

## PENERAPAN ALGORITMA NAÏVE BAYES UNTUK PREDIKSI PENYAKIT TUBERCULOSIS (TB)

**Rizal Amegia Saputra**

Program Studi Manajemen Informatika  
AMIK BSI Sukabumi

Jl. Cemerlang No. 8 Sukakarya, Sukabumi

E-mail: [rizal.rga@bsi.ac.id](mailto:rizal.rga@bsi.ac.id)

### ABSTRACT

*The disease Tuberculosis (TB) is a contagious and deadly diseases in the world, even the World Health Organization (WHO) declared as the world's emergency disease (global emergency), some research fields of health including one disease TB has been widely carried out to detect the disease early, but it is not yet known which algorithm is quite good in predicting disease TB. On this research will apply the Algorithm Naïve Bayes, in predicting diagnosis of TB disease to Naïve Bayes algorithm so that the destination is the most accurate algorithm in the prediction of the disease TUBERCULOSIS. The test results using the method of Confusion Matrix and the ROC Curve, the naïve bayes algorithm is known that has a value of 91,61%, accuracy and value of the AUC of 0,995. See the value of AUC, the naïve bayes methods including group classification is very good, because the results of his AUC values between 0.90-1.00.*

*Keywords: Naïve Bayes, Tuberculosis.*

### I. PENDAHULUAN

Pencemaran polusi udara yang disebabkan dari pertumbuhan industri, mobil dan lainnya, itu sudah mencapai tingkat yang sangat tinggi, dan penduduk setempat biasanya mengalami berbagai macam penyakit pernafasan termasuk beberapa penyakit paru yang berbahaya salah satunya yaitu Tuberkulosis (TB)[1], penyakit TB adalah penyakit menular yang disebabkan oleh bakteri yang disebut *Mycobacterium tuberculosis* dan merupakan penyebab kematian paling tinggi yang terjadi pada usia produktif 15-50 tahun, kelompok ekonomi lemah, dan berpendidikan rendah [2].

Penyakit TB merupakan masalah penting untuk kesehatan masyarakat didunia, bahkan penyakit ini termasuk salah satu penyakit yang dicanangkan oleh *World Health Organization* (WHO) sebagai kedaruratan dunia (*global emergency*), berdasarkan laporan WHO tahun 2009, indonesia termasuk negara terbesar ketiga setelah India dan Cina yang sebagian penduduknya terkena penyakit TB, dan setiap tahunnya di indonesia terjadi sekitar 245.000 kasus tuberkulosis baru, dengan jumlah tuberkulosis menular (BTA+) sejumlah 107.000 kasus, sedangkan kematian karena TB sekitar 46.000 setiap tahunnya [3]. Tahun 2007 di Indonesia penyakit TB merupakan salah satu

dari sepuluh besar penyakit penyebab kematian terbesar di indonesia[4].

Barikut daftar tabel penyebab kematian terbesar yaitu:

Tabel 1.1. Daftar Penyebab Kematian Terbesar Tahun 2007[5],[4]:

No	Penyebab Kematian	Proporsi Kematian %
1	Stroke	15,4
2	Tuberculosis	7,5
3	Hipertensi	6,8
4	Kecelakaan Lalu Lintas	6,5
5	Perinatal	6
6	Diabetes Melitus	5,7
7	Tumor Ganas	5,7
8	Penyakit Hati	5,1
9	Jantung Koroner	5,1
10	Penyakit Saluran Nafas Bawah	5,1

Tabel 1.1 menunjukkan penyakit TB menjadi urutan kedua, jumlah kasus TB meningkat dan banyak yang tidak berhasil disembuhkan, terutama pada negara yang dikelompokkan dalam 22 negara dengan masalah TB besar (*high burden countries*) dan indonesia termasuk diantaranya [6].

Klasifikasi data penyakit TB pada medis merupakan tugas penting dalam memprediksi penyakit, bahkan dapat membantu dokter dalam mengambil keputusan diagnosis penyakit tersebut[5], dengan demikian sangat penting melakukan diagnosis secara dini agar dapat mengurangi penularan TB kepada masyarakat luas. Banyak peneliti yang sudah melakukan kegiatan dalam memprediksi penyakit TB dengan metode klasifikasi *data mining*, namun belum diketahui metode apa yang paling akurat dalam memprediksi penyakit TB.

Pada penelitian ini akan menerapkan Algoritma *Naïve Bayes*, dalam memprediksi diagnosis penyakit TB dengan tujuan agar algoritma *Naïve Bayes* merupakan algoritma yang paling akurat dalam prediksi penyakit TB, sehingga dapat melakukan diagnosa penyakit TB secara dini [8], algoritma *naïve bayes* adalah salah satu metode *data mining* termasuk kedalam sepuluh klasifikasi *data mining* yang paling populer diantara algoritma-algoritma lainnya [8].

## II. LANDASAN TEORI

### 2.1 Tuberculosis (TB)

TB (*Tuberculosis*) adalah suatu penyakit infeksi yang disebabkan oleh bakteri *mycobacteria tuberculosis*. Bakteri ini merupakan bakteri yang sangat kuat sehingga memerlukan waktu lama untuk mengobatinya. Bakteri ini lebih sering menginfeksi organ paru-paru (90%) dibandingkan mengenai organ tubuh lainnya [9].

### 2.2 Data Mining

*Data mining* adalah suatu disiplin ilmu yang bertujuan untuk menemukan, menggali atau menambahkan pengetahuan dari data atau informasi yang kita miliki[8]. Menurut *Gartner Group* menyebutkan bahwa *Data Mining* adalah proses menelusuri pengetahuan baru, pola dan tren yang dipilih dari jumlah data yang besar yang disimpan dalam repositori atau tempat penyimpanan dengan menggunakan teknik pengenalan pola serta statistik dan tehnik matematika [11].

### 2.3 Algoritma Klasifikasi Data Mining

Klasifikasi *Data mining*[12] adalah suatu metode pembelajaran, untuk memprediksi nilai dari sekelompok atribut dalam menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui.

### 2.4 Naïve Bayes

*Naive Bayes* merupakan metode yang tidak memiliki aturan, *Naive Bayes* menggunakan cabang matematika yang dikenal dengan teori *probabilitas* untuk mencari peluang terbesar dari kemungkinan klasifikasi, dengan cara melihat frekuensi tiap klasifikasi pada *data training*. Klasifikasi *Naive Bayes* adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi *probabilitas* keanggotaan suatu *class*. Klasifikasi *bayesian* didasarkan pada teorema *Bayes*, diambil dari nama seorang ahli matematika yang juga menteri Prebysterian Inggris, Thomas Bayes (1702-1761) [13]. Klasifikasi bayesian memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network*[14].

*Bayes rule* digunakan untuk menghitung probabilitas suatu *class*. Algoritma *Naive Bayes* memberikan suatu cara mengkombinasikan peluang terdahulu dengan syarat kemungkinan menjadi sebuah formula yang dapat digunakan untuk menghitung peluang dari tiap kemungkinan yang terjadi. Bentuk umum dari teorema *bayes* seperti dibawah ini:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (2.1)$$

Dimana:

X :Data dengan class yang belum diketahui

H :Hipotesis data X merupakan suatu class spesifik.

$P(H|X)$  :Probabilitas hipotesis H berdasar kondisi X (*posteriori probability*)

$P(H)$  :Probabilitas hipotesis H (*prior probability*)

$P(X|H)$ :Probabilitas X berdasar kondisi pada hipotesis H

$P(X)$  :Probabilitas dari X

*Naïve bayes* adalah penyederhanaan metode bayes. *Teorema bayes* disederhanakan menjadi:

$$P(H|X)=P(X|H)P(X) \quad (2.4)$$

*Bayes rule* diterapkan untuk menghitung posterior dan probabilitas dari data sebelumnya. Dalam analisis *bayesian*, klasifikasi akhir dihasilkan dengan menggabungkan kedua sumber informasi (*prior* dan *posterior*) untuk menghasilkan probabilitas menggunakan aturan *bayes*.

### III. METODOLOGI PENELITIAN

#### 3.1. Jenis Penelitian

Didalam penelitian ini digunakan metode penelitian *kuantitatif* dan desain riset yang digunakan yaitu *eksperimen komparatif*.

#### 3.2. Metode Pemilihan Populasi dan Sampel

Populasi dalam penelitian ini merupakan pasien TB positif dan pasien TB negative tahun 2011-2013 yang berasal dari data internal Puskesmas Karawang Sukabumi, dan Sampel yang digunakan yaitu data rekap medik penderita penyakit TB Positif dan TB Negatif, data tersebut bersifat intern yang belum dipublikasikan oleh pihak puskesmas sebanyak 300 sampel yang digunakan, 200 sampel pasien positif TB dan 100 Sampel pasien negative TB.

#### 3.3. Metode Pengumpulan Data

Dalam penelitian ini metode pengumpulan data untuk menggunakan data sekunder. Data utama diperoleh dari data rekap medik puskesmas pasien yang positif TB dan pasien yang negatif TB, sedangkan data pendukung lainnya didapat dari buku, jurnal dan publikasi lainnya.

#### 3.4. Instrumen Penelitian

Instrumen penelitian yaitu sebagai berikut:

1. Penelitian ini menggunakan Data sekunder berupa data pasien yang menderita positif TB dan negatif TB yang akan digunakan sebagai instrumen guna memperoleh data dalam proses mendiagnosa penyakit TB.
2. Data disajikan dalam bentuk Tabulasi model dan variabel masing-masing sebanyak 300 pasien terdiri dari pasien TB sebanyak 200 dan pasien bukan TB 100 yang diambil dari tahun 2011/2013.
3. Perangkat lunak yang digunakan untuk menganalisis adalah *Rapidminer* dan *Graphical User Interface* (GUI) untuk menguji *rule* algoritma terpilih adalah PHP.

#### 3.5. Metode Analisis dan Pengujian Data

Teknik Analisis data menggunakan data *Kuantitatif* berupa matematika terhadap angka atau numerik dan nominal. Pada penelitian ini, analisa data dilakukan melalui data rekap medik puskesmas dengan nilai rata-rata pasien TB dan bukan TB, data diolah dan di uji dalam pengujian pada algoritma *Naïve Bayes*. Kemudian pengujian *Rule* yang diperoleh algoritma *Naïve Bayes* tersebut kemudian diuji dengan *Confusion Matrix* dan *Kurva Receiver Operating Characteristic* (ROC) untuk mengukur tingkat akurasi yang akan dihasilkan.

Dengan pengujian diatas dapat diperoleh *rule* algoritma yang mempunyai akurasi yang lebih baik. Sehingga dapat diterapkan pada *Graphical User Interface* (GUI) dengan baik.

Selanjutnya aplikasi yang dibuat akan dievaluasi untuk menghasilkan pengetahuan (*knowledge*) baru.

Dalam penelitian ini pengujian data menggunakan metode *eksperimen* dengan model *Cross Industry Standard Process for Data Mining* (CRISP-DM) yang terdiri dari enam tahap, yaitu [11]:

#### 1. Tahap Pemahaman Bisnis (*Business Understanding*)

Berdasarkan dari hasil laporan puskesmas tahun 2011-2013 yang menderita penyakit TB itu sekitar 300 jiwa orang, untuk mengurangi jumlah resiko penyakit TB positif harus dilakukan pemeriksaan terhadap pasien agar dapat diketahui penyakit secara dini, hal ini dapat dibantu dengan memanfaatkan teknik klasifikasi pada *data mining* model *Naïve Bayes*.

#### 2. Tahap Pemahaman Data (*Data Understanding*)

Data yang digunakan adalah data sekunder, yang didapat dari data rekap medik puskesmas, didalam data tersebut dapat diketahui status pasien yang positif TB dan negatif TB yang terdiri dari 9 atribut *predictor* dan 1 atribut hasil. Atribut-atribut yang menjadi parameter terlihat pada Tabel 3.

Tabel 3.1 Atribut dan Nilai Kategori Penyakit TB

Atribut	Kategori	Nilai Konversi RapidMiner
Lama Batuk	Tidak Batuk	3
	1 Minggu - < 1 Bulan	2
	1 Bulan - < 3 Bulan	1
	> 3 Bulan	0
Jenis Batuk	Tidak Batuk	3
	Batuk Kering	2
	Batuk Berdahak	0
	Batuk Berdarah	1
Keringat Malam	Tidak	1
	Ya	0
Lemas	Tidak Lemas	2
	1-2 Minggu	0
	> 2 Minggu	1
Sesak Nafas	Tidak Sesak Nafas	3
	< 1 Minggu	1
	1 Minggu - 1 Bulan	0
	> 1 Bulan	2
Sakit Dinding Dada	Tidak Sakit Dada	2
	Sakit Dada Setiap Batuk	0
	Sakit Dada Terus	1
Demam	Tidak Demam	3
	kadang-kadang	2
	Agak Demam	0
	Demam Panas Sekali	1
Hilang Napsu Makan	Tidak	1
	Ya	0
Berat Badan Menurun	Tidak	1
	Ya	0
Kelas	Positif	
	Negatif	

#### 3. Tahap Data Persiapan (*Data Preparation*)

Jumlah data yang diperoleh pada penelitian ini sebanyak 300 *record*, baik pasien TB dan pasien tidak TB, akan tetapi data tersebut masih mengandung duplikasi dan anomali atau inkonsisten data. Untuk mendapatkan data yang berkualitas, ada beberapa teknik *preprocessing* yang digunakan yaitu [15]:

a. *Data Validation*, untuk mengidentifikasi dan menghapus anomali dan inkonsistensi data. Anomali dan inkonsistensi data dapat terjadi dikarenakan adanya *missing values* dan juga karena data yang dipakai mengandung nilai-nilai yang salah atau disebut *noise* atau *outlier*.

b. *Data Integration and Transformation*, untuk meningkatkan akurasi dan efisiensi algoritma. Data yang digunakan dalam penulisan ini kategorikal.

c. *Data size reduction and discretization*, untuk memperoleh *data set* dengan jumlah atribut dan *record* yang lebih sedikit tetapi bersifat *informative*. Didalam data *training* yang digunakan dalam penelitian ini, dilakukan penghapusan data duplikasi angka menggunakan *tools* SPSS17.0.

Setelah dilakukan *preprocessing* data yang didapat dari data internal Puskesmas sebanyak 300 *record* direduksi dengan menghilangkan duplikasi menjadi 261 *record*, 92 *record* Negatif TB dan 169 Positif TB, setelah itu data akan dibagi menjadi dua bagian yaitu 70% data untuk proses *Tranning* dan 30% data untuk proses *Testing*. Pemilihan sample ini menggunakan *tools* SPSS17.0 untuk memilih data mana yang akan terpilih untuk data *tranning* dan data *testing*[14], pemilihan data ini menghasilkan data secara *random* berdasarkan *variate Bernoulli* dengan kemungkinan parameter (*probability parameter*) 0,7. Dengan modifikasi ini akan mendapatkan nilai +1 atau -1, Nilai positif digunakan untuk kasus *training* dan nilai negatif digunakan untuk kasus *testing*. Pada *variate Bernoulli* ini digunakan variabel *Partition* dengan rumus  $2 * rv.bernoulli(0.7)$ -1. Hasil dari perhitungan diatas didapat 183 untuk data *tranning* dan 78 untuk data *testing*.

4. Tahap Pemodelan (*Modelling*)

Tahap ini juga disebut tahap *learning* karena pada tahap ini data *training* diklasifikasikan oleh model dan kemudian menghasilkan sejumlah aturan. Pada penelitian ini, pembuatan model menggunakan algoritma *Naïve Bayes*.

5. Tahap Evaluasi (*Evaluation*)

Pada tahap ini dilakukan pengujian terhadap model-model untuk mendapatkan informasi model yang akurat. Evaluasi dan validasi menggunakan metode *Confision Matrix* dan *kurva ROC*.

6. Tahap *Deployment*

Pada tahap ini diterapkan algoritma *Naïve Bayes* pada GUI yang akan diterapkan pada instansi terkait.

3.6. Uji Ketahanan GUI

1.Kuesioner Metode SQA (*Software Quality Assurance*).

Untuk memiliki perangkat lunak yang kualitas standar minimal, maka diterapkan salah satu metode untuk pengukuran kualitas perangkat lunak secara kuantitatif yaitu metode SQA (*Software Quality Assurance*) [16]. Ada sepuluh orang yang akan menguji ketahanan GUI dengan cara memberikan nilai kuesioner yaitu Admin Puskesmas, Dokter, Admin Apotek, Developer, Programmer dan Pasien.

1. Cara Hasil Evaluasi GUI

Pada penelitian ini dalam mengevaluasi GUI, menggunakan pendekatan sebagai berikut [16]:

$$\text{Skor} < \text{Skor Auditability} > * 0.10 + \text{masing-masing} < \text{Skor Accuracy} > * 0.10 + \text{Skor Completeness} > * 0.15 + < \text{Skor Error Tolerance} > * \text{responden} = 0.10 + < \text{Skor Execution Efficiency} > * 0.10 + < \text{Skor Operability} > * 0.15 + < \text{Skor Simplicity} > * 0.15 + < \text{Skor Training} > * 0.15$$

3. Evaluasi Berdasarkan Kriteria/Rata-Rata Skor Responden

Dalam mengevaluasi berdasarkan kriteria, dalam penelitian ini menggunakan persamaan:

$$\text{Rata-rata Responden} = \frac{\text{Jumlah Skor Responden}}{\text{Jumlah Responden}} \tag{3.1}$$

IV. PEMBAHASAN

4.1. Hasil Pengujian Model

Nilai *accuracy*, *precision* dan *recall* dari data *training* dapat dihitung dengan menggunakan *Rapid Miner*. Setelah diuji coba dengan metode *cross validation*, didapat hasil pengukuran terhadap data *training* yaitu *accuracy*=92,89% *precision*=98.33% *anrecall*=83.39%.

4.2. Evaluasi Model *Confusion Matrix*

Tabel 4.1 adalah perhitungan berdasarkan data *training*, diketahui dari 183 data, 109 diklasifikasikan Positif sesuai dengan prediksi yang dilakukan dengan metode Algoritma *Naïve Bayes*, 12 data diprediksi Positif tetapi ternyata hasilnya Negatif, 61 data *class Negatif* diprediksi sesuai, dan 1 data diprediksi Negatif ternyata Positif

Tabel 4. 1 Model *Confusion Matrix* untuk Metode Algoritma *Naïve Bayes*

Accuracy : 92.89%			
	true Positif	true Negatif	class precision
Pred.Positif	109	12	90.08%
pred.Negatif	1	61	98.39%
class recall	99.09%	83.56%	

Dibawah ini adalah perhitungan manual untuk mencari nilai *accuracy*, *sensitivity*, *specificity*, *PPV* dan *NPV* :

$$\text{Accuracy} = \frac{TP+TN}{TP+TN+FP+FN} = \frac{109+61}{109+61+1+12} = 0,928962$$

$$\text{Sensitivity} = \frac{TP}{TP+FN} = \frac{109}{109+12} = 0,900826$$

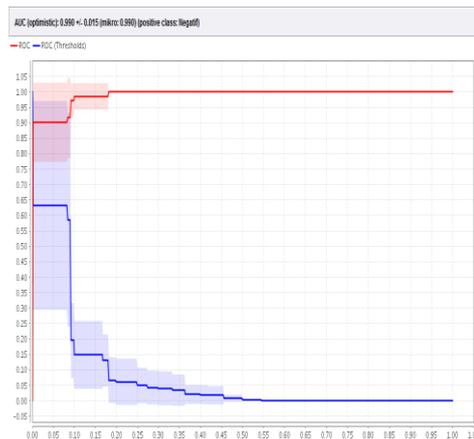
$$\text{Specificity} = \frac{TN}{TN+FP} = \frac{61}{61+1} = 0,983871$$

$$\text{PPV} = \frac{TP}{TP+FP} = \frac{109}{109+1} = 0,990909$$

$$\text{NPV} = \frac{TN}{TN+FN} = \frac{61}{61+12} = 0,835616$$

### 4.3. Kurva ROC (*Receiver Operating Characteristic*)

Hasil perhitungan kurva ROC untuk Algoritma *Naïve Bayes* dengan menggunakan data *training* hasilnya sebesar 0.990 dan dapat dilihat bentuk dari kurvanya seperti dibawah ini:



Gambar 4.1. Kurva ROC Menggunakan Data *Training*

Gambar 4.1. menunjukkan bahwa hasil algoritma *Naïve Bayes* mendapatkan nilai **0.990**, dimana nilai tersebut menunjukkan nilai AUC klasifikasi sangat baik, dikarenakan memiliki nilai diantara 0.90 – 1.00 [16].

### 4.4. Deployment

Dalam penerapan pada GUI yang dibuat, data yang digunakan sebanyak 25 *record* yang diambil pada data *testing* secara *random* menggunakan rumus Bernoulli, hasil dari perhitungan tersebut didapat 9 data pasien Negatif TB dan 16 data Positif TB, Gambar 4.2 menggambarkan bentuk GUI untuk memprediksi penyakit TB, berikut bentuk dari GUI yang dibuat:

Gambar 4.1 GUI Sistem Prediksi Diagnosa Penyakit TB

### 4.5. Hasil Pengukuran Uji Keandalan GUI

Skor rata-rata yang dihasilkan dari perhitungan uji keandalan GUI adalah 81,66, sedangkan nilai optimal untuk sebuah perangkat lunak yang memenuhi standar kualitas berdasarkan uji SQA adalah 80. Jadi GUI yang dibuat sudah memenuhi standar uji kelayakan SQA

### 4.4. Implikasi Penelitian

Hasil dari penelitian ini akan membawa dampak pada aspek sistem, aspek manajerial maupun terhadap penelitian selanjutnya.

#### 1. Aspek Sistem

Hasil penelitian ini dapat diterapkan dengan adanya dukungan sistem yang berjalan dengan baik, sehingga pihak yang berkepentingan atau ahli medis lain dapat menggunakan hasil penelitian ini untuk memprediksi penyakit TB. Diperlukan sarana dan prasana yang memadai yang terdiri dari *Hardware*, *Software* (sistem operasi dan aplikasi yang dibuat pada tahap *deployment*), dan infrastruktur lainnya guna memberikan hasil yang terbaik

## 2. Aspek Manajerial

Agar hasil penelitian ini dapat diterapkan dengan baik, perlu adanya langkah-langkah yang harus dilakukan diantaranya adalah:

- Pembuatan *Standard Operasional Procedure* perlu dilakukan agar penerapan hasil penelitian ini sesuai dengan tujuan yang diharapkan.
- Berdasarkan *Standard Operasional Procedure* yang dibuat, perlu adanya penyuluhan (*Desiminasi*) agar hasil dari penelitian ini dapat tersosialisasi dengan baik.
- Pelatihan terhadap calon pengguna diperlukan, agar hasil penelitian ini dapat dimengerti dan diterapkan.
- Penerapan hasil penelitian ini merupakan langkah yang terpenting. Jika langkah ini tidak dilakukan, maka hasil penelitian yang telah dilakukan tidaklah berguna.

## 3. Penelitian Lanjutan

Penelitian lebih lanjut dapat dilakukan dengan cara seperti berikut:

- Hasil penelitian dapat diterapkan pada instansi medis.
- Melakukan proses *feature selection* yang dapat meningkatkan akurasi dari algoritma *data mining*.
- Mengkomprasi dengan metode *data mining* lainnya seperti *Support Vector Machine*, *K-Nearest Neighbor*, *CART* dan lainnya.
- Penelitian dapat dikembangkan dengan mengoptimasikan algoritma terpilih. Seperti *Particle Swarm Optimization*, *Genetic Algorithm* dan lainnya.
- Graptic User Interface (GUI)* perlu dikembangkan mengikuti perkembangan IPTEK, salah satunya yaitu GUI berbasis *Mobile Computing*.

## V. KESIMPULAN

Dalam penelitian ini dilakukan pengujian model algoritma *Naïve Bayes*, hasil evaluasi dan validasi, diketahui bahwa *Naïve Bayes* memiliki nilai *accuracy* dan AUC cukup tinggi. Dengan demikian, metode *Naïve Bayes* merupakan metode yang cukup baik dalam memprediksi secara dini diagnosis penyakit TB.

*Graptic User Interface (GUI)* yang dibuat dapat digunakan dengan cukup akurat dalam memprediksi penyakit TB, berdasarkan diuji kelayakan standar kualitas SQA, GUI tersebut sudah memenuhi standar, karena nilai rata-rata dari beberapa responden itu memiliki nilai lebih dari 80, untuk itu GUI ini dapat diterapkan pada pihak PUSKESMAS untuk mendiagnosa penyakit TB secara dini.

Agar penelitian ini bisa ditingkatkan, berikut adalah saran-saran yang diusulkan:

- Hasil penelitian ini diharapkan dapat digunakan pada Puskesmas atau pada instansi dinas kesehatan, khususnya pada perusahaan riset dilakukan, untuk lebih meningkatkan efisiensi dan kecepatan dalam pengambilan suatu keputusan pada pasien.
- Menambahkan jumlah data yang lebih besar dan atribut yang lebih banyak, sehingga hasil pengukuran yang akan didapatkan lebih baik lagi.
- Melakukan pengembangan dengan *feature selection* seperti *genetic algorithm*, *chi square*, dan metode *feature selection* lainnya. untuk menyeleksi atribut yang berpengaruh kuat, sehingga atribut yang dipakai hanya sedikit namun tidak mengurangi akurasi dari algoritma yang digunakan.
- Penelitian ini dapat dikembangkan dengan mengoptimalkan parameter dengan *Particle Swarm Optimization*, *Genetic Algorithm* dan lainnya.
- Graptic User Interface (GUI)* perlu dikembangkan mengikuti kemajuan IPTEK, seperti GUI yang berbasis *mobile computing* dan lainnya.

## DAFTAR PUSTAKA

- Atif Imran Toor, "Decision Support System for Lung Diseases," 2006.
- Er Orhan, Feyzullah Temurtas, A Çetin Tanrikulu, "Tuberculosis Disease Diagnosis Using Artificial Neural Networks," *Springer*, pp. 299-302, 2010.
- Anonim. (2012, Mei) <http://tcbatam.com>. [Online]. <http://tcbatam.com/alur-diagnosis-tb/>
- Anonim. (2009, Mei) [www.hukor.depkes.go.id](http://www.hukor.depkes.go.id). [Online]. <http://www.hukor.depkes.go.id>
- Setiadi and Widodo, "Kajian Penerapan Model Neural Network Untuk Prediksi Penyakit

- Hati" , Jakarta, 2012.
- [6] Retno Asti Werdhani, PATOFISIOLOGI, DIAGNOSIS, DAN KLAFISIKASI TUBERKULOSIS, 2011.
- [7] Jason Fine, *An Overview Of Statistical Methods in Diagnostic Medicine*. Chapel Hill, 2012.
- [8] Wu and Kumar, *The Top Ten Algorithms in Data Mining*. USA: CRC Press, 2009.
- [9] Anonim, *Pedoman Penerapan DOTS di Rumah Sakit*. Jakarta, 2007.
- [10] S Susanto and D Suryadi, *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: C.V ANDI OFFSET, 2010.
- [11] Daniel T Larose, *Discovering Knowledge In Data*. Canada: Wiley-Interscience, 2005.
- [12] Carlo Vercellis, *Data Mining and Optimization for Decision Making*. Italy: WILEY, 2009.
- [13] Max Bramer, *Principles Of Data Mining*. London: Springer, 2007.
- [14] Kusrini & luthfi, *Algoritma Data mining*. Yogyakarta: Andi Offset, 2009.
- [15] J Han and M Kamber, *Data Mining: Concepts and Techniques*. San Fransisco: Morgan kauffman, 2006.
- [16] Prabowo,P,W., "GEOGRAPHICAL INFORMATION SYSTEM ENGINEERING IN THE IMPLEMENTATION OF SPATIAL DYNAMIC MODEL ON FORESTED AREAS ," in *ISIT*, Jakarta, 2009.
- [17] Florin Gorunescu, *Data Mining, Concepts, Models and Techniques*. Berlin: Springer, 2011.
- [18] Anonim, *Pedoman Penerapan DOTS di Rumah Sakit*. Jakarta, 2006.
- [19] Redho Pati. (2010, Oktober) ittelkom.ac.id. [Online] [http://digilib.ittelkom.ac.id/index.php?option=com\\_content&view=article&id=687:svm&catid=15:pemrosesan-sinyal&Itemid=14](http://digilib.ittelkom.ac.id/index.php?option=com_content&view=article&id=687:svm&catid=15:pemrosesan-sinyal&Itemid=14)
- [20] Witten, I. H., Frank, E., & Hall, M. A., *Data Mining Practical Machine Learning Tools and Techniques*. USA: Morgan Kaufmann Publishers, 2007.
- [21] Xu & Donald, *CLUSTERING*. Canada: A JOHN WILEY & SONS, INC, 2009.
- [22] Xiao Hua Zhou, Nancy A Obuchowski, and Donna K Mcclish, *Statistical Methods in Diagnostic Medicine*. New York: Canada, 2002.