

# KOMPARASI ALGORITMA KLASIFIKASI DATA MINING UNTUK MEMPREDIKSI PENYAKIT TUBERCULOSIS (TB): STUDI KASUS PUSKESMAS KARAWANG SUKABUMI

Rizal Amegia Saputra<sup>1)</sup>, Prabowo Pudjo Widodo<sup>2)</sup>

<sup>1)</sup> Program Studi Manajemen Informatika  
AMIK BSI Sukabumi  
Jl. Veteran No.20A,Sukabumi, Indonesia  
*e-mail: rizal.rga@bsi.ac.id*

<sup>2)</sup> Program Pascasarjana Magister Ilmu Komputer  
STMIK Nusa Mandiri Jakarta  
Jl. Salemba Raya No.5, Jakarta Pusat 10430, Indonesia  
*e-mail: prabowopw@yahoo.com*

**Abstrak-** Penyakit Tuberculosis(TB) merupakan penyakit menular dan mematikan didunia, bahkan *World Health Organization* (WHO) mencanangkan sebagai penyakit kedaruratan dunia (*global emergency*), beberapa penelitian bidang kesehatan termasuk salah satunya penyakit TB telah banyak dilakukan untuk mendeteksi penyakit secara dini. Dalam penelitian ini dilakukan komparasi algoritma *C4.5*, *naïve bayes*, *neural network* dan *logistic regression* yang diaplikasikan terhadap data pasien yang dinyatakan positif TB dan negatif TB . Dari hasil pengujian dengan mengukur kinerja dari keempat algoritma tersebut menggunakan metode pengujian *Confusion Matrix* dan *Kurva ROC*, diketahui bahwa algoritma *naïve bayes* memiliki nilai *accuracy* paling tinggi, yaitu **91,61%** diikuti oleh algoritma *C4.5* dengan *accuracy* sebesar 89,77% metode *neural network* sebesar 84,07% dan yang terendah adalah metode *logistic regression* dengan nilai *accuracy* 80,02%, Nilai AUC untuk metode *naïve bayes* juga menunjukkan nilai tertinggi, yaitu **0,995**, disusul algoritma *C4.5* dengan nilai AUC sebesar 0,982, metode *logistic regression* dengan nilai AUC 0,968 dan yang terendah adalah nilai AUC *neural network*, yaitu 0,940. Melihat nilai AUC dari keempat metode tersebut, maka semua metode termasuk kelompok klasifikasi sangat baik, karena hasil nilai AUC-nya antara 0.90-1.00.

*Keywords* : *Algorithm C4.5*, *naïve bayes*, *neural network* dan *logistic regression*

## 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) [34], 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 [25].

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. [5]. Tahun 2007 di Indonesia penyakit TB merupakan salah satu dari sepuluh besar penyakit penyebab kematian terbesar di indonesia [3].

Klasifikasi data penyakit TB pada medis merupakan tugas penting dalam memprediksi penyakit, bahkan dapat membantu dokter dalam mengambil keputusan diagnosis penyakit tersebut [11], 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 kesempatan ini penelitian yang akan dilakukan adalah membandingkan beberapa metode klasifikasi *data mining*, diantaranya yaitu Algoritma

C4.5, *Naïve Bayes*, *Neural Network*, dan *Logistic Regression*, keempat metode tersebut digunakan dalam memprediksi diagnosis penyakit TB dengan tujuan agar algoritma terpilih merupakan algoritma yang paling akurat sehingga dapat melakukan diagnosa penyakit TB secara dini, ke empat metode tersebut merupakan sepuluh klasifikasi *data mining* paling populer [39].

## 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 [2].

### 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 [33]. 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 [18].

### 2.3. Algoritma Klasifikasi Data Mining

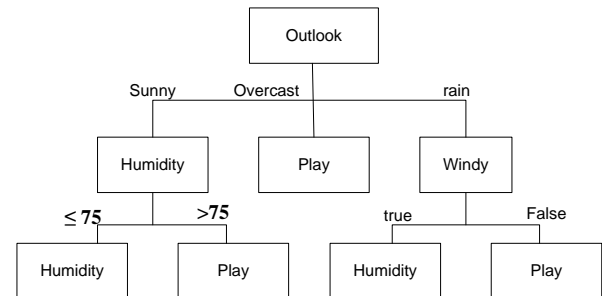
Klasifikasi *Data mining* [35] adalah suatu metode pembelajaran, untuk memprediksi nilai dari sekelompok atribut dalam menggambarkan dan membedakan kelas data atau konsep yang bertujuan untuk memprediksi kelas dari objek yang label kelasnya tidak diketahui.

### 2.4. Algoritma Tree C4.5

Algoritma C4.5 merupakan bagian dari kelompok algoritma *decision trees* dan merupakan katerogi 10 algoritma yang paling populer [13]. Algoritma C4.5 diperkenalkan oleh J. Ross Quinlan diakhir tahun 1970 hingga awal tahun 1980-an. J. Ross Quinlan seorang peneliti dibidang mesin pembelajaran yang merupakan pengembangan dari algoritma ID3 (*Iterative Dichotomiser*), algoritma tersebut digunakan untuk membentuk pohon keputusan. Pohon keputusan dianggap sebagai salah satu pendekatan yang paling populer, dalam klasifikasi pohon keputusan terdiri dari sebuah node yang membentuk akar, node akar tidak memiliki inputan. Node lain yang bukan sebagai akar tetapi memiliki tepat satu inputan disebut node internal atau test node, sedangkan node lainnya dinamakan daun. Daun mewakili nilai target yang paling tepat dari salah satu *class* [21].

Pada dasarnya konsep dari algoritma C4.5 adalah mengubah data menjadi pohon keputusan

dan aturan-aturan keputusan (*rule*). C4.5 adalah algoritma yang cocok untuk masalah klasifikasi dan data mining, C4.5 memetakan nilai atribut menjadi *class* yang dapat diterapkan untuk klasifikasi baru [39]. Seperti gambar 1 dibawah ini:



Sumber : [8]

**Gambar 1. Pohon Keputusan**

Ada beberapa tahapan dalam membangun sebuah pohon keputusan dengan Algoritma C4.5 yaitu[17]:

1. Menyiapkan *data training*. Data training biasanya diambil dari data histori yang pernah terjadi sebelumnya dan sudah dikelompokkan ke dalam kelas-kelas tertentu.
2. Menentukan akar dari pohon. Akar akan diambil dari atribut yang terpilih, dengan cara menghitung nilai *gain* dari masing-masing atribut, nilai *gain* yang paling tinggi yang akan menjadi akar pertama. Sebelum menghitung nilai *gain* dari atribut, hitung dahulu nilai *entropy*. Untuk menghitung nilai *entropy* digunakan rumus:

$$Entropy(S) = \sum_{i=1}^n -p_i * \log_2 p_i \quad (1)$$

Keterangan:

S : himpunan kasus

A : atribut

n : jumlah partisi S

pi : proporsi dari Si terhadap S

3. Kemudian hitung nilai *gain* yang menggunakan rumus:

$$Gain(S, A) = Entropy(S) - \sum_{i=1}^n \frac{|S_i|}{|S|} * Entropy(S_i) \quad (2)$$

Keterangan:

S = himpunan kasus

A = fitur

n = jumlah partisi atribut A

$\frac{|S_i|}{|S|}$  = proporsi Si terhadap S

|S| = jumlah kasus dalam S

4. Ulangi langkah ke-2 hingga semua record terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat:

- Semua *record* dalam simpul N mendapat kelas yang sama.
- Tidak ada atribut di dalam *record* yang dipartisi lagi.
- Tidak ada *record* di dalam cabang yang kosong

### 2.5. 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) [8]. Klasifikasi bayesian memiliki kemampuan klasifikasi serupa dengan *decision tree* dan *neural network* [17].

*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:

$$P(H|X) = \frac{P(X|H)P(H)}{P(X)} \quad (3)$$

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

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

$$P(H|X)=P(X|H)P(X) \quad (4)$$

*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*.

### 2.6. Neural Network

*Neural network* adalah satu set unit input/output yang terhubung dimana tiap relasinya memiliki bobot, Selama *fase* pembelajaran, *neural network* menyesuaikan bobot sehingga dapat memprediksi *class* yang benar dari *tuple* [13]. *Neural network* dimaksudkan untuk mensimulasikan perilaku sistem biologi susunan syaraf manusia, yang terdiri dari sejumlah besar unit pemroses yang disebut *neuron*, yang beroperasi secara parallel (Alpayandin, 2010). Neuron mempunyai relasi dengan *synapse* yang mengelilingi *neuron-neuron* lainnya. Susunan syaraf tersebut dipresentasikan dalam *neural network* berupa graf yang terdiri dari simpul (*neuron*) yang dihubungkan dengan busur, yang berkorespondensi dengan *synapse*. Sejak tahun 1950-an, *neural network* telah digunakan untuk tujuan prediksi, bukan hanya klasifikasi tapi juga regresi dengan atribut target kontinu [39].

Langkah pembelajaran algoritma *backpropagation* adalah sebagai berikut [17] :

- Inialisasikan bobot jaringan secara acak (biasanya antara -1.0 s/d 1.0)
- Untuk setiap data pada data training, hitung input untuk simpul berdasarkan nilai input dan bobot jaringan saat itu. Dengan menggunakan rumus :

$$Input_j = \sum_{i=1}^n O_i w_{ij} + \square_j \quad (5)$$

Keterangan :

$O_i$  = Output simpul I dari layer sebelumnya

$W_{ij}$  = bobot relasi dari simpul I pada layer sebelumnya ke simpul j

$\square_j$  = bias (sebagai pembatas)

- Berdasarkan input dari langkah kedua, selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktifitas sigmoid:

$$Output = \frac{1}{1+e^{-input}} \quad (6)$$

- Hitung nilai error antara nilai yang diprediksi dengan nilai yang sesungguhnya menggunakan rumus :

$$Error_j = output_j * (1-Output_j)*(Target_j-Output_j) \quad (7)$$

Keterangan :

$Output_j$  = Output actual dari simpul j

$Target_j$  = nilai target yang sudah diketahui pada data training

- Setelah nilai *error* dihitung, selanjutnya dibalik ke layer sebelumnya (*backpropagation*). Untuk menghitung nilai *error* pada hidden layer, menggunakan rumus :

$$Error_j = output_j * (1 - Output_j) * \sum_{k=1}^n error_k w_{jk} \quad (8)$$

Keterangan :

$Output_j$  = Output actual dari simpul j

$Error_k$  = Error dari simpul  $k$

$W_{jk}$  = Bobot relasi dari simpul  $j$  ke simpul  $k$  pada layer berikutnya

6. Nilai *error* yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi, dengan menggunakan rumus :

$$W_{ij} = W_{ij} + l * Error_j * Output_i$$

(9)

Keterangan :

$W_{ij}$  = bobot relasi dari unit  $i$  pada layer sebelumnya ke unit  $j$

$l$  = *learning rate* (konstantan, nilainya 0 sampai dengan 1)

$Error_j$  = *Error* pada *output layer* simpul  $j$

$Output_i$  = Output dari simpul  $i$ .

### 2.7 Logistic Regression

Regresi logistik (*Logistic regression*) adalah bagian dari analisis regresi yang digunakan ketika variabel dependen (*respon*) merupakan variabel dikotomi. Variabel dikotomi biasanya hanya terdiri atas dua nilai [30] yang mewakili kemunculan atau tidak adanya suatu kejadian yang biasanya diberi angka 0 atau 1. Tidak seperti *regresi linier* biasa, regresi logistik tidak mengasumsikan hubungan antara variabel independen dan dependen secara linier. Regresi logistik merupakan *regresi non linier* dimana model yang ditentukan akan mengikuti pola kurva linier.

$$\text{Log}(P/1-P) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k \quad (10)$$

### 2.8 Uji Keandalan GUI

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].

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

$$\begin{aligned} \text{Skor masing-masing responden} = & \langle \text{SkorAuditability} \rangle * 0.10 + \\ & \langle \text{SkorAccuracy} \rangle * 0.10 + \\ & \langle \text{SkorCompleteness} \rangle * 0.15 + \\ & \langle \text{SkorErrorTolerance} \rangle * 0.10 + \\ & \langle \text{SkorExecutionEfficiency} \rangle * 0.10 + \\ & \langle \text{SkorOperability} \rangle * 0.15 + \\ & \langle \text{SkorSimplicity} \rangle * 0.15 + \\ & \langle \text{SkorTraining} \rangle * 0.15 \end{aligned} \quad (11)$$

Dalam mengevaluasi berdasarkan kriteria, dalam penelitian ini menggunakan persamaan [16]:

$$\text{Rata-rata Responden} = \frac{\text{Jumlah Skor Responden}}{\text{Jumlah Responden}} \quad (12)$$

## III. HASIL PENELITIAN DAN PEMBAHASAN

### 3.1. Analisis Hasil Komparasi

Komparasi hasil perhitungan nilai AUC untuk metode *C4.5*, *naïve bayes*, *neural network* dan *logistic regression* dapat dilihat pada Tabel 1 berikut ini:

Tabel.1 Komparasi Nilai *Accuracy* dan AUC

Metode	Confussion Matrix		AUC		Perbandingan Nilai Akurasi	Perbandingan Nilai AUC
	Tranning	Testing	Tranning	Testing		
Algoritma C4.5	86,87%	92,67%	0,964	1,000	89,77%	0,982
Naïve Bayes	<b>92,89%</b>	<b>90,33%</b>	<b>0,990</b>	<b>1,000</b>	<b>91,61%</b>	<b>0,995</b>
Neural Network	83,13%	85,00%	0,940	0,940	84,07%	0,940
Logistic Regression	83,04%	77,00%	0,963	0,972	80,02%	0,968

Tabel.1 membandingkan *accuracy* dan AUC dari tiap metode dan diambil nilai rata-rata hasil perhitungan dari data *tranning* dan *testing*. Terlihat bahwa nilai *accuracy Naïve Bayes* paling tinggi dengan nilai 91,61% dan AUC 0,995. Diketahui dari keempat metode yang dibandingkan, semuanya termasuk kedalam klasifikasi sangat baik, dikarenakan memiliki nilai AUCnya diantara 0.90 – 1.00.

### 3.2. 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 2 menggambarkan bentuk GUI untuk

memprediksi penyakit TB, berikut bentuk dari GUI yang dibuat:



Gambar.2. GUI Sistem Prediksi Diagnosa Penyakit TB

### 3.3. Hasil Pengukuran Uji Keandalan GUI

Skor rata-rata yang dihasilkan dari perhitungan 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.

### 3.4. Implikasi Penelitian

Implikasi penelitian pada penelitian ini terarah pada tiga aspek, yaitu:

#### 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 dapat 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.

d. 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*.
- 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*.

## IV. KESIMPULAN

Dalam penelitian ini dilakukan pengujian model dengan membandingkan empat metode *data mining* yaitu algoritma *C4.5*, *naive bayes*, *neural network*, dan *logistic regression*, hasil evaluasi dan validasi, diketahui bahwa *Naive Bayes* memiliki nilai *accuracy* dan AUC paling tinggi diantara metode yang dikomprasikan, diikuti oleh algoritma *C4.5*, *neural network*, dan *logistic regression* memiliki akurasi yang paling rendah. Dengan demikian, metode *Naive 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.

4. Penelitian ini dapat dikembangkan dengan membandingkan algoritma *data mining* lainnya misalkan saja dengan metode *Support Vector Machine*, *K-Nearest Neighbor*, *CART* dan lainnya atau dapat mengoptimalkan parameter dengan *Particle Swarm Optimization*, *Genetic Algorithm* dan lainnya.
5. *Graphic User Interface (GUI)* perlu dikembangkan mengikuti kemajuan IPTEK, seperti GUI yang berbasis *mobile computing* dan lainnya.

#### DAFTAR PUSTAKA

- [1] Anonim. (2006, Maret). Pedoman Penerapan DOTS di Rumah Sakit. Jakarta, Jakarta.
- [2] Anonim. (2007). Pedoman Penerapan DOTS di Rumah Sakit. Jakarta.
- [3] Anonim.(2009, Mei 13). Retrieved Mei 01, 2013, from [www.hukor.depkes.go.id](http://www.hukor.depkes.go.id): <http://www.hukor.depkes.go.id>
- [4] Anonim.(2012, Mei 20). Retrieved Maret 23, 2013, from <http://www.depkes.go.id>: <http://www.depkes.go.id/index.php/berita/press-release/1922-pengendalian-kasus-tuberkulosis-harus-berkualitas-untuk-mencegah-terjadinya-tb-mdr.html>
- [5] Anonim.(2012.). *Alur Diagnosis TB*. Retrieved April 13, 2013, from <http://tcbatam.com>: <http://tcbatam.com/alur-diagnosis-tb/>
- [6] Asha, Natarajan & Murty. (2011). *A Hybrid Approach to the Diagnosis of Tuberculosis by Cascading Clustering and Classification*. *JOURNAL OF COMPUTING*, 2151-9617.
- [7] Bennett, K. P., & Campbell, C. (2000). *Support Vector Machines: Hype or Hallelujah*. *SIGKDD*, vol-2.
- [8] Bramer, M. (2007). *Principles Of Data Mining*. London: Springer.
- [9] C, R Kothari. (2004). *Research Methodology Methods and Techniques*. India: New Age International Limited.
- [10] Elveren, E., & Yumusak, N. (2011). *Tuberculosis Disease Diagnosis Using Artificial Neural Network Trained with Genetic Algorithm*. *Springer Science*, 35:329–332.
- [11] Fine, J. (2012). *An Overview Of Statistical Methods in Diagnostic Medicine*. Chapel Hill.
- [12] Gorunescu, F. (2011). *Data Mining, Concepts, Models and Techniques*. Berlin: Springer.
- [13] Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. San Fransisco: Morgan kauffman.
- [14] Han, J., Kamber, M., & Pei, J. (2011). *Data Mining: Concepts and Techniques, 3rd.ed*. Boston: Morgan Kaufmann.
- [15] Ioan, P. (2007, August 19). *An approach of the Naive Bayes classifier for the document classification*. pp. 135–138.
- [16] Jha, G. K. (2007). *Artificial Neural Network*. Pusa, New Delhi: Indian Agricultural Research Institute.
- [17] Kusriani, & Luthfi, E. T. (2009). *Algoritma Data Mining*. Yogyakarta: Andi Offset.
- [18] Larose, D. T. (2005). *Discovering Knowledge In Data*. Canada: Wiley-Interscience.
- [19] -----(2006). *Data Mining Methods and Models*. Canada: John Wiley & Sons.
- [20] Lio & Triantaphyllou. (2008). *Recent Advances in Data Mining of Enterprise Data: Algorithms and Application*. USA: World Scientific Publishing.
- [21] Maimon & Rokach. (2010). *Data Mining and knowledge Discovery Handbook*. New York: Springer.
- [22] Meyer, D. (2012, September 12). Retrieved Februari 20, 2013, from <http://cran.r-project.org>: <http://cran.r-project.org/web/packages/e1071/vignettes/svmdoc.pdf>.
- [23] Noble, W. S. (2006, Desember 12). Retrieved Februari 20, 2013, from <http://www.raetschlab.org>: <http://raetschlab.org/lectures/ismb09tutorial/images/WhatIsASVM.pdf>
- [24] Nugroho, A. S., Witarto, A. B., & Handoko, D. (2003, Desember 20). Retrieved Februari 22, 2013, from <http://ilmukomputer.com>: <http://ilmukomputer.com>
- [25] Orhan, E., Temurtas, F., & Tanrikulu, A. Ç. (2010). *Tuberculosis Disease Diagnosis Using Artificial Neural Networks*. Springer, 299-302.
- [26] Pati, R. (2010, Oktober 18). *digilib*. Retrieved Februari 20, 2013, from [ittelkom.ac.id](http://ittelkom.ac.id): [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).
- [27] Prabowo, P.W. (2009). *Geographical Information System Engineering In The Implementation Of Spatial Dynamic Model On Forested Areas Case Study: The Province Bali, Riau, And South Kalimantan*. Jakarta.10-11.
- [28] Prasetyo, E. (2013). *Data Mining Konsep & Aplikasi Menggunakan Matlab*. Yogyakarta: Andi Publisher.
- [29] Samanta, Al-Balushi, & Al-Araimi. (2003). *Artificial neural networks and support vector machines with genetic algorithm for bearing fault detection*. *Engineering Applications of Artificial Intelligenc*, 657–665.
- [30] Santoso. (2007). *Data Mining Terapan dengan Matlab*. Yogyakarta: Graha Ilmu.
- [31] Setiadi & Prabowo. (2012). *Kajian Penerapan Model Neural Network Untuk Prediksi Penyakit Hati*. Jakarta.
- [32] Silwattananusarn, T., & Tuamsuk, K. (2012). *Data Mining and Its Application For Knowledge Management : A literature Riview From 2007 to 2012 (Vol. Vol.2)*. Thailand: IJDKP.
- [33] Susanto, S., & Suryadi, D. (2010). *Pengantar Data Mining Menggali Pengetahuan dari Bongkahan Data*. Yogyakarta: C.V ANDI OFFSET.
- [34] Toor, A. I. (2006). *Decision Support System for Lung Diseases*. Turki.
- [35] Vercellis, C. (2009). *Data Mining and Optimization for Decision Making*. Italy: WILEY.

- [36] Wahono, R.S. (2012, Agustus). Retrieved Mei 12, 2013, from <http://romisatriawahono.net>:  
<http://romisatriawahono.net/2012/08/07/kiat-menyusun-kerangka-pemikiran-penelitian/>
- [37] Werdhani, R. A. (2011). Patofisiologi, Diagnosis, Dan Klafikasi Tuberkulosis. Jakarta.
- [38] Witten, I. H., Frank, E., & Hall, M. A. (2007). *Data Mining Practical Machine Learning Tools and Techniques*. USA: Morgan Kaufmann Publishers.
- [39] Wu & Kumar. (2009). *The Top Ten Algorithms in Data Mining*. USA: CRC Press.
- [40] Xu & Donald. (2009). *Clustering*. Canada: A JOHN WILEY & SONS, INC.
- [41] Zhou, X. H., Obuchowski, N. A., & Mcclish, D. K. (2002). *Statistical Methods in Diagnostic Medicine*. New York: Canada.

#### **Biodata Penulis**

**Rizal Amegia Saputra**, memperoleh gelar Sarjana Komputer (S.Kom), Jurusan Sistem Informasi STMIK Nusa Mandiri Sukabumi, lulus tahun 2010. Memperoleh gelar Magister Komputer (M.Kom) Program Pasca Sarjana Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta, lulus tahun 2013. Saat ini menjadi Staff Akademik di AMIK BSI Sukabumi.