

Aplikasi Diagnosa Penyakit Tuberculosis Menggunakan Algoritma Data Mining

Amrin¹, Hafdiarsya Saiyar²

¹Program Studi Teknik Komputer AMIK Bina Sarana Informatika Jakarta
amrin.ain@bsi.ac.id

²Program Studi Teknik Komputer AMIK Bina Sarana Informatika Jakarta
hafdiarsya.hyr@bsi.ac.id

Cara Sitasi: Amrin, & Saiyar, H. (2018). Aplikasi Diagnosa Penyakit Tuberculosis Menggunakan Algoritma Data Mining. *Paradigma*, XX(2), 91-97.

Abstract - It is important for doctors to make an early diagnosis of tuberculosis in order to reduce the transmission of the disease to the wider community. In this study, the authors will apply and compare several methods of data mining classification, including Algoritma C4.5, Naïve Bayes, and Neural Network to diagnose tuberculosis disease, then compare which of the three methods are the most accurate. Based on the performance measurement results of the three models using Cross Validation, Confusion Matrix and ROC Curve methods, it is known that Naïve Bayes method is the best method with accuracy of 94.18% and under the curva (AUC) value of 0.977, then neural network method with accuracy 89,89% and under the curva value (AUC) 0,975, and then C4.5 method with accuracy level equal to 84,56% and under the curva value (AUC) equal to 0,938. This shows that the three models that are produced including the category of classification is very good because it has an AUC value between 0.90-1.00.

Keywords: C4.5, naïve bayes, neural network, confusion matrix, ROC Curva

PENDAHULUAN

Tuberculosis yang disingkat TBC atau TB adalah penyakit menular yang disebabkan oleh bakteri *Mycobacterium Tuberculosis* yang ditularkan melalui udara (*droplet nuclei*) saat seorang pasien TBC batuk dan percikan ludah yang mengandung bakteri tersebut terhirup oleh orang lain saat bernapas (Widoyono, 2011). Penyakit TB (Orhan dan Tanrikulu, 2010) 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. Penyakit ini dapat menular sehingga perlu penanganan yang intensif, setidaknya diperlukan pengobatan minimal 6 bulan secara rutin dan terus menerus. Sedangkan Indonesia menempati peringkat ke-2 di dunia setelah India dengan pasien TBC terbanyak dan diperkirakan ada 1.020.000 kasus TB di Indonesia (Kemenkes, 2018). Penularan *tuberculosis* (TBC) sangat cepat melalui udara. Bagi penderita diharapkan selalu melakukan pemeriksaan dan pengobatan sampai tuntas. TBC ditularkan melalui udara. Percikan ludah atau dahak yang dikeluarkan menjadi media penularan yang sangat cepat di dunia ini. Penularan TBC melalui udara akan sangat rentan terjadi di ruang publik.

Dari berbagai penelitian akan ada puluhan ribu kuman yang keluar dari batuk dan bersin. Oleh karenanya diharapkan masyarakat untuk menggunakan masker di tempat-tempat umum dan senantiasa berperilaku hidup bersih dan sehat (Kemenkes, 2018).

Klasifikasi data penyakit TB pada medis merupakan tugas penting dalam memprediksi penyakit, bahkan dapat membantu dokter dalam mengambil keputusan diagnosis penyakit tersebut (Fine, 2012), dengan demikian sangat penting melakukan diagnosa secara dini agar dapat mengurangi penularan TB kepada masyarakat luas. Pada penelitian ini, penulis akan menerapkan dan membandingkan beberapa metode klasifikasi data mining, diantaranya yaitu Algoritma C4.5, Naïve Bayes, dan Neural Network untuk mendiagnosa penyakit *Tuberculosis*. Data yang penulis gunakan adalah data pasien puskesmas bojonggede yang terdiagnosa *tuberculosis*.

Menurut Han dan Kamber dalam (Amrin, 2016) *Data mining* adalah rangkaian proses untuk menggali nilai tambah berupa informasi yang belum terekplorasi dari sebuah basis data, melakukan eksplorasi dengan cara-cara tertentu untuk memanipulasi data menjadi informasi yang lebih berharga dengan cara mengekstraksi dan mengenali pola penting dari basis data. Menurut Daryl Pregibons dalam (Gorunescu, 2011) "*Data mining*

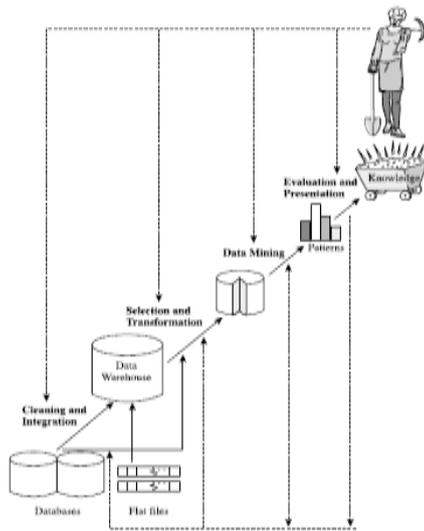
adalah perpaduan dari ilmu statistik, kecerdasan buatan, dan penelitian bidang *database*". Nama *data mining* berasal dari kemiripan antara pencarian informasi yang bernilai dari *database* yang besar dengan menambang sebuah gunung untuk sesuatu yang bernilai (Sumathi & Sivanandam, 2006). Keduanya memerlukan penyaringan melalui sejumlah besar material, atau menyelidiki dengan cerdas untuk mencari keberadaan sesuatu yang disebut bernilai tadi.

Data Mining merupakan teknologi baru yang sangat berguna untuk membantu perusahaan-perusahaan menemukan informasi yang sangat penting dari gudang data mereka. Beberapa aplikasi data mining fokus pada prediksi, mereka meramalkan apa yang akan terjadi dalam situasi baru dari data yang menggambarkan apa yang terjadi di masa lalu (Witten, Frank, & Hall, 2011)

METODOLOGI PENELITIAN

A. Tahapan Proses Data Mining

Data mining sering disebut juga Knowledge Discovery in Database atau disingkat menjadi KDD, adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola atau hubungan dalam set data berukuran besar (Santosa, 2007). Gambar tahapan pembuatan aplikasi data mining ditunjukkan pada gambar 1 berikut ini:



Sumber: Han & Kamber (2006)

Gambar 1. Tahapan Proses KDD

Gambar 1 menunjukkan langkah dalam proses *data mining*. Proses dalam tahap *data mining* terdiri dari tiga langkah utama, yaitu (Sogala, 2006):

1. Data Preparation

Pada langkah ini, data dipilih, dibersihkan, dan dilakukan *preprocessed* mengikuti pedoman dan *knowledge* dari ahli domain yang menangkap dan mengintegrasikan data internal dan eksternal ke dalam tinjauan organisasi secara menyeluruh.

2. Algoritma *data mining*

Penggunaan algoritma *data mining* dilakukan pada langkah ini untuk menggali data yang terintegrasi untuk memudahkan identifikasi informasi bernilai.

3. Fase analisa data

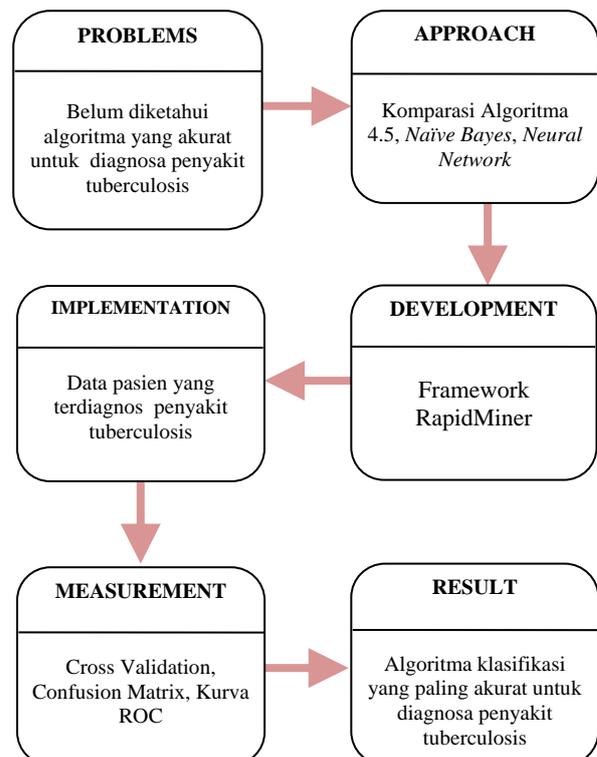
Keluaran dari data mining dievaluasi untuk melihat apakah *knowledge* domain ditemukan dalam bentuk *rule* yang telah diekstrak dari jaringan.

B. Kerangka Pemikiran Pemecahan Masalah

Jenis penelitian yang dilakukan dalam penelitian ini adalah jenis penelitian eksperimen komparatif. Eksperimen komparatif yaitu membandingkan dua objek yang berbeda, misalnya membandingkan dua algoritma yang berbeda dengan melihat hasil statistik masing-masing mana yang lebih baik (Khotari, 2004).

Penelitian ini terdiri dari beberapa tahap seperti terlihat pada kerangka pemikiran Gambar 2 Permasalahan (*problem*) pada penelitian ini adalah Belum diketahui algoritma yang akurat untuk diagnosa penyakit tuberculosis.

Untuk itu dibuat *approach* (model) yaitu algoritma C4.5, *Naive Bayes*, dan *Neural Network* untuk memecahkan permasalahan kemudian dilakukan pengujian terhadap kinerja dari ketiga metode tersebut. Pengujian menggunakan metode *Cross Validation*, *Confusion Matrix* dan kurva ROC. Untuk mengembangkan aplikasi (*development*) berdasarkan model yang dibuat, digunakan Rapid Miner.



Gambar 2. Kerangka Pemikiran Pemecahan Masalah

C. Algoritma C4.5

Algoritma C4.5 merupakan struktur pohon dimana terdapat simpul yang mendeskripsikan atribut-atribut, setiap cabang menggambarkan hasil dari atribut yang diuji, dan setiap daun menggambarkan kelas. Algoritma C4.5 secara rekursif mengunjungi setiap simpul keputusan, memilih pembagian yang optimal, sampai tidak bisa dibagi lagi. Algoritma C4.5 menggunakan konsep *information gain* atau *entropy reduction* untuk memilih pembagian yang optimal (Han & Kamber, 2006).

Ada beberapa tahap dalam membuat sebuah pohon keputusan dengan algoritma C4.5 (Kusrini & Luthfi, 2009), yaitu :

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:

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

Keterangan:

- S = himpunan kasus
- n = jumlah partisi S
- p_i = proporsi S_i terhadap S

3. Kemudian hitung nilai *gain* menggunakan rumus:

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

Keterangan:

- S = himpunan kasus
- A = fitur
- n = jumlah partisi atribut A
- $|S_i|$ = proporsi S_i terhadap S
- $|S|$ = jumlah kasus dalam S

4. Ulangi langkah ke-2 hingga semua *record* terpartisi.
5. Proses partisi pohon keputusan akan berhenti saat :
 - a. Semua *record* dalam simpul N mendapat kelas yang sama.
 - b. Tidak ada atribut di dalam *record* yang dipartisi lagi.
 - c. Tidak ada *record* di dalam cabang yang kosong.

D. Naïve Bayes

Klasifikasi Bayes (Kusrini & Luthfi, 2009) adalah pengklasifikasian statistik yang dapat digunakan untuk memprediksi probabilitas keanggotaan suatu kelas. Klasifikasi Bayes juga dikenal dengan *Naïve Bayes*.

$$P(x|y) = \frac{P(y|x) P(x)}{P(y)}$$

keterangan :

y = data dengan kelas yang belum diketahui

x = hipotesis data y merupakan suatu kelas spesifik

$P(x | y)$ = probabilitas hipotesis x berdasar kondisi y (*posteriori probability*)

$P(x)$ = probabilitas hipotesis x (*prior probability*)

$P(y | x)$ = probabilitas y berdasarkan kondisi pada hipotesis x

$P(y)$ = probabilitas dari y

E. Neural Network

Neural network adalah (Han & Kamber, 2006) satu set unit *input/output* yang terhubung dimana tiap relasinya memiliki bobot. *Multi Layer Perceptron* disebut juga *multilayer feedforward neural network* merupakan sebuah kelas *neural network* Algoritma *backpropagation* untuk *multilayer perceptron*, merupakan metode yang sistematis untuk *training* sehingga bisa dilakukan dan lebih efisien (Maimon & Rokach, 2010). MLP terdiri dari *input layer*, satu atau lebih *hidden layer*, dan *output layer*. Langkah pembelajaran dalam algoritma *backpropagation* adalah sebagai berikut (Myatt, 2007):

1. Inisialisasi bobot jaringan secara acak (biasanya antara -0.1 sampai 1.0)
2. Untuk setiap data pada data *training*, hitung input untuk simpul berdasarkan nilai input dan bobot jaringan saat itu, menggunakan rumus:

$$Input_j = \sum_{i=1}^n O_i w_{ij} + \Theta_j$$

Keterangan:

O_i = *Output* simpul *i* dari layer sebelumnya

w_{ij} = bobot relasi dari simpul *i* pada layer sebelumnya ke simpul *j*

Θ_j = bias (sebagai pembatas)

3. Berdasarkan input dari langkah dua, selanjutnya membangkitkan output untuk simpul menggunakan fungsi aktivasi sigmoid:

$$Output = \frac{1}{1 + e^{-Input}}$$

4. Hitung nilai *Error* antara nilai yang diprediksi dengan nilai yang sesungguhnya menggunakan rumus:

$$Error_j = Output_j \cdot (1 - Output_j) \cdot (Target_j - Output_j)$$

Keterangan:

$Output_j$ = Output aktual dari simpul j

$Target_j$ = Nilai target yang sudah diketahui pada data *training*

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

$$Error_j = Output_j(1 - Output_j) \sum_{k=1}^n Error_k w_{jk}$$

Keterangan:

$Output_j$ = Output aktual dari simpul j

$Error_k$ = error simpul k

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

- Nilai *Error* yang dihasilkan dari langkah sebelumnya digunakan untuk memperbarui bobot relasi menggunakan rumus

$$w_{ij} = w_{ij} + l \cdot Error_j \cdot Output_i$$

Keterangan:

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

l = *learning rate* (konstanta, nilainya antara 0 sampai dengan 1)

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

$Output_i$ = *Output* dari simpul i

F. Evaluasi dan Validasi Model

Untuk mengukur akurasi model maka dilakukan evaluasi dan validasi menggunakan teknik:

1. Confusion matrix

Confusion Matrix adalah alat (*tools*) visualisasi yang biasa digunakan pada supervised learning. Tiap kolom pada matriks adalah contoh kelas prediksi, sedangkan tiap baris mewakili kejadian di kelas yang sebenarnya (Gorunescu, 2011). *Confusion matrix* berisi informasi aktual (*actual*) dan prediksi (*predicted*) pada sistem klasifikasi.

2. Kurva ROC (*Receiver Operating Characteristic*)

Kurva ROC menunjukkan akurasi dan membandingkan klasifikasi secara visual. ROC mengekspresikan *confusion matrix*. ROC adalah grafik dua dimensi dengan *false positives* sebagai garis horizontal dan *true positives* sebagai garis vertikal (Vercellis, 2009). *The area under curve* (AUC) dihitung untuk mengukur perbedaan performansi metode yang digunakan. AUC dihitung menggunakan rumus: (Liao, 2007)

$$\theta^r = \frac{1}{mn} \sum_{j=1}^n \sum_{i=1}^m \psi(x_i^r, x_j^r)$$

Dimana

$$\psi(X,Y) = \begin{cases} 1 & Y < X \\ \frac{1}{2} & Y = X \\ 0 & Y > X \end{cases}$$

Performance keakurasian AUC dapat diklasifikasikan menjadi lima kelompok yaitu (Gorunescu, 2011):

0.90 – 1.00 = *Excellent Clasification*

0.80 – 0.90 = *Good Clasification*

0.70 – 0.80 = *Fair Clasification*

0.60 – 0.70 = *Poor Clasification*

0.50– 0.60 = *Failure*

HASIL DAN PEMBAHASAN

A. Analisa Data

Pada penelitian ini data yang digunakan sebanyak 136 data pasien tuberculosis (TBC) baik yang positif maupun yang negatif. Variabel input pada penelitian ini terdiri dari enam variabel, yaitu: 1. Keringat pada malam hari tanpa aktivitas fisik, 2. Berat badan turun, 3. Nafsu makan berkurang 4. Mudah lelah dan lemah, 5. Demam, 6. Batuk berdahak lebih dari tiga minggu disertai batuk darah, Sedangkan variabel output adalah variabel penyakit TBC. Perangkat lunak yang digunakan untuk menganalisa adalah RapidMiner *versi 5.3*.

B. Pengujian Model

Model yang telah dibentuk diuji tingkat akurasi dengan memasukan data uji yang berasal dari data *training*. Karena data yang didapat dalam penelitian ini setelah proses *preprocessing* hanya 136 data maka digunakan metode *cross validation*, *Confusion Matrix*, dan Kurva ROC untuk menguji tingkat akurasi. Untuk nilai akurasi model untuk metode C4.5 sebesar 84.56%, metode *naïve bayes* sebesar 94.18%, dan metode *neural network* sebesar 89.89%.

1. Confusion Matrix

Berdasarkan Tabel 1 berikut ini, diketahui dari 136 data, 66 diklasifikasikan tidak (negatif) sesuai dengan prediksi yang dilakukan dengan metode C4.5, lalu 11 data diprediksi tidak (negatif) tetapi ternyata ya (positif), 49 data *class* ya (positif) diprediksi sesuai, dan 10 data diprediksi ya (positif) ternyata tidak (negatif).

Tabel 1. Model *Confusion Matrix* untuk Metode C4.5

accuracy: 84.56% +/- 4.01% (mikro: 84.56%)			
	true T	true Y	class precision
pred T	66	11	85.71%
pred Y	10	49	83.05%
class recall	86.84%	81.67%	

Sumber: Hasil Pengolahan Menggunakan RapidMiner 5.3 (2018)

Tabel 2 adalah *confusion matrix* untuk metode *naïve bayes*. diketahui dari 136 data, 74 diklasifikasikan tidak (negatif) sesuai dengan prediksi yang dilakukan dengan metode *naïve bayes*, lalu 6 data diprediksi tidak (negatif) tetapi ternyata ya (positif), 54 data *class* ya (positif) diprediksi sesuai, dan 2 data diprediksi ya (positif) ternyata tidak (negatif).

Tabel 2. Model *Confusion Matrix* untuk Metode *naïve bayes*

accuracy: 94.18% +/- 6.38% (mikro: 94.12%)			
	true T	true Y	class precision
pred. T	74	6	92.50%
pred. Y	2	54	96.43%
class recall	97.37%	90.00%	

Sumber: Hasil Pengolahan Menggunakan RapidMiner 5.3 (2018)

Tabel 3 adalah *confusion matrix* untuk metode *neural network*. diketahui dari 136 data, 69 diklasifikasikan tidak (negatif) sesuai dengan prediksi yang dilakukan dengan metode *naïve bayes*, lalu 7 data diprediksi tidak (negatif) tetapi ternyata ya (positif), 53 data *class* ya (positif) diprediksi sesuai, dan 7 data diprediksi ya (positif) ternyata tidak (negatif).

Tabel 3. Model *Confusion Matrix* untuk Metode *Neural Network*

accuracy: 89.89% +/- 8.63% (mikro: 89.71%)			
	true 0.0	true 1.0	class precision
pred. 0.0	69	7	90.78%
pred. 1.0	7	53	88.33%
class recall	90.78%	88.33%	

Sumber: Hasil Pengolahan Menggunakan RapidMiner 5.3 (2018)

Dari tiga table *confusion matrix*, selanjutnya dilakukan perhitungan nilai *accuracy*, *precision*, dan *recall*. Perbandingan nilai *accuracy*, *precision*, dan *recall* yang telah dihitung untuk metode C4.5, *naïve bayes*, dan *neural network* dapat dilihat pada Tabel 4

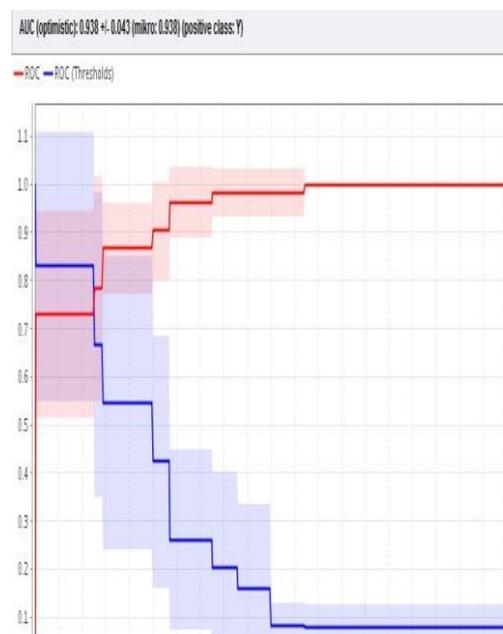
Tabel 4 Komparasi Nilai *Accuracy*, *Precision*, dan *Recall*

	C4.5	<i>Naïve bayes</i>	<i>Neural Network</i>
<i>Accuracy</i>	84.56%	94.18%	89.89%
<i>Precision</i>	85.40%	97.14%	91.07%
<i>Recall</i>	81.71%	90.48%	88.81%

Sumber: Hasil Pengolahan Menggunakan RapidMiner 5.3 (2018)

2. Kurva ROC

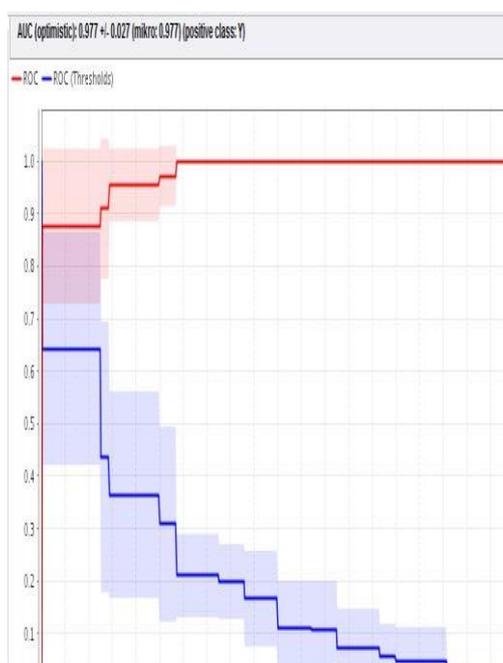
Hasil perhitungan divisualisasikan dengan kurva ROC. Perbandingan ketiga metode komparasi bisa dilihat pada Gambar 3 yang merupakan kurva ROC untuk algoritma C4.5.



Sumber: Hasil Pengolahan Menggunakan RapidMiner 5.3 (2018)

Gambar 3 Kurva ROC dengan algoritma C4.5

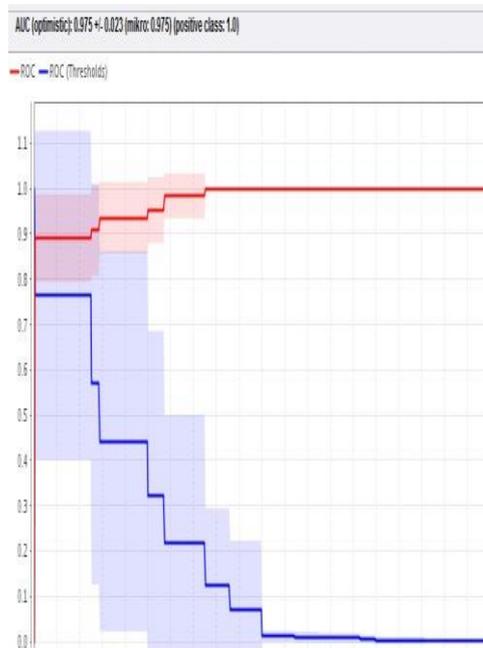
Kurva ROC pada gambar 3 mengekspresikan *confusion matrix*. Garis horizontal adalah *false positives* dan garis vertikal *true positives*.



Sumber: Hasil Pengolahan Menggunakan RapidMiner 5.3 (2018)

Gambar 4 Kurva ROC dengan Metode *Naïve Bayes*

Seperti terlihat pada Gambar 3, Gambar 4, dan Gambar 5, area di bawah kurva pada Gambar 4 paling luas diantara ketiga metode.



Sumber: Hasil Pengolahan Menggunakan RapidMiner 5.3 (2018)

Gambar 5 Kurva ROC dengan Metode *Neural Network*

Pebandingan hasil perhitungan nilai AUC untuk metode C4.5, *naïve bayes*, dan *neural network* dapat dilihat pada Tabel 5.

Tabel 5 Komparasi Nilai AUC

	C4.5	<i>Naïve Bayes</i>	<i>Neural network</i>
AUC	0.938	0.977	0.975

Sumber: Hasil Pengolahan Menggunakan RapidMiner 5.3 (2018)

3. Analisis Hasil Komparasi

Model yang dihasilkan metode C4.5, *naïve bayes*, dan *neural network* diuji menggunakan metode *Cross Validation*, terlihat perbandingan nilai *accuracy*, *precision*, dan *recall* pada Tabel 6, untuk metode *naïve bayes* C4.5 memiliki nilai *accuracy*, *precision*, *sensitivity*, dan *recall* yang paling tinggi, diikuti dengan metode *neural network*, dan yang terendah adalah metode C.45.

Tabel 6 Komparasi Nilai *Accuracy* dan AUC

	C4.5	<i>Naïve Bayes</i>	<i>Neural network</i>
<i>Accuracy</i>	84.56%	94.18%	89.89%
AUC	0.938	0.977	0.975

Sumber: Hasil Pengolahan Menggunakan RapidMiner 5.3 (2018)

Tabel 6 membandingkan *accuracy* dan AUC dari tiap metode. Terlihat bahwa nilai *accuracy naïve bayes* paling tinggi begitu pula dengan nilai AUC-nya. Untuk metode *neural network* dan C4.5 juga

menunjukkan nilai yang sesuai. Untuk klasifikasi *data mining*, nilai AUC dapat dibagi menjadi beberapa kelompok (Gorunescu, 2011).

- 0.90-1.00 = klasifikasi sangat baik
- 0.80-0.90 = klasifikasi baik
- 0.70-0.80 = klasifikasi cukup
- 0.60-0.70 = klasifikasi buruk
- 0.50-0.60 = klasifikasi salah

Berdasarkan pengelompokan di atas dan Tabel 6 maka dapat disimpulkan bahwa metode C4.5, *naïve bayes*, dan *neural network* termasuk klasifikasi sangat baik karena memiliki nilai AUC antara 0.90-1.00.

C. Rancangan Aplikasi Data Mining

Berdasarkan hasil di atas diperoleh bahwa metode terbaik adalah metode *naïve bayes*, selanjutnya berdasarkan metode tersebut kemudian dirancang dan dibuatlah aplikasi diagnosa penyakit *tuberculosis* menggunakan algoritma data mining, dalam hal ini *naïve bayes*, maka implementasi hasil rancangan aplikasi seperti terlihat pada gambar 6 berikut ini:



Gambar 6. GUI Sistem Prediksi Penyakit Tuberculosis

KESIMPULAN

Kesimpulan yang dapat diambil berdasarkan penelitian ini adalah bahwa:

1. Performa model C4.5 untuk diagnosa penyakit TBC memberikan tingkat akurasi kebenaran sebesar 84,56% dengan nilai area under the curve (AUC) sebesar 0,938. Performa model *naïve bayes* memberikan tingkat akurasi kebenaran sebesar 94,18% dengan nilai area under the curve (AUC) sebesar 0,977. Sedangkan Performa model *neural network* memberikan tingkat akurasi kebenaran sebesar 89,89% dengan nilai area under the curve (AUC) sebesar 0,975. Hal ini menunjukkan bahwa ketiga model tersebut termasuk katagori klasifikasi sangat baik karena memiliki nilai AUC antara 0.90-1.00.

2. Berdasarkan tingkat akurasi dan nilai area under the curve (AUC), maka performa metode *naïve bayes* adalah yang paling baik untuk mendiagnosa penyakit TBC, disusul metode neural network, dan berikutnya metode C4.5.

REFERENSI

- Amrin, A. (2016). Data Mining Dengan Regresi Linier Berganda Untuk Peramalan Tingkat Inflasi. *Jurnal Techno Nusa Mandiri*, XIII(1), 74–79. Retrieved from <http://ejournal.nusamandiri.ac.id/ejurnal/index.php/techno/article/view/268>
- Fine, J. (2012). *An Overview Of Statistical Methods in Diagnostic Medicine*. Chapel Hill.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.
- Han, J., & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Soft Computing (Vol. 54). San Fransisco: Morgan Kauffman. <https://doi.org/10.1007/978-3-642-19721-5>
- Khotari. (2004). *Data Mining Concepts and Technique*. San Fransisco: Morgan Kauffman.
- Kusrini, & Luthfi, E. . (2009). *Algoritma Data Mining*. Yogyakarta: Andi Publishing.
- Liao, T. W. (2007). *Recent Advances in Data Mining of Enterprise Data: Algorithms and Application*. Singapore: World Scientific Publishing.
- Maimon, O., & Rokach, L. (2010). *Data Mining And Knowledge Discovery Handbook*. New York: Springer.
- Myatt, G. J. (2007). *Making Sense of Data: A Practical Guide to Exploratory Data Analysis and Data Mining*. New Jersey: John Wiley & Sons, Inc.

- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Sogala, S. S. (2006). *Comparing the Efficacy of the Decision Trees with Logistic Regression for Credit Risk Analysis*. India.
- Sumathi, S., & Sivanandam, S. N. (2006). *Introduction to Data Mining and its Applications*. Berlin Heidelberg New York: Springer.
- Vercellis, C. (2009). *Business Intelligent: Data Mining and Optimization for Decision Making*. Southern Gate, Chichester, West Sussex: John Willey & Sons, Ltd.
- Witten, I. H., Frank, E., & Hall, M. A. (2011). *Data Mining: Practical Machine Learning and Tools*. Burlington: Morgan Kaufmann.

PROFIL PENULIS

- Amrin, S.Si, M.Kom. Dompu 10 Agustus 1980. Tahun 2003 lulus dari Program Strata Satu (S1) Jurusan Matematika Universitas Diponegoro Semarang. Tahun 2014 lulus dari Program Strata Dua (S2) Jurusan Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta. Pekerjaan saat ini sebagai Dosen AMIK BSI Jakarta sejak tahun 2007.
- Hafdiarsya Saiyar, ST, M.Kom. Jakarta 07 April 1983. Tahun 2005 Lulus dari Program Strata Satu (S1) Jurusan Teknologi Informatika Universitas Muhammadiyah Prof.DR HAMKA Jakarta. Tahun 2015 Lulus dari Program Strata Dua (S2) Jurusan Magister Ilmu Komputer STMIK Nusa Mandiri Jakarta.