

MODEL PENGAMBILAN KEPUTUSAN DALAM PREDIKSI KASUS TUBERKULOSIS MENGGUNAKAN REGRESI LOGISTIK BERBASIS *BACKWARD ELIMINATION*

Ratih Sari Wardani *, Purwanto**

* Program Studi Kesehatan Masyarakat, Universitas Muhammadiyah Semarang
email: ratihsw@gmail.com

** Program Studi Teknik Informatika, Universitas Dian Nuswantoro, Semarang
email: purwanto@dsn.dinus.ac.id

ABSTRACT

The purpose of this study is to obtain a predictive model in cases of Tuberculosis (TB) using a logistic regression based on backward elimination. The study is conducted through literature review, collecting secondary data of 200 patients of Pulmonary Tuberculosis aged more than 15 years. Data is collected from BKPM Semarang, Central Java. The data consist of characteristics, anamnesis, physical examination, laboratory test results data and radiological examination. Logistic regression based on backward elimination model has better accuracy than logistic regression without attribute selection.

Kata kunci : Tuberkulosis, Regresi Logistik, Backward Elimination, Prediksi

PENDAHULUAN

Tuberkulosis adalah penyakit menular yang dapat menyerang Paru maupun organ lain. TB sampai sekarang masih menjadi masalah global kesehatan dunia, tahun 2011 diperkirakan 8,7 juta kasus TB (13% ko-infeksi dengan HIV) dan membunuh 1,4 juta orang (1 juta HIV negative dan 430.000 orang dengan HIV positif) dimana Indonesia menempati posisi nomor empat dari seluruh kasus TB dunia (WHO, 2012).

TB memerlukan penegakan diagnosis yang akurat, ketidakakuratan diagnosis menyebabkan pengobatan yang diberikan pada pasien menjadi tidak tepat, hal itu berpotensi untuk meningkatkan kasus resistensi terhadap obat TB (*multidrug-resistant tuberculosis/MDR-TB*) yang sampai sekarang semakin meningkat menjadi 60.000 kasus pada tahun 2011(WHO, 2012). Laporan JEMM (*Joint External Monitoring Mission*) tahun 2007, sebanyak 10-30% pasien TB di rumahsakit DOTS tidak melaksanakan proses diagnosis dan sputum secaralengkap (Depkes& Stop TB partnership, 2007), selanjutnya penelitian pada 61 rumah sakit di Jawa 13-53% pasien TB paru dewasa tidak mendapatkan tatakelola diagnosis sesuai standar (Probandari et.al, 2010). Oleh karena itu permasalahan ketidakakuratan diagnosis TB menjadi tantangan para peneliti saat ini.

Banyak model yang dapat digunakan untuk membantu manajemen dalam pengambilan keputusan, diantaranya dengan model regresi logistik. Model regresi logistik telah banyak digunakan untuk prediksi di berbagai bidang. Regresi logistik untuk prediksi kematian dalam trauma kepala (Eftekharet. al., 2005). Model ini juga telah digunakan untuk analisis resiko kredit (Satchidananda dan Simha, 2006). Escaño et. al. (2006) menggunakan regresi logistik untuk data medis. Untuk kasus TB, Permasasari et. al. (2009) telah memprediksi insiden TB, tetapi hanya menggunakan satu variabel (*univariate*) data.

Dalam penelitian ini, dikembangkan model pengambilan keputusan dalam prediksi kasus TB berdasarkan banyak variabel (*multivariate*) yang meliputi data yang diperoleh dari karakteristik pasien, anamnesis, pemeriksaan fisik, data pemeriksaan laboratorium dan data pemeriksaan radiologi menggunakan regresi logistik berbasis *backward elimination*.

METODE PENELITIAN

Data dikumpulkan dari rekammedis pasien Tuberkulosis Paru usia 15 tahun ke atas dari Balai Kesehatan Penyakit Paru Masyarakat (BKPM) wilayah Semarang tahun 2013 sebanyak 200 pasien. Data sekunder diperoleh dari data rekam medis pasien terdiri dari data karakteristik, anamnesis, pemeriksaan fisik, data pemeriksaan laboratorium dan data pemeriksaan radiologi. Alur penelitian dimulai dari:

1. Pengumpulan data-data yang diperlukan untuk diagnosis Tuberkulosis;

2. Pengolahan awal (*pre-processing*) data dilakukan untuk pembersihan data TB. Pengolahan awal data dalam penelitian ini meliputi menghilangkan semua nilai *missing value* sebelum dilakukan analisis data lebih lanjut. Selanjutnya dilakukan pembagian data untuk digunakan dalam proses pembelajaran (*training*) dan pengujian (*testing*). Pada penelitian ini, 90% record digunakan untuk training dan 10 % record digunakan untuk testing.
3. Membangun model regresi logistik dipergunakan untuk prediksi kasus TB. Pada penelitian ini digunakan regresi logistik berbasis *Bakward Elimination*. Langkah *Bakward Elimination* dimulai dengan menggunakan semua variabel bebas X_i ($i=1,2,\dots,n$) sebagai input model regresi logistik dan selanjutnya secara bertahap mengurangi jumlah variabel di dalam persamaan model.
4. Evaluasi model menggunakan *confusion matrix*.

HASIL DAN PEMBAHASAN

Data yang Digunakan

Penelitian ini dimulai dengan mengumpulkan data 200 orang pasien Tuberkulosis Paru usia 15 tahun ke atas terdiri dari variabel karakteristik, pemeriksaan laboratorium, pemeriksaan radiologi dan hasil diagnosis yang dikumpulkan dari laporan TB03 dan buku bantu suspek TB pada tahun 2013, sedangkan untuk variabel anamnesis dan hasil pemeriksaan fisik diperoleh dengan cara membuka rekam medis pasien satu persatu. Selanjutnya pada data yang terkumpul dilakukan pengolahan awal data dengan tujuan untuk menghilangkan *missing* data maupun *outlier* data sehingga data siap untuk digunakan sebagai dasar pembangunan model prediksi. Variabel yang digunakan untuk prediksi terdiri 16 variabel bebas terdiri dari karakteristik pasien (jenis kelamin dan umur), anamnesis (gejala yang dialami pasien : batuk, berdahak atau tidak, ada darah atau tidak, panas badan, sesak nafas, berat badan, nafsu makan dan ada tidaknya anggota keluarga yang batuk), pemeriksaan fisik (denyut nadi dan tekanan darah), pemeriksaan laboratorium (dahak) dan pemeriksaan radiologi (rontgen thorax atau tidak dan hasilnya) sesuai dengan standard penatalaksanaan TB yang ditetapkan. Sedangkan variabel terikatnya adalah hasil diagnosis (TB Paru (Basil Tahan Asam (BTA) + dan TB Paru BTA-). Hasilnya adalah data yang dapat digunakan sebagai data dasar analisis sebesar 180 kasus, sebagai berikut :

Karakteristik

1. Jenis Kelamin (jk)

Tabel 1 Distribusi Frekuensi Berdasarkan jenis kelamin

Variabel	Frekuensi	Persentase
Jenis kelamin		
1. Laki-laki	105	58.3
2. Perempuan	75	41.7
Jumlah	180	100

Sebagian besar responden berjenis kelamin laki-laki (58,3%)

2. Umur

Umur responden berkisar antara 16 tahun sampai dengan 74 tahun dengan rata-rata 41,19 tahun dan simpangan baku 15,011 tahun

Anamnesis

1. Batuk (batuk)

Mayoritas responden mengalami gejala batuk (96,7%) dan 3,3 % tidak batuk.

2. Batuk berdahak (berdahak)

Responden yang batuk mayoritas mengalami batuk berdahak (81,7%) dan 19,3% tidak berdahak.

3. Keberadaan darah (darah)

Responden yang batuk hanya 15,6 % yang dahaknya mengandung darah, sedangkan 84,4 % batuknya tidak mengandung darah.

4. Sesak nafas (sesak)
Sebagian besar responden mengalami sesak nafas (57,8%), sedangkan 42,2% tidak sesak nafas.
5. Panas Badan (panas)
Sebagian besar responden pada saat periksa badannya panas (54,4%) dan 45,6% tidak mengalami panas badan.
6. Nafsu Makan (naf_makan)
Nafsu makan responden sebagian besar kurang (51,1%) dan 48,9% nafsu makannya baik.
7. Berat badan (bb)
Berat badan responden sebagian besar turun (60%), 38,3% nafsu makan tetap dan yang naik hanya 1,7 %
8. Batuk anggota keluarga (batuk_angg)
Mayoritas responden tidak mempunyai anggota keluarga yang batuk (84,4%), sedangkan 15,6 % mempunyai anggota keluarga yang batuk

Pemeriksaan Fisik

1. Denyut nadi (denyut_nadi)
Denyut nadi responden berkisar antara 63 kali/menit sampai dengan 201 kali/menit dengan rata-rata 106,07 kali/menit dan simpangan baku 20,926 kali/menit
2. Tekanan darah (sistole dan diastole)
Tekanan darah sistole berkisar 81 mmHg sampai dengan 210 mmHg dengan rata-rata 123,32 mmHg dan simpangan baku 19,162. Sedangkan untuk diastole minimum 81 mmHg dan maksimum 127 mmHg dengan rata-rata $75,65 \pm 12,138$ mmHg.

Pemeriksaan Radiologi

1. Pemeriksaan Thorax (thorax)
Mayoritas responden dilakukan pemeriksaan Thorax (97,2%) dan 2,8 % tidak dilakukan pemeriksaan Thorax.
2. Hasil pemeriksaan Thorax (hasil)
Hasil pemeriksaan thorax 85,5% positif TB, sedangkan 14,5% hasilnya negatif TB.

Pemeriksaan Laboratorium (dahak)

Pemeriksaan laboratorium yang dilakukan adalah pemeriksaan dahak Sewaktu, Pagi dan Sewaktu (S-P-S), hasilnya 42,8 % BTA dan positif 57,2% BTA negatif.

Diagnosis TB (diagnosis)

Hasil diagnosis TB diperoleh 42,8 % TB Paru BTA positif dan 57,2% TB Paru BTA negatif

Model Regresi Logistik

Model regresi logistik (Hosmer dan Lemeshow, 2000) dengan variabel bebas sebanyak n buah dapat dinyatakan sebagai:

$$\text{Logit } P(Y = 1) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

Persamaan (1) dapat ditulis dengan:

$$P(Y = 1) = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n}} \quad (2)$$

Dimana:

- P = probabilitas /kemungkinan
 X_1, X_2, \dots, X_n = variabel bebas/independen
 β_0 = konstanta
 $\beta_1, \beta_2, \dots, \beta_n$ = koefisien variabel

Persamaan (2), sebagai model regresi logistik dapat disederhanakan menjadi:

$$P(Y = 1) = \frac{1}{1 + e^{-(\beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n)}} \quad (3)$$

Bobot model regresi logistik tanpa menggunakan *backward elimination* dihasilkan sebagai berikut:

Bias (konstanta): 1.125

Bobot [jk = 0] = 0.006

Bobot [jk = 1] = -0.008

Bobot [batuk = 0] = -0.000

Bobot [batuk = 1] = 0.008

Bobot [berdahak = 0] = -0.016

Bobot [berdahak = 1] = 0.072

Bobot [darah = 1] = 0.019

Bobot [darah = 0] = -0.103

Bobot [sesak = 1] = -0.042

Bobot [sesak = 0] = 0.031

Bobot [panas = 1] = -0.009

Bobot [panas = 0] = 0.007

Bobot [naf_makan = 0] = 0.013

Bobot [naf_makan = 1] = -0.014

Bobot [bb = 0] = 0.005

Bobot [bb = 1] = -0.075

Bobot [batuk_angg = 1] = -0.005

Bobot [batuk_angg = 0] = 0.026

Bobot [thorax = 1] = 0.001

Bobot [thorax = 0] = -0.036

Bobot [hasil = 0] = -0.000

Bobot [hasil = 1] = 0.000

Bobot [dahak = 0] = -3.614

Bobot [dahak = 1] = 2.702

Bobot [umur] = 0.062

Bobot [denyut_nadi] = -0.254

Bobot [sistole] = 0.022

Bobot [diastole] = -0.026

Untuk mengevaluasi akurasi prediksi, pada penelitian ini menggunakan *confusion matrix* (Gorunescu, 2011).

		PREDICTED CLASS	
		Class= Yes/Positive	Class= No/ Negative
OBSERVED CLASS	Class= Yes/Positive	a (TP=True Positive)	b (FN=False Negative)
	Class= No/ Negative	c (FP=False Positive)	d (TN=True Negative)

Gambar 1. Confusion Matrix untuk akurasi prediksi

Rumus akurasi prediksi adalah sebagai berikut:

$$Accuracy = \left(\frac{a + d}{a + b + c + d} \right) = \frac{TP + TN}{TP + TN + FP + FN} \quad (4)$$

Untuk menentukan akurasi, percobaan dilakukan dengan menggunakan 90% record digunakan untuk training dan 10% record digunakan untuk testing. Hasil confusion matrix adalah sebagai berikut:

Tabel 2. Hasil Confusion Matrix

	true 0.0	true 1.0
pred. 0.0	9	9
pred. 1.0	0	0

Dari tabel 2 dan persamaan (4), maka akurasi yang dihasilkan adalah

$$Accuracy = \frac{9}{18} \times 100\%.$$

Dengan menggunakan model regresi logistik berbasis *backward elimination*, maka tidak semua variabel bebas digunakan. Variabel bebas *sesak nafas* akan tereliminasi dan variabel bebas yang berpengaruh terhadap diagnosis TB yaitu *jenis kelamin, umur, batuk, berdahak, darah, panas, nafsu makan, berat badan, batuk anggota keluarga, denyut nadi, sistole, diastole, thorax, hasil, dan dahak* akan dipergunakan sebagai input model regresi logistik. Akurasi yang dihasilkan dengan menggunakan model regresi logistik berbasis *backward elimination* adalah sebesar 72,22%. Perbandingan akurasi model regresi logistik dan regresi logistik berbasis *backward elimination* ditunjukkan pada tabel 3.

Tabel 3. Perbandingan Akurasi Prediksi

Model	Akurasi (%)	Peningkatan (%)
Regresi Logistik	50	
Regresi Logistik + Backward Elimination	72,22	44,44

KESIMPULAN

Pembuatan model diagnosis TB didasarkan pada sistem penatalaksanaan TB sesuai dengan strategi DOTS menggunakan variabel karakteristik pasien, anamnesis, pemeriksaan fisik, pemeriksaan laboratorium dan pemeriksaan radiologi. Hasil eksperimen menunjukkan bahwa model regresi logistik berbasis *backward elimination* menghasilkan akurasi sebesar 72,22%. Model ini dapat digunakan untuk membantu pengambilan keputusan klinis terutama dalam memprediksi kasus TB.

UCAPAN TERIMAKASIH

Ucapan terima kasih disampaikan kepada Direktorat Pendidikan Tinggi Kementerian Pendidikan dan Kebudayaan Republik Indonesia yang telah memberikan dana hibah penelitian tahun anggaran 2014.

DAFTAR PUSTAKA

- Depkes dan Stop TB Partnership. 2007. Report of the Joint External TB Monitoring Mission Indonesia (16-27 April 2007). Jakarta: Depkes
- Escaño, L.M.E., Saiz, G. S., Lorente, F.J.L., Fernando, A.B. dan Ugarriza, J.M.V. 2006. Logistic Regression Versus Neural Networks for Medical Data, Monografías del Seminario Matemático García de Galdeano 33:245–252
- Eftekhar, J., Mohammad, K., Ardebili, H.E., Ghodsi, M. dan Ketabchi, E. 2005. Comparison Of Artificial Neural Network And Logistic Regression Models For Prediction Of Mortality In Head Trauma Based On Initial Clinical Data. BMC Medical Informatics and Decision Making. 5(3): 1-8.
- Gorunescu, F. 2011. Data Mining: Concepts, Models and Techniques, Intelligent Systems. Reference Library, 12, 319–330
- Hosmer, D. W., dan Lemeshow, S. 2000. Applied Logistic Regression (2nd ed.). New York, USA: John Wiley & Sons, Inc.

- Permanasari, A. E., Rambli, D. R. A., Dominic, D.D., 2009. A Comparative Study of Univariate Forecasting Methods for Predicting Tuberculosis Incidence on Human, Proceeding of 2009 Student Conference on Research and Development (SCORed 2009), Malaysia, pp 188-191
- Probandari, A., Utarini, A. dan Karin, A.H., 2010, Missed Opportunity for Standardized Diagnosis and Treatment among Adult Tuberculosis Patient in hospital Involved in Public Private Mix for Directly Observed Treatment Short Course Strategy in Indonesia : a cross sectional study. BMC Health Service Research.
- Satchidananda, S.S. dan Simha, J.B. 2006. Comparing Decision Trees with Logistic Regression for Credit Risk Analysis (SAS APAUGC)
- WHO, 2012. Global Tuberculosis Control Report. Swizerland: WHO Press