes, 2014). Sedangkan untuk mengobati penyakit TBC dapat dilakukan secara intensif. Tetapi membutuhkan waktu lama, yaitu 12 minggu (fase awal) dan 4 – 6 bulan (fase lanjutan) untuk pemulihan (Ethica et al, 2014).

Dapat disimpulkan bahwa penyakit TBC ini tergolong penyakit menular berbahaya terutama di Indonesia. Untuk penyembuhannya membutuhkan waktu yang cukup lama, serta belum adanya alat yang memadai untuk melakukan prediksi penyakit TBC (Deliu & Manea, 2014). Maka dari itu, perlu adanya aplikasi yang dapat digunakan untuk

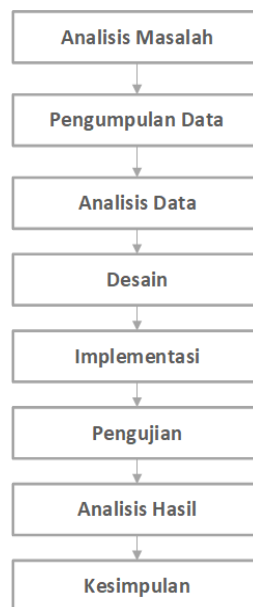
mempermudah prediksi penyakit TBC. Salah satu teknik yang dapat digunakan untuk melakukan prediksi penyakit adalah klasifikasi. Klasifikasi adalah salah satu teknik yang ada dalam data mining, membutuhkan pohon keputusan untuk membuat data dalam grup atau kelas (Adhatrao et al, 2013).

Banyak penelitian prediksi penyakit dengan teknik klasifikasi, diantaranya dilakukan oleh Purushottam et al (2016) dalam prediksi penyakit jantung menggunakan *Decision Tree* (C4.5) dan hasil dari penelitian tersebut algoritma *Decision Tree* (C4.5) menghasilkan nilai terbaik. Selain itu, penelitian lain yang dilakukan Kumar & Umatejaswi (2017) dengan melakukan komparasi 4 algoritma, yaitu *Naive Bayes*, *Random Tree*, C4.5 dan *Simple Logistics* untuk diagnosis penyakit diabetes dengan total 650 kasus dan 14 atribut. Hasil dari penelitian tersebut juga menyatakan algoritma C4.5 menghasilkan nilai terbaik, dengan nilai akurasi 100 dan *error rate* 0. Algoritma C4.5 banyak digunakan di dalam klasifikasi karena sangat mudah di interpretasikan, cepat, serta memiliki akurasi presisi tinggi (Kamagi & Hansun, 2014).

Berdasarkan telaah diatas, dalam penelitian ini akan menggunakan algoritma C4.5 dalam prediksi penyakit TBC. Dari pengolahan algoritma C4.5 didapatkan pohon keputusan, kemudian akan diinterpretasikan kedalam aturan-aturan keputusan (*rules*) yang dapat digunakan sebagai acuan untuk melakukan prediksi penyakit TBC.

## II. METODE PENELITIAN

Metodologi yang akan digunakan dalam penelitian ini dapat dilihat pada Gambar 1.



Gambar 1. Tahapan Metodologi Penelitian

### 2.1 Analisis Masalah

Pada tahap ini bertujuan untuk mengidentifikasi permasalahan-permasalahan yang muncul dari adanya penyakit TBC. Diantaranya, penyakit TBC ini tergolong penyakit menular berbahaya terutama di Indonesia. Untuk penyembuhannya membutuhkan waktu yang cukup lama, serta belum adanya alat yang memadai untuk melakukan prediksi penyakit TBC. Berdasarkan permasalahan yang ada, dibutuhkan aplikasi yang dapat digunakan untuk melakukan prediksi penyakit TBC.

## 2.2 Pengumpulan Data

Data yang akan digunakan dalam penelitian ini yaitu data medis pasien yang didapatkan dari Balai Kesehatan Paru Masyarakat Surakarta, *total* ada 280 kasus dengan jumlah pasien terkena penyakit TBC ada 167 kasus, dan pasien yang tidak terkena TBC ada 113 kasus.

## 2.3 Analisis Data

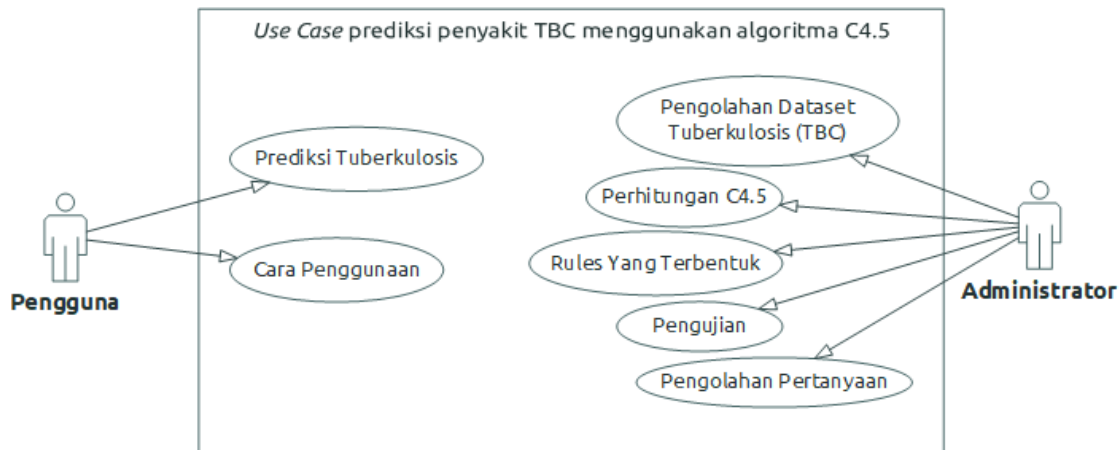
Data yang didapatkan, kemudian dianalisa untuk menentukan variabel – variabel yang digunakan dalam pembuatan aplikasi. Variabel yang diperlukan dalam penelitian ini, dapat dilihat pada Tabel 1.

Tabel 1. Variabel Penelitian

Variabel	Tipe	Value
X1 Merokok	<i>Binomial</i>	YA, TIDAK
X2 Kelembaban Udara	<i>Binomial</i>	LEMBAB, TIDAKLEMBAB
X3 Keadaan Rumah	<i>Binomial</i>	BERSIH, KOTOR
X4 Diabetes Melitus	<i>Binomial</i>	YA, TIDAK
X5 HIV	<i>Binomial</i>	YA, TIDAK
X6 Batuk $\geq$ 3 Minggu	<i>Binomial</i>	YA, TIDAK
X7 Sesak Nafas	<i>Binomial</i>	YA, TIDAK
X8 Nyeri Dada	<i>Binomial</i>	YA, TIDAK
X9 Dahak Disertai Darah	<i>Binomial</i>	YA, TIDAK
X10 Demam	<i>Binomial</i>	YA, TIDAK
X11 Nafsu Makan Menurun	<i>Binomial</i>	YA, TIDAK
X12 Berat Badan Menurun	<i>Binomial</i>	YA, TIDAK
X13 Keringat di Malam Hari	<i>Binomial</i>	YA, TIDAK
Y Terdiagnosa <i>TBC</i>	<i>Binomial</i>	YA, TIDAK

## 2.4 Desain

Tahap ini bertujuan untuk menggambarkan seperti apa rancangan dari aplikasi yang akan dibuat menggunakan algoritma C4.5. Rancangan aplikasi akan diinterpretasikan dalam *use case diagram*. *Use case diagram* dapat dilihat pada Gambar 2.



Gambar 2. Use case diagram aplikasi

Ada 2 aktor yang terlibat ke dalam aplikasi, yaitu Pengguna dan Administrator. Berikut ini penjelasan mengenai *use case diagram* dari kedua aktor diatas :

**a. Pengguna**

1. Prediksi Tuberkulosis  
Pengguna dapat melakukan proses prediksi penyakit TBC.
2. Cara Penggunaan  
Pengguna dapat mengetahui cara penggunaan aplikasi.

**b. Administrator**

1. Pengolahan Dataset Tuberkulosis (TBC)  
Administrator dapat melakukan pengolahan dataset sesuai atribut yang dibutuhkan meliputi *create, read, update, delete* (CRUD) dengan cara *import file excel* ataupun memasukkan data satu – satu.
2. Perhitungan C4.5  
Administrator dapat melakukan proses perhitungan dari *dataset* yang ada menggunakan algoritma C4.5.
3. Rules Yang Terbentuk  
Administrator dapat melihat *rules* yang didapatkan dari proses perhitungan C4.5.
4. Pengujian  
Administrator dapat melakukan pengujian *rules* yang didapatkan menggunakan data training meliputi *create, read, update, delete* (CRUD), dengan cara *import file excel* ataupun memasukkan data satu – satu.
5. Pengolahan Pertanyaan  
Administrator dapat melakukan pengolahan pertanyaan – pertanyaan yang akan diajukan kepada pengguna.

**2.5 Implementasi**

Tahap ini aplikasi mulai dibangun dengan mengimplementasikan desain yang sudah dibuat. Aplikasi akan dibuat dalam model *web based* menggunakan bahasa pemrograman PHP. Sedangkan untuk *Database Management System* (DBMS) akan menggunakan MySQL. Aplikasi ini dapat dibuka menggunakan web browser maupun menggunakan handphone dan tablet yang terhubung ke internet (Kurniawan & Dwiyatmika, 2017). Serta algoritma yang akan digunakan adalah algoritma C4.5.

Menurut Marwana (2014) algoritma C4.5 merupakan algoritma yang digunakan untuk membangun *decision tree* (pengambilan keputusan). Algoritma C4.5 adalah salah satu

algoritma induksi pohon keputusan yaitu ID3 (*Iterative Dichotomiser 3*). ID3 dikembangkan oleh J. Ross Quinlan. Dalam prosedur algoritma ID3, *input* berupa sampel *training*, label *training* dan atribut. Algoritma C4.5 ini merupakan pengembangan dari ID3. Selain itu menurut Harryanto & Hansun (2017) ide dasar dari algoritma ini adalah pembuatan pohon keputusan berdasarkan pemilihan atribut yang memiliki prioritas tertinggi atau dapat disebut memiliki nilai *gain* tertinggi berdasarkan nilai *entropy* atribut tersebut sebagai poros atribut klasifikasi. Kemudian secara rekursif cabang-cabang pohon diperluas sehingga seluruh pohon terbentuk. Untuk memudahkan penjelasan mengenai algoritma C4.5. Berikut ini disertakan contoh dalam Tabel 2. Diambil 10 kasus secara acak, dari data medis pasien yang didapatkan dari Balai Kesehatan Paru Masyarakat Surakarta.

Tabel 2. Contoh Keputusan Penyakit TBC

merokok	kelembabanudara	keadaanrumah	diabetes	HIV	batuk ≥ 3 minggu	sesaknafas	nyeridada	dahak	demam	nafsumakan	berat bedan	keringat	TBC
TIDAK	TIDAKLEMBAB	BERSIH	TIDAK	YA	YA	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	YA
TIDAK	TIDAKLEMBAB	BERSIH	TIDAK	TIDAK	YA	YA	YA	YA	YA	YA	YA	YA	YA
TIDAK	TIDAKLEMBAB	BERSIH	TIDAK	YA	YA	TIDAK	TIDAK	YA	TIDAK	YA	YA	TIDAK	YA
TIDAK	TIDAKLEMBAB	BERSIH	TIDAK	YA	YA	YA	YA	YA	YA	YA	YA	YA	YA
TIDAK	TIDAKLEMBAB	BERSIH	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK	YA	YA	YA	YA	YA
YA	TIDAKLEMBAB	BERSIH	YA	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
YA	TIDAKLEMBAB	KOTOR	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
TIDAK	LEMBAB	KOTOR	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	TIDAK	YA	YA	TIDAK	TIDAK
TIDAK	TIDAKLEMBAB	BERSIH	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK	TIDAK
YA	TIDAKLEMBAB	BERSIH	TIDAK	TIDAK	TIDAK	TIDAK	YA	TIDAK	YA	TIDAK	TIDAK	TIDAK	TIDAK

Dalam contoh kasus yang tertera dalam Tabel 2, akan didapatkan *rules* hasil dari pembuatan pohon keputusan untuk menentukan pasien terkena TBC atau tidak. Berdasarkan apa yang ditulis oleh Hanik Umi (2011), terdapat empat langkah dalam proses pembuatan pohon keputusan pada algoritma C4.5, yaitu:

- Memilih atribut sebagai akar.
- Membuat cabang untuk masing-masing nilai.
- Membagi setiap kasus dalam cabang.
- Mengulangi proses dalam setiap cabang sehingga semua kasus dalam cabang memiliki kelas yang sama.

Data yang dimiliki harus disusun menjadi sebuah tabel berdasarkan kasus sebelum dilakukan perhitungan untuk mencari nilai *entropy* dan *gain*. Berikut ini rumus untuk mencari nilai *entropy* dan *gain*.

$$Entropy(S) = \sum_{j=1}^k - p_j \log_2 p_j \dots\dots\dots(1)$$

Rumus (1) merupakan rumus yang digunakan dalam perhitungan *entropy* yang digunakan untuk menentukan *heterogeneity* dari sebuah kumpulan *data sample* (Amin et al, 2015). Berikut keterangannya :

- S : Himpunan kasus
- k : Jumlah partisi S
- p<sub>j</sub> : Jumlah kasus pada partisi ke-j

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

Rumus (2) merupakan rumus yang digunakan dalam perhitungan *gain* setelah melakukan perhitungan *entropy*. Berikut keterangannya :

- A : Atribut dari dataset
- k : Jumlah partisi S
- S : Himpunan kasus

Dengan mengetahui rumus-rumus diatas, data yang telah diperoleh dapat dimasukkan dan diproses dengan algoritma C4.5 untuk proses pembuatan *decision tree*. Berikut ini adalah penjelasan lebih rinci mengenai masing – masing langkah dalam pembentukan pohon keputusan dengan menggunakan algoritma C4.5 untuk menyelesaikan permasalahan pada Tabel 2.

- a. Menghitung jumlah kasus, jumlah kasus untuk keputusan YA, jumlah kasus untuk keputusan TIDAK, dan *entropy* dari semua kasus. Kasus dibagi berdasarkan atribut yang ada. Setelah itu dilakukan perhitungan *gain* untuk masing – masing atribut. Hasil perhitungan ditunjukkan dalam Tabel 3.

Tabel 3. Contoh Perhitungan Node 1

Node	TOTAL		Jumlah Kasus (S)	YA (S1)	TIDAK (S2)	Entropy	Gain
1			10	5	5	1	
	Merokok						0.396
		YA	3	0	3	0	
		TIDAK	7	5	2	0.863120569	
	Kelembabanudara						0.108
		LEMBAB	1	0	1	0	
		TIDAKLEMBAB	9	5	4	0.99107606	
	Keadaanrumah						0.236
		BERSIH	8	5	3	0.954434003	
		KOTOR	2	0	2	0	
	Diabetes						0.108
		YA	1	0	1	0	
		TIDAK	9	5	4	0.99107606	
	HIV						0.396
		YA	1	0	1	0	
		TIDAK	7	2	5	0.863120569	
	BATUK $\geq$ 3 minggu						1
		YA	5	5	0	0	
		TIDAK	5	0	5	0	
	Sesaknafas						0.396
		YA	3	3	0	0	
		TIDAK	7	2	5	0.863120569	
	Nyeridada						0.03
		YA	5	3	2	0.970950594	
		TIDAK	5	2	3	0.970950594	
	Dahak						0.396
		YA	3	3	0	0	
		TIDAK	7	2	5	0.863120569	
	Demam						0.124
		YA	4	3	1	0.811278124	
		TIDAK	6	2	4	0.918295834	
	Nafsumakan						0.278
		YA	5	4	1	0.721928095	
		TIDAK	5	1	4	0.721928095	
	Beratbadan						0.278
		YA	5	4	1	0.721928095	
		TIDAK	5	1	4	0.721928095	
	Keringat						0.396
		YA	3	3	0	0	
		TIDAK	7	2	5	0.863120569	

Baris *TOTAL* kolom *entropy* pada Tabel 3 dihitung dengan rumus 1, sebagai berikut:

$$Entropy(Total) = (-\frac{5}{10} \times \log_2(\frac{5}{10})) + (-\frac{5}{10} \times \log_2(\frac{5}{10}))$$

$$Entropy(Total) = 1$$

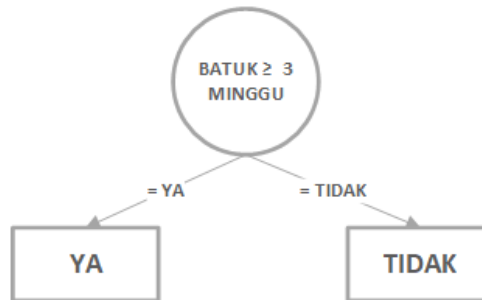
Sementara itu nilai *gain* pada baris Merokok dihitung dengan menggunakan rumus 2, sebagai berikut :

$$Gain(A) = Entropy(Total) - \sum_{i=1}^k \frac{|Merokok_i|}{|Total|} \times Entropy(Merokok_i)$$

$$Gain(A) = 1 - ((\frac{3}{10} \times 0) + (\frac{7}{10} \times 0.863120569))$$

$$Gain(A) = 0.396$$

- b. Dari hasil pada Tabel 3 dapat diketahui bahwa atribut dengan nilai *gain* tertinggi adalah  $BATUK \geq 3$  MINGGU yaitu sebesar 1. Dengan demikian  $BATUK \geq 3$  MINGGU dapat menjadi *node* akar. Ada 2 nilai atribut dari  $BATUK \geq 3$  MINGGU yaitu YA dan TIDAK. Dari kedua nilai atribut tersebut, nilai atribut YA dan TIDAK sudah mengklasifikasikan kasus menjadi satu. Sehingga tidak perlu dilakukan perhitungan lebih lanjut. Dan hasil pohon keputusan yang terbentuk, seperti berikut ini :



Gambar 3. Contoh Pohon Keputusan Hasil Perhitungan *Node* 1

- c. Dengan memperhatikan pohon keputusan pada Gambar 3, diketahui bahwa semua kasus sudah masuk dalam kelas. Dengan demikian, pohon keputusan pada Gambar 3 merupakan pohon keputusan terakhir yang terbentuk. Dari pohon keputusan pada Gambar 3, didapatkan 2 *rules* pada Tabel 4.

Tabel 4. Contoh *rules* yang terbentuk dari pohon keputusan

no	rules yang terbentuk
1	<b>IF</b> $BATUK \geq 3$ MINGGU = YA <b>THEN</b> TBC = YA
2	<b>IF</b> $BATUK \geq 3$ MINGGU = TIDAK <b>THEN</b> TBC = TIDAK

## 2.6 Pengujian

Pada tahap ini, pengujian dilakukan bertujuan untuk mengetahui apakah aplikasi yang telah dibuat sesuai dengan *design* yang direncanakan, menguji fitur – fitur yang ada sesuai sebagaimana semestinya, menguji aplikasi dengan beberapa model data uji untuk membandingkan nilai *precision*, *recall*, dan *accuracy*. Menurut Han et al (2011) untuk menghitung akurasi dapat digunakan tabel *confusion matrix* pada Tabel 5 berikut :

Tabel 5. Panduan Menghitung *Precision*, *Recall*, dan *Accuracy*

Kelas Sebenarnya	Kelas Prediksi		Total
	yes	no	
yes	TP ( <i>True Positive</i> )	FN ( <i>False Negative</i> )	P
no	FP ( <i>False Positive</i> )	TN ( <i>True Negative</i> )	N
Total	P'	N'	P + N

Untuk nilai *precision*, *recall*, dan *accuracy* dapat dihitung dengan rumus sebagai berikut ini:

$$Precision = \frac{TP}{TP+FP} \dots\dots\dots(3)$$

$$Recall = \frac{TP}{TP+FN} \dots\dots\dots(4)$$

$$Accuracy = \frac{TP+TN}{P+N} = \frac{TP+TN}{TP+FN+FP+TN} \dots\dots\dots(5)$$

## 2.7 Analisis Hasil

Setelah pengujian pada aplikasi berhasil dilakukan, selanjutnya akan dilakukan pengukuran tingkat akurasi prediksi dari aplikasi yang telah dibuat.

## 2.8 Kesimpulan

Pada tahap ini, penulis akan memberikan hasil simpulan dari semua proses yang dilakukan dalam penelitian ini.

## III. HASIL DAN PEMBAHASAN

Hasil dan pembahasan dari penelitian yang dilakukan adalah sebagai berikut :

### 3.1 Implementasi Aplikasi

Pada bagian implementasi aplikasi akan dijelaskan beberapa menu yang ada pada aplikasi yang telah dibuat. Aplikasi prediksi penyakit tuberkulosis (TBC) ini terdapat 2 aktor. Dimana tiap aktor memiliki fitur masing – masing, diantaranya sebagai berikut :

#### a. Administrator

##### 1. Dataset Tuberkulosis (TBC)

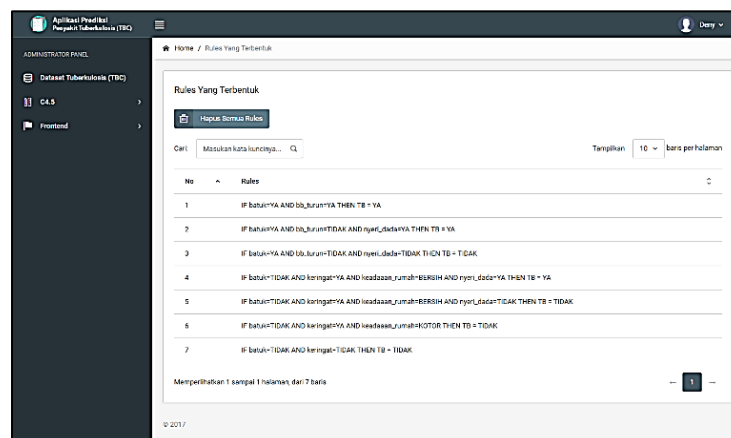
Pada menu *Dataset Tuberkulosis (TBC)*, administrator dapat mengolah dataset tuberkulosis (TBC). Meliputi *edit*, hapus, dan tambah data kasus. Khusus untuk tambah data kasus bisa melalui *import file excel* ataupun menambahkannya satu per-satu.

##### 2. Perhitungan C4.5

Pada bagian perhitungan C4.5 ini, dapat dilakukan proses perhitungan C4.5 dari data *training* yang telah ada. Untuk mendapatkan *rules* yang akan digunakan dalam proses Prediksi Penyakit Tuberkulosis (TBC).

##### 3. Rules Yang Terbentuk

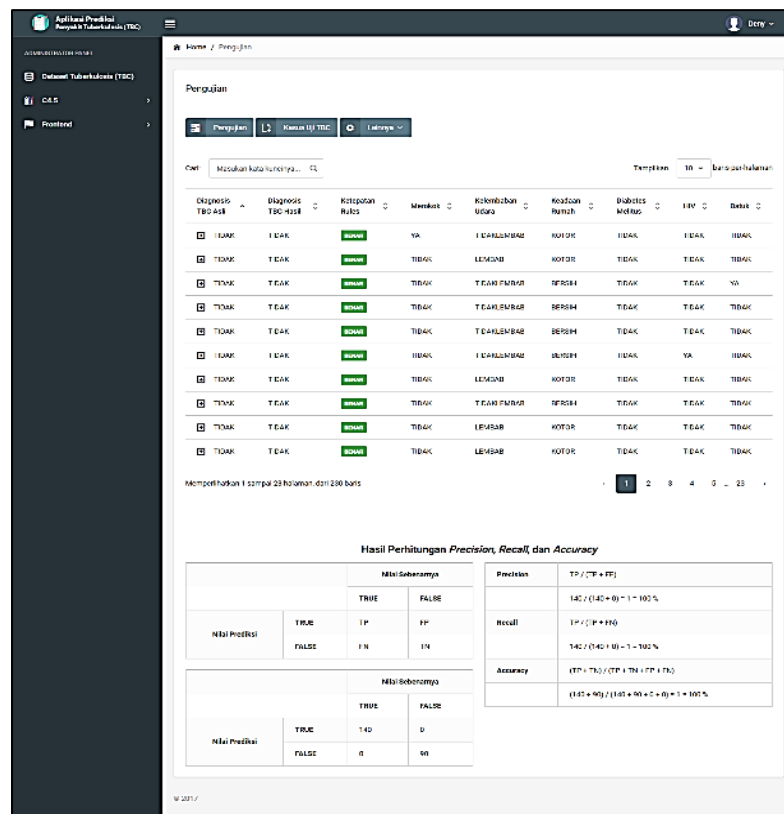
Pada menu ini, akan didapatkan beberapa *rules* atau aturan yang dapat digunakan untuk melakukan Prediksi Penyakit Tuberkulosis (TBC).



Gambar 4. Tampilan Menu *Rules Yang Terbentuk*

4. Pengujian

Pada menu Pengujian, administrator dapat mengolah data uji. Meliputi *edit*, hapus, dan tambah kasus uji. Khusus tambah kasus uji dapat dilakukan dengan 2 cara, yaitu melalui *import file excel* ataupun menambahkan kasus uji satu per-satu. Setelah selesai menambahkan data uji, administrator dapat melakukan pengujian dari data uji yang telah dimasukkan berdasarkan *rules* yang telah diperoleh. Dari proses pengujian, akan diperoleh seberapa ketepatan proses pengujian dan hasil nilai *Precision*, *Recall*, dan *Accuracy*.



Gambar 5. Tampilan Menu Pengujian

5. Pengolahan Pertanyaan

Untuk bagian pengolahan pertanyaan, administrator dapat merubah pertanyaan – pertanyaan yang akan diajukan ke pada pengguna.

b. Pengguna

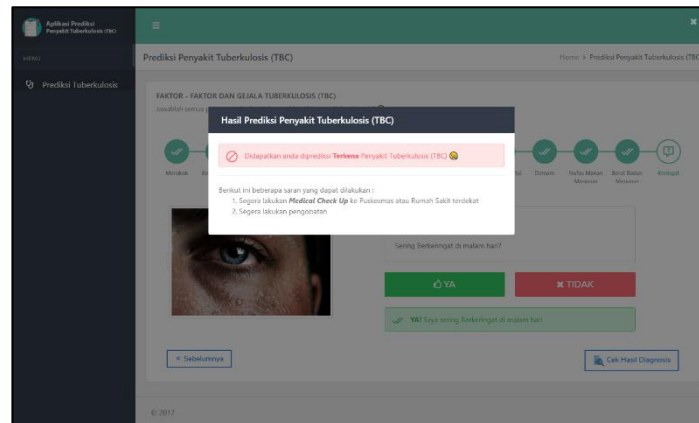
1. Halaman Depan (*frontend*)

Di Halaman Depan (*frontend*) ini, pengguna dapat melihat beberapa menu yang tersedia di aplikasi ini antara lain Beranda, Apa itu Tuberkulosis?, Penanganan Tuberkulosis, Cek Tuberkulosis dan Tentang. Halaman ini dibuat dalam model *one page*.

2. Prediksi Tuberkulosis (TBC)

Pada menu ini, pengguna dapat melakukan prediksi tuberkulosis (TBC) dengan cara menjawab semua pertanyaan yang ada (*total* ada 13 pertanyaan) sesuai kondisi yang dialami pengguna saat ini. Tiap jawaban yang diberikan, akan

digunakan sebagai *input* dalam proses prediksi tuberkulosis berdasarkan *rules* yang telah diperoleh. Setelah menjawab semua pertanyaan yang ada, pengguna akan memperoleh hasil prediksi tuberkulosis (TBC). Apakah pengguna diprediksi terkena penyakit tuberkulosis atau tidak.



Gambar 6. Hasil Prediksi apabila pengguna diprediksi terkena Penyakit Tuberkulosis (TBC)

### 3. Cara Penggunaan

Di bagian menu Cara Penggunaan ini, pengguna dapat melihat seperti apa cara penggunaan dari aplikasi Prediksi Penyakit Tuberkulosis (TBC).

### 3.2 Pengujian

Pada bagian pengujian akan dilakukan beberapa tipe pengujian, meliputi pengujian *blackbox* dan pengujian *Precision, Recall, dan Accuracy*.

#### 3.2.1 Pengujian *Blackbox*

Pada bagian pengujian *blackbox* ini akan dilakukan pengujian mengenai fitur yang ada pada aplikasi, apakah sudah sesuai dengan hasil yang diharapkan atau tidak.

Tabel 6. Hasil Pengujian *Blackbox*

Module/Menu/Fitur	Test Case	Hasil Yang Diharapkan	Hasil
Administrator - Login	User & Password Benar	Muncul Notifikasi Berhasil Login	Sesuai
	User & Password Salah	Muncul Notifikasi Gagal Login	Sesuai
Administrator – Dataset Tuberkulosis (TBC)	Klik Tambah Kasus TBC	Menampilkan Halaman Tambah Kasus TBC	Sesuai
	Klik Import Data TBC	Menampilkan <i>Modals</i> Import Data TBC	Sesuai
	Klik Hapus Kasus TBC	Menghapus Kasus TBC yang terpilih	Sesuai
	Klik Hapus Semua Kasus TBC	Menghapus Semua Kasus TBC yang ada	Sesuai
	Klik Edit Kasus TBC	Menampilkan halaman untuk mengedit Kasus TBC yang terpilih	Sesuai
Administrator Perhitungan C4.5	Klik Perhitungan C4.5	Memproses Kasus TBC yang telah ada, untuk dihitung menggunakan Algoritma C4.5 sehingga didapatkan <i>rules</i> atau aturan.	Sesuai

Module/Menu/Fitur	Test Case	Hasil Yang Diharapkan	Hasil
Administrator – Rules Yang Terbentuk	Klik Hapus Semua Rules	Menghapus rules yang telah diperoleh dari proses perhitungan C4.5	Sesuai
Administrator Pengujian	Klik Tambah Kasus Uji TBC	Menampilkan halaman tambah kasus uji TBC	Sesuai
	Klik Import Data Uji TBC	Menampilkan Modals import data uji TBC	Sesuai
	Klik Hapus Data Uji TBC	Menghapus data uji TBC yang terpilih	Sesuai
	Klik Hapus Semua Data Uji TBC	Menghapus semua data uji TBC yang ada	Sesuai
	Klik Edit Data Uji TBC	Menampilkan halaman untuk mengedit data uji TBC	Sesuai
	Klik Pengujian	Melakukan Proses Pengujian ke data kasus uji TBC yang telah ada, sehingga mendapatkan hasil Ketepatan Prediksi. Serta nilai Precision, Recall, Accuracy dan Error Rate	Sesuai
Administrator – Pertanyaan Yang Ada	Klik Pertanyaan Yang Ada	Menampilkan halaman berisi pertanyaan dari tiap topik yang akan diajukan ke Pengguna, dan dapat melakukan edit pertanyaan tersebut.	Sesuai
Pengguna – Prediksi Tuberkulosis	Klik Opsi Jawaban di tiap topik pertanyaan	Muncul Message, mengenai jawaban yang dipilih	Sesuai
	Klik Selanjutnya	Untuk lanjut ke pertanyaan selanjutnya	Sesuai
	Klik Sebelumnya	Untuk balik ke pertanyaan sebelumnya	Sesuai
	Klik Cek Hasil Diagnosis	Untuk memproses jawaban yang telah dipilih pengguna, sehingga didapatkan hasil prediksi Tuberkulosis	Sesuai
Pengguna – Cara Penggunaan	-	Menampilkan halaman cara penggunaan aplikasi	Sesuai

### 3.2.2 Pengujian *Precision*, *Recall*, dan *Accuracy*

Dalam pengujian ini dilakukan dengan membagi data *training* dan data *testing* yang telah diperoleh dari Balai Kesehatan Paru Masyarakat Surakarta berjumlah 280 kasus menjadi beberapa bagian secara acak. Di tiap bagian terdapat 3 kali proses pengujian, dan hasil yang didapatkan adalah nilai rata-rata dari ke 3 proses pengujian tersebut.

Tabel 7. Hasil Pengujian *Precision*, *Recall*, dan *Accuracy*

Data Training	Data Testing	Precision	Recall	Accuracy
50	230	97,67%	100%	98,67%
100	180	99%	100%	99,3%
150	130	100%	100%	100%
200	80	100%	100%	100%
250	30	100%	100%	100%

Dari hasil pengujian *Precision*, *Recall*, dan *Accuracy* yang dilakukan didapatkan hasil pada Tabel 7 bahwa, pada pengujian dengan jumlah data *training* 250 dan data *testing* 30 diperoleh nilai *Precision* 100%, *Recall* 100% dan *Accuracy* 100%. Berikut ini *rules* yang terbentuk pada proses pengujian ini :

Tabel 8. *Rules* Yang Terbentuk Pada Pengujian Dengan Jumlah Data *Training* 250 kasus

<b>Rules ke</b>	<b>Detail Rules</b>
1	IF (BATUK $\geq$ 3 MINGGU = YA) AND (BB TURUN = YA) THEN TBC = YA
2	IF (BATUK $\geq$ 3 MINGGU = YA) AND (BB TURUN = TIDAK) AND (NYERI DADA = YA) THEN TBC = YA
3	IF (BATUK $\geq$ 3 MINGGU = YA) AND (BB TURUN = TIDAK) AND (NYERI DADA = YA) THEN TBC = TIDAK
4	IF (BATUK $\geq$ 3 MINGGU = TIDAK) AND (KERINGAT MALAM = YA) AND (KEADAAN RUMAH = BERSIH) AND (NYERI DADA = YA) THEN TB = YA
5	IF (BATUK $\geq$ 3 MINGGU = TIDAK) AND (KERINGAT MALAM = YA) AND (KEADAAN RUMAH = BERSIH) AND (NYERI DADA = TIDAK) THEN TB = TIDAK
6	IF (BATUK $\geq$ 3 MINGGU = TIDAK) AND (KERINGAT MALAM = YA) AND (KEADAAN RUMAH = KOTOR) THEN TB = TIDAK
7	IF (BATUK $\geq$ 3 MINGGU = TIDAK) AND (KERINGAT MALAM = TIDAK) THEN TB = TIDAK

Maka dari itu pada pengujian menggunakan data *training* 250 kasus dan data *testing* 30 kasus diperoleh hasil terbaik dengan nilai *Precision*, *Recall*, dan *Accuracy* maksimal dapat mencapai 100%.

### 3.3 Analisa Hasil

Aplikasi prediksi penyakit Tuberkulosis (TBC) yang dibuat dengan menerapkan algoritma C4.5 memiliki 2 aktor utama yaitu administrator dan pengguna dengan hak akses yang berbeda. Administrator dapat mengelola semua data yang ada pada aplikasi meliputi pengolahan *dataset* tuberkulosis, perhitungan C4.5, melihat *rules* yang terbentuk, pengujian *rules*, dan pengolahan pertanyaan ke pengguna. Sedangkan untuk hak akses pengguna dapat melakukan prediksi penyakit tuberkulosis dengan cara menjawab beberapa pertanyaan yang diajukan. Dari jawaban yang diberikan oleh pengguna, aplikasi akan melakukan proses klasifikasi berdasarkan *rules* yang sudah terbentuk. Dari proses klasifikasi ini, akan didapatkan hasil apakah pengguna diprediksi terkena tuberkulosis atau tidak.

Pengujian dilakukan dengan 2 cara, yaitu pengujian *blackbox* dan pengujian *Precision*, *Recall*, dan *Accuracy*. Pada pengujian *blackbox* menunjukkan bahwa semua modul yang ada pada aplikasi dapat berjalan sesuai dengan yang diharapkan. Pada pengujian *Precision*, *Recall*, dan *Accuracy* yang dilakukan dengan menggunakan data yang diperoleh dari Balai Kesehatan Paru Masyarakat Surakarta berjumlah 280 kasus menunjukkan bahwa nilai rata – rata maksimal yang didapatkan mencapai 100%. Selain itu dari semua pengujian *Precision*, *Recall*, dan *Accuracy* yang telah dilakukan didapatkan bahwa semakin banyaknya data *training* maka nilai *Precision*, *Recall*, dan *Accuracy* yang diperoleh juga semakin meningkat. Sedangkan dari *rules* yang diperoleh didapatkan sebuah *strong* variabel yaitu Batuk  $\geq$  3 Minggu dikarenakan variabel tersebut berada pada *root node* dari semua *rules* yang diperoleh.

#### IV. KESIMPULAN DAN SARAN

##### 4.1 Kesimpulan

Dari hasil penelitian yang telah dilakukan, didapatkan beberapa kesimpulan sebagai berikut :

1. Algoritma C4.5 berhasil digunakan untuk melakukan Prediksi Penyakit Tuberkulosis (TBC), serta aplikasi yang dibuat berjalan dengan baik dan hasil yang didapatkan cukup baik.
2. Pada pengujian *blackbox* yang dilakukan, semua menu modul yang ada pada aplikasi berjalan dengan sangat baik, serta sesuai dengan fungsi yang diharapkan.
3. Pada pengujian *Precision*, *Recall*, dan *Accuracy* menunjukkan bahwa semakin banyaknya data *training* maka nilai *Precision*, *Recall*, dan *Accuracy* yang diperoleh juga semakin meningkat. Nilai rata – rata maksimal dari pengujian *Precision*, *Recall*, dan *Accuracy* yang didapatkan mencapai 100%.

##### 4.2 Saran

1. Memperbanyak sebaran data yang dapat diperoleh dari rumah sakit lainnya.
2. Untuk dapat mengoptimasi aplikasi ini, dapat digunakan metode yang berbeda ataupun digabungkan dengan metode lainnya. Supaya proses *mining* dan hasil yang didapatkan bisa lebih baik lagi.

#### DAFTAR PUSTAKA

- Adhatrao, K., Gaykar, A., Dhawan, A., Jha, R., Honrao, V. (2013). Predicting Students' Performance Using ID3 And C4.5 Classification Algorithms. *International Journal of Data Mining & Knowledge Management Process*, 3(5), 39-52. Doi: 10.5121/ijdkp.2013.3504
- Amin, R. K., Indwiarti, Sibaroni, Y. (2015). Implementation of Decision Tree Using C4.5 Algorithm in Decision Making of Loan Application by Debtor (Case Study: Bank Pasar of Yogyakarta Special Region). *Tersedia dari Information and Communication Technology (ICoICT) : 2015 3rd International Conference*, Nusa Dua, Bali. IEEE. Doi: 10.1109/ICoICT.2015.7231400
- Cheon, S. A., Cho, H. H., Kim, J., Lee, J., Kim, H., Park, T. J. (2016). Recent tuberculosis diagnosis toward the end TB strategy. *Journal of Microbiological Methods*, 123, 51-61. Doi: 10.1016/j.mimet.2016.02.007
- Deliu, I., Manea, E. (2014). The Incidence Of Mycobacterium Tuberculosis Infections And Resistance To AntiTuberculosis Drugs In Giurgiu Population. *Current Trends in Natural Sciences*, 3(6), 27-31.
- Ekata., Tyagi, P. K., Gupta, N. K., Gupta, S. (2016). Diagnosis of Pulmonary Tuberculosis using Fuzzy Inference System. *Innovative Applications of Computational Intelligence on Power, Energy and Controls with their Impact on Humanity (CIPECH-16)*, 3-7. Doi: 10.1109/CIPECH.2016.7918726
- Eticha, B. M., Atomsa, A., Tsehaineh, B., Berheto, T. M. (2014). Patients' perspectives of the quality of tuberculosis treatment services in South Ethiopia. *American Journal of Nursing Science*, 3(4), 48-55. Doi: 10.11648/j.ajns.20140304.12
- Hanik, U. (2011). *Fuzzy Decision Tree dengan Algoritma C4.5 pada Data Diabetes Indian Pima*, Institut Teknologi Sepuluh November. Surabaya.
- Han, J., Kamber, M., Pei, J. (2011). *Data mining : concepts and techniques (3rd ed.)*. USA: Morgan Kaufmann.

- Harryanto, F., Hansun, S. (2017). Penerapan Algoritma C4.5 untuk memprediksi Penerimaan Calon Pegawai Baru di PT WISE. *Jatiji*, 3(2), 95-103.
- J. R. Quinlan, C4.5: Programs for Machine Learning, San Mateo: Morgan Kaufmann, 1993.
- Kamagi, D. H., Hansun, S. (2014) Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa. *ULTIMATICS*, 6(1), 15-20.
- Kementrian Kesehatan Republik Indonesia. (2014). Pedoman Nasional Pengendalian Tuberkulosis. Tulisan pada : <http://spiritia.or.id/dokumen/pedoman-tbnasional2014.pdf> [Diakses pada 17 oktober 2017]
- Kementrian Kesehatan Republik Indonesia. (2016). Berapa Kasus TBC di Indonesia?. Tulisan pada : <http://databoks.katadata.co.id/datapublish/2017/03/22/berapa-kasus-tbc-di-indonesia> [Diakses pada 17 oktober 2017]
- Kumar, P. S., Umatejaswi, V. (2017). Diagnosing Diabetes using Data Mining Techniques. *International Journal of Scientific and Research Publications*, 7(6), 705-709.
- Kurniawan, Y. I., Dwiyatmika, W. (2017). Aplikasi Diagnosa Retardasi Mental Pada Anak. Tersedia dari Prosiding SEMNAS Penguatan Individu di Era Revolusi Informasi.
- Marwana. (2014). Algoritma C4.5 untuk Simulasi Prediksi Kemenangan Dalam Pertandingan Sepakbola, *STIMED*, Nusa Palapa, Makassar.
- Padila. (2013). Asuhan Keperawatan Penyakit Dalam. Nuha Medika. Yogyakarta.
- Purushottam., Saxena, K., Sharma, R. (2016). Efficient Heart Disease Prediction System using Decision Tree. *International Conference on Computing, Communication and Automation (ICCCA2015)*, 85, 962-969. Doi:10.1016/j.procs.2016.05.288
- Sandi, G., Supangkat, S. H., Slamet C. (2016). Health risk prediction for treatment of hypertension. Tersedia dari *Cyber and IT Service Management: International Conference*, Bandung, Indonesia. IEEE. Doi : 10.1109/CITSM.2016.7577584
- World Health Organization. (2015). TB comorbidities and risk factors. Tulisan pada : <http://www.who.int/tb/areas-of-work/treatment/risk-factors/en/> [Diakses pada 17 oktober 2017]
- World Health Organization. (2016). The top 10 causes of death. Tulisan pada : <http://www.who.int/mediacentre/factsheets/fs310/en/> [Diakses pada 17 oktober 2017]
- World Health Organization. (2016). Global tuberculosis report 2016. Tulisan pada : [http://www.who.int/tb/publications/global\\_report/en/](http://www.who.int/tb/publications/global_report/en/) [Diakses pada 17 oktober 2017]
- Zulvia, F., Kuo, R., Roflin, E. (2017). Efficient Heart Disease Prediction System using Decision Tree. *IEEE*, 204-209. Doi:10.1109/COMPSAC.2017.57.