

Model *Neural Network* dengan Inisialisasi Pembobot Awal Menggunakan Regresi Logistik Biner untuk Memprediksi Jenis Penyakit *Erythemato-Squamous*

Aziz Kustiyo¹, Handayani Tjandrasa²

¹ Staf Pengajar Departemen Ilmu Komputer FMIPA IPB

² Staf Pengajar Departemen Teknik Informatika FTIF ITS

Abstrak

Diagnosa untuk membedakan penyakit-penyakit yang termasuk ke dalam golongan *erythemato-squamous* merupakan masalah besar dalam dermatologi. Hal tersebut disebabkan karena penyakit-penyakit tersebut memiliki ciri-ciri *erythema* dan *scaling* yang hampir serupa. Di samping itu, pada tahap awal perjalanannya, satu penyakit dapat menunjukkan ciri-ciri penyakit lain dan pada tahap berikutnya baru menunjukkan gejala penyakit yang sebenarnya. Pada penelitian ini, model *feedforward neural network* dengan *training propagasi balik* dengan inisialisasi pembobot awal menggunakan regresi logistik biner digunakan untuk melakukan prediksi terhadap jenis penyakit tersebut. Hasil yang diperoleh menunjukkan bahwa model *neural network* tersebut memiliki ketepatan 91,04% lebih tinggi dibandingkan dengan model regresi logistik nominal (74,50%) dan model *neural network* dengan inisialisasi pembobot awal secara acak (53,63%).

Kata kunci: *neural network*, regresi logistik, ketepatan prediksi

PENDAHULUAN

Latar Belakang

Diagnosa untuk membedakan penyakit-penyakit *erythemato-squamous* merupakan masalah besar dalam dermatologi. Penyakit-penyakit yang masuk ke dalam kelompok tersebut adalah : *psoriasis*, *seboreic dermatitis*, *lichen planus*, *pityriasis rosea*, *chronic dermatitis* dan *pityriasis rubra pilaris*. Pada pengamatan awal, penyakit-penyakit tersebut memiliki ciri-ciri klinis *erythema* dan *scaling* dengan perbedaan-perbedaan yang sangat kecil (Merz dan Murphy, 1996).

Dalam dunia kedokteran, langkah-langkah yang dilakukan untuk menentukan jenis penyakit tersebut terdiri dari dua tahap yaitu pemeriksaan secara klinis dan secara histopatologis menggunakan mikroskop. Kesulitan

yang timbul pada evaluasi tersebut adalah bahwa satu penyakit mungkin menunjukkan ciri-ciri histopatologis penyakit lain pada tahap permulaan dan baru menunjukkan karakteristiknya pada tahap berikutnya (Merz dan Murphy, 1996).

Beberapa penelitian telah dilakukan untuk memprediksi jenis penyakit tersebut, antara lain menggunakan algoritma *nearest neighbor classifier*, *naïve bayesian classifier*, *voting feature intervals-5*, algoritma genetika, dan *neural network* dengan algoritma propagasi balik.

Ketepatan prediksi model *neural network* tersebut jauh lebih rendah dibandingkan dengan ketepatan prediksi algoritma lainnya. Padahal Penggunaan model *neural network* pada beberapa penelitian lain menunjukkan bahwa *neural network* mampu menghasilkan ketepatan prediksi yang cukup tinggi.

Tujuan Penelitian

Tujuan penelitian ini adalah sebagai berikut:

1. Membangun model neural network untuk memprediksi jenis penyakit *erythematous-squamous* dengan menggunakan regresi logistik biner dan regresi linier untuk menentukan pembobot awal (*inisialisasi*) dan arsitektur neural network.
2. Membangun model regresi logistik nominal untuk memprediksi jenis penyakit *erythematous-squamous*.
3. Membandingkan model neural network dengan model regresi logistik nominal dalam memprediksi jenis penyakit *erythematous-squamous*.

Beberapa penelitian terkait

Beberapa penelitian yang terkait dengan penelitian ini dapat dibagi menjadi dua kelompok, yaitu penelitian-penelitian yang terkait dengan prediksi jenis penyakit *erythematous-squamous* dan penelitian-penelitian yang terkait perbandingan antara neural network (*NN*) dengan regresi logistik (*RL*).

Penelitian-penelitian tentang prediksi jenis penyakit erythematous-squamous

Guvendir dan Ekmesiz (2000) membuat sistem pakar untuk memprediksi jenis penyakit *erythematous-squamous* berdasarkan pada tiga algoritma klasifikasi, yaitu *nearest neighbor classifier*, *naïve bayesian classifier* dan *voting feature intervals-5*. *Guvendir et al (1998)* mempelajari diagnosa untuk membedakan jenis penyakit *erythematous-squamous* menggunakan *voting feature interval*. *Fidelis et al (2000)* melakukan klasifikasi terhadap jenis penyakit *erythematous-squamous* menggunakan algoritma klasifikasi berbasis algoritma genetika. Algoritma genetika yang diajukan memiliki encoding kromosom yang fleksibel di mana setiap kromosom bersesuaian dengan sebuah rule klasifikasi. *Bojarczuk et al (2001)* melakukan klasifikasi terhadap jenis

penyakit *erythematous-squamous* menggunakan *algoritma constrained-syntax genetic programming* dengan konsep-konsep *data mining* sebagai dasarnya. *Lay et al (1999)* menggunakan neural network dengan algoritma *backpropagation* dilanjutkan dengan algoritma apriori untuk mengklasifikasikan jenis penyakit *erythematous-squamous*. *Castelano et al (2003)* menggunakan sistem neuro-fuzzy, yang disebut *KERNEL*, untuk mengklasifikasikan jenis penyakit *erythematous-squamous*.

Penelitian-penelitian tentang perbandingan antara neural network dengan regresi logistik

Warner dan Manavendra (1996) membandingkan *NN* dan *RL* untuk memprediksi peluang kematian pasien setelah menjalani operasi cangkok jantung bypass (*coronary artery bypass grafting surgery*) berdasarkan pada faktor-faktor resiko individual. *Schumacher (1996)* membandingkan *NN* dan *RL* untuk mengklasifikasikan tumor payudara jinak dan ganas (*benign and malignant breast tumor*) berdasarkan pengukuran sonografi dan karakteristik pasien. *Wu dan Su (1998)* membandingkan *GANN (genetic algorithm neural network)* regresi logistik untuk memprediksi pola berjalan pada pasien yang mengalami pembedahan sendi (*arthrodesis*) pergelangan kaki.

Manfaat Penelitian

Manfaat dari penelitian ini adalah mengetahui pengaruh inisialisasi pembobot awal menggunakan regresi logistik biner terhadap ketepatan prediksi neural network. Manfaat lainnya adalah membandingkan model neural network dengan regresi logistik nominal.

TINJAUAN PUSTAKA

Artificial neural network

Menurut *Fu (1994)*, *neural*

network terdiri dari elemen pemroses yang menyerupai neuron dan sejumlah koneksi terboboti di antara elemen-elemen tersebut. *Neural network* memiliki arsitektur paralel-terdistribusi dengan sejumlah besar node dan koneksi. Setiap koneksi menghubungkan satu node dengan node yang lain dan memiliki pembobot tertentu.

Topologi *neural network* ditentukan oleh struktur dan skema interkoneksi. Struktur sering dinyatakan sebagai jumlah *layer* dan jumlah node per *layer*.

Terdapat 3 tipe *layer*, yaitu *layer input*, *layer hidden* dan *layer output*. *Layer input* mengkodekan *input* yang akan diproses dalam jaringan, *layer output* mengkodekan nilai yang mungkin bagi *input* tertentu, sedangkan *layer hidden* adalah *layer* yang berada di antara kedua *layer* tersebut.

Berdasarkan skema interkoneksinya, *neural network* dapat dibedakan menjadi *feedforward* atau *recurrent* dan koneksi-koneksinya simetris atau asimetris.

Pembobot pada koneksi-koneksi tersebut dapat berupa bilangan real atau bilangan bulat. Pembobot tersebut dapat dibatasi pada suatu selang tertentu. Selama proses *training*, pembobot-pembobot tersebut akan dikoreksi tetapi dapat juga nilainya ditetapkan. Ketika proses *training* selesai, semua pembobot harus sudah ditetapkan.

Beberapa model *neural network* yang digunakan untuk klasifikasi antara lain: *single layer perceptron*, *adaline*, dan *multilayer perceptron (MLP)*. *MLP* merupakan *feedforward neural network* dengan sedikitnya satu *hidden layer*. Model ini dapat menangani klasifikasi non linear karena dapat membentuk daerah keputusan yang lebih kompleks (daerah keputusan pada *perceptron* berupa *hyperplane*).

Backpropagation network merupakan *multilayer feedforward network* dengan *learning rule* disebut *backpropagation* yang merupakan jenis dari teknik *gradient descent* dengan

backward error (gradient) propagation. Himpunan contoh-contoh *training* untuk *network* tersebut harus diberikan berkali-kali agar diperoleh pembobot koneksi-koneksi di antara neuron yang menghasilkan klasifikasi *input* dengan benar.

Backpropagation network dapat mengaproksimasi pemetaan hanya jika *network* cukup besar. Pada umumnya, tidak diketahui berapa ukuran *network* yang akan memberikan hasil terbaik untuk satu masalah tertentu. Jika *network* terlalu kecil maka tidak dapat belajar untuk menangani data *trainig* dengan baik. Di sisi lain, jika *network* terlalu besar maka *network* dapat mempelajari berbagai solusi yang konsisten dengan data *training* tetapi memberikan aproksimasi yang tidak tepat terhadap model aktual.

Validasi Silang

Validasi silang (*cross validation*) merupakan salah satu teknik untuk menduga *error rate*. Beberapa teknik yang lain di antaranya adalah: *holdout*, *leave one out* dan *bootstrapping* (Weiss dan Kapulous dalam Fu (1994)). *K-fold cross-validation* (Stone dalam Fu (1994)) membagi data menjadi *k* bagian terpisah (*disjoint*), satu bagian menjadi data *testing* dan *k-1* bagian menjadi data *training* sehingga terdapat *k* pasang data *training-testing*.

Leave one out dan *bootstrapping* digunakan untuk contoh berukuran kecil (*small sample*), *holdout* untuk contoh besar (misalnya 1000) dan *k-fold cross-validation* dapat digunakan untuk contoh kecil maupun contoh besar.

Regresi logistik

Menurut Hosmer dan Lemeshow (1989), tujuan melakukan analisis data menggunakan regresi logistik adalah untuk mendapatkan model terbaik dan sederhana, namun model tersebut sejalan dengan tinjauan dari ilmu biologi untuk menjelaskan hubungan di antara hasil (*variabel respon*) dengan variabel-variabel bebas (*variabel penjelas*).

Variabel bebas ini sering disebut dengan kovariat (*covariates*).

Pada regresi logistik, variabel respon berupa bilangan biner (0 dan 1) atau bilangan dengan skala nominal maupun ordinal. Berdasarkan ke-3 tipe variabel respon tersebut, regresi logistik dibedakan menjadi regresi logistik biner (*binary logistic regression*), regresi logistik nominal dan regresi logistik ordinal.

Membangun model regresi logistik dengan metode stepwise

Prosedur stepwise untuk pemilihan dan penghapusan variabel-variabel dari model didasarkan pada algoritma statistika yang memeriksa "penting atau tidaknya" (*importance*) variabel-variabel tersebut serta memasukkan atau mengeluarkan variabel-variabel tersebut berdasarkan pada aturan keputusan yang pasti.

Penting atau tidaknya sebuah variabel diukur berdasarkan signifikansi statistik koefisien variabel tersebut. Pada regresi logistik, error diasumsikan mengikuti distribusi binomial dan signifikansi diuji melalui uji khi-kuadrat rasio likelihood. Jadi pada setiap langkah, variabel terpenting secara statistik, adalah variabel yang menghasilkan perubahan terbesar terhadap log-likelihood relatif terhadap model tanpa variabel tersebut.

Model Regresi logistik nominal

Regresi logistik kebanyakan diterapkan pada model relasi antara variabel hasil yang berupa bilangan biner (*binary*) dengan sekumpulan variabel bebas, tetapi dengan sedikit modifikasi regresi logistik juga dapat diterapkan ketika variabel hasil merupakan politomi (*polytomous*), yaitu memiliki lebih dari 2 kategori.

Dalam membuat model regresi logistik untuk variabel hasil politomi, yang harus diperhatikan adalah skala pengukuran variabel hasil. Diasumsikan, variabel hasil yang dibahas adalah variabel hasil dengan skala nominal.

Skala variabel bebas dapat berupa kategorik maupun kontinu. Variabel dengan skala ordinal dapat dianalisis sebagai variabel kontinu atau kategorik.

Kesetaraan neural network dan regresi logistik biner

Model *neural network* dengan fungsi aktivasi sigmoid adalah sebagai berikut:

$$F(x) = \frac{1}{1 + e^{-g(x)}}$$

Model *regresi logistik biner* adalah sebagai berikut :

$$\pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

Berikut ini akan dibuktikan bahwa kedua model tersebut adalah sama.

Apakah $\pi(x) = F(x)$?

$$\begin{aligned} \pi(x) &= \frac{e^{g(x)}}{1 + e^{g(x)}} \\ &= \frac{e^{g(x)}}{1 + e^{g(x)}} : \frac{e^{g(x)}}{e^{g(x)}} \\ &= \frac{e^{g(x)}}{e^{g(x)} + e^{g(x)}} \\ &= \frac{1}{e^{-g(x)} + 1} \\ &= \frac{1}{1 + e^{-g(x)}} \\ &= F(x) \text{ (terbukti)} \end{aligned}$$

Regresi linier

Seperti halnya dalam regresi logistik, regresi linier digunakan untuk mendapatkan model terbaik dan

sederhana untuk menjelaskan hubungan di antara hasil (*variabel respon*) dengan variabel-variabel bebas (*variabel penjelas*). Perbedaannya terdapat pada variabel respon, di mana pada regresi linier, variabel respon merupakan bilangan dengan skala selang (*kontinyu*).

Menurut *Walpole (1990)* bentuk umum dari regresi linier adalah

$$Y_i = \alpha + \beta X_i, \quad i = 1, 2, \dots, n.$$

dengan Y_i merupakan variabel respon, α merupakan konstanta, dan β koefisien variabel bebas (X_i).

Penduga untuk α dan β diperoleh dengan menggunakan metode kuadrat terkecil. Bila diberikan contoh $\{(x_i, y_i); i = 1, 2, \dots, n\}$, maka nilai dugaan kuadrat terkecil bagi parameter dalam garis regresi $\hat{y} = a + bx$ dapat diperoleh dari rumus

$$b = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}$$

$$\text{dan} \quad a = \bar{y} - b\bar{x}$$

Kesetaraan neural network dan regresi linier

Bentuk umum dari regresi linier adalah $Y_i = \alpha + \beta X_i, i = 1, 2, \dots, n$ atau dapat ditulis sebagai $y_i = a + \sum b_i x_i, i = 1, 2, 3, \dots, n$, dengan a merupakan konstanta, dan b_i koefisien regresi.

Model neural network dengan fungsi aktivasi linier (*Schalkoff, 1992*) adalah sebagai berikut: $o_i = f(\text{net}) = \text{net}_i$ dengan $\text{net}_i = \sum w_i x_i + b, i = 1, 2, 3, \dots, n$, dengan w_i merupakan pembobot dan b merupakan bias.

Dengan demikian kedua model memiliki kesetaraan, yaitu konstanta pada regresi linier setara dengan bias pada model neural network ($a = b$) dan suku $\sum b_i x_i = \sum w_i x_i$.

Diagnosis penyakit-penyakit erythemato-squamuous

Diagnosa untuk membedakan penyakit-penyakit *erythemato-squamuous* merupakan masalah yang nyata dalam dermatologi. Penyakit-penyakit tersebut secara bersama menunjukkan ciri-ciri klinis erythema dan scaling dengan perbedaan yang sangat kecil. Penyakit-penyakit yang termasuk ke dalam kelompok ini adalah: *psoriasis, seboreic dermatitis, lichen planus, pityriasis rosea, cronic dermatitis* dan *pityriasis rubra pilaris* (*Merz dan Murphy, 1996*).

Biasanya biopsi diperlukan untuk mendiagnosa, tetapi sayangnya penyakit-penyakit tersebut secara bersama-sama memiliki ciri-ciri histopatologis yang mirip. Kesulitan lain ketika melakukan diagnosa untuk membedakan penyakit-penyakit tersebut adalah satu jenis penyakit mungkin menunjukkan ciri-ciri histopatologis jenis penyakit lain pada tahap permulaan dan baru menunjukkan karakteristiknya pada tahap berikutnya (*Merz dan Murphy, 1996*).

Dalam dunia kedokteran, langkah-langkah yang dilakukan untuk menentukan jenis penyakit tersebut terdiri dari dua tahap. Tahap pertama dilakukan pemeriksaan secara klinis terhadap pasien. Setelah itu, contoh kulit diambil untuk dievaluasi terhadap ciri-ciri histopatologis. Nilai-nilai ciri-ciri histopatologis ditentukan dengan analisis contoh kulit tersebut menggunakan mikroskop (*Merz dan Murphy, 1996*).

METODE PENELITIAN

Pengumpulan Data

Data yang digunakan berasal dari (*Merz dan Murphy, 1996*) yang merupakan hasil pemeriksaan 366 pasien yang menderita penyakit *erythemato-squamuous* dengan distribusi seperti yang tertera pada **Tabel 1**.

Tabel 1. Distribusi kelas jenis penyakit erythemato-squamous

Kode kelas	Kelas	Jumlah
1	Psoriasis	112
2	seboreic dermatitis	61
3	lichen planus	72
4	Pityriasis rosea	49
5	cronic dermatitis	52
6	Pityriasis rubra pilaris	20

Dalam dunia kedokteran, langkah-langkah yang dilakukan untuk menentukan jenis penyakit tersebut terdiri dari dua tahap.

Tabel 2. Daftar atribut hasil pemeriksaan klinis

Var	Keterangan	Nilai
X1	Erythema	0,1,2,3
X2	Scaling	0,1,2,3
X3	Definite borders	0,1,2,3
X4	Itching	0,1,2,3
X5	Koebner phenomenon	0,1,2,3
X6	polygonal papules	0,1,2,3
X7	follicular papules	0,1,2,3
X8	oral mucosal involvement	0,1,2,3
X9	knee and elbow involvement	0,1,2,3
X10	scalp involvement	0,1,2,3
X11	family history	0 atau 1
X34	age	linier

Tahap pertama dilakukan pemeriksaan secara klinis terhadap pasien. Pemeriksaan tersebut meliputi 12 ciri-ciri klinis (Tabel 2.), diantaranya adalah umur, erythema dan scaling. Setelah itu, contoh kulit diambil untuk dievaluasi terhadap 22 ciri-ciri histopatologis (Tabel 3.). Nilai-nilai ciri-ciri histopatologis ditentukan dengan analisis contoh kulit tersebut menggunakan mikroskop.

Secara keseluruhan terdapat 34 ciri-ciri klinis dan histopatologis yang digunakan

untuk menentukan jenis penyakit erythemato-squamous. Family history memiliki nilai 1 jika terdapat riwayat jenis penyakit tersebut pada keluarga pasien yang sedang didiagnosis dan nilai 0 jika tidak ditemukan. Age menunjukkan umur pasien. Setiap ciri-ciri (klinis dan histopatologis) diberikan derajat dari 0 sampai dengan 3. Nilai 0 menunjukkan tidak terdapat ciri-ciri tertentu, nilai 3 menunjukkan kemungkinan terbesar adanya ciri-ciri tertentu, sedangkan nilai 1 dan 2 menunjukkan nilai pertengahan.

Tabel 3. Daftar atribut hasil pemeriksaan histopatologi

Var	Keterangan
X12	melanin incontinence
X13	eosinophils in the infiltrate
X14	PNL infiltrate
X15	fibrosis of the papillary dermis
X16	exocytosis
X17	acanthosis
X18	hyperkeratosis
X19	parakeratosis
X20	clubbing of the rete ridges
X21	elongation of the rete ridges
X22	thinning of the suprapapillary epidermis
X23	spongiform pustule
X24	munro microabcess
X25	focal hypergranulosis
X26	disappearance of the granular layer
X27	vacuolisation and damage of basal layer
X28	spongiosis
X29	saw-tooth appearance of retes
X30	follicular horn plug
X31	perifollicular parakeratosis
X32	inflammatory mononuclear infiltrate
X33	band-like infiltrate

Dari data 366 pasien tersebut, terdapat 8 pasien yang tidak tercatat umurnya (age) sehingga data yang digunakan hanya data yang lengkap yaitu data 358 pasien.

Membagi data menjadi 3 bagian untuk melakukan validasi silang

Sebelum dibuat model neural network maupun model regresi logistik nominal, terlebih dahulu data dibagi menjadi 3 bagian terpisah. Dua bagian akan digunakan sebagai data training

dan satu bagian lagi digunakan sebagai data *testing* sehingga terdapat tiga pasang data *training-testing*. Ketiga pasang data tersebut selanjutnya disebut data **bagian 1**, data **bagian 2** dan data **bagian 3**.

Model neural network

Model neural network yang akan dibuat merupakan *multi layer perceptron* dengan *learning rule backpropagation* terdiri dari *layer input* dengan 34 *node*, *layer hidden* 6 *node* dan *layer output* terdiri dari 6 *node*.

Menetapkan pembobot awal

Pembobot awal yang akan ditetapkan adalah pembobot awal pada koneksi antara *layer input* dan *layer hidden (layer ke-1)* serta antara *layer hidden* dan *layer output (layer ke-2)*.

Terdapat 4 cara penentuan pembobot awal, yaitu:

1. Pembobot awal *layer ke-1* ditetapkan dengan menggunakan *regresi logistik biner* dan pembobot awal *layer ke-2* ditetapkan dengan *regresi linier* berbentuk $y_k = f(h_1, h_2, h_3, h_4, h_5, h_6)$ dengan $k=1,2,3,4,5,6$, sedangkan h_i adalah *output* dari unit *hidden ke-i*.
2. Pembobot awal *layer ke-1* ditetapkan dengan menggunakan *regresi logistik biner* dan pembobot awal *layer ke-2* ditetapkan dengan *regresi linier* berbentuk $y_k = f(h_k)$ dengan $k=1,2,3,4,5,6$.
3. Pembobot awal *layer ke-1* ditetapkan dengan menggunakan *regresi logistik biner* dan pembobot awal *layer ke-2* ditetapkan nilainya sama dengan nol.
4. Pembobot awal *layer ke-1* dan *layer ke-2* ditetapkan secara acak, demikian juga dengan nilai *threshold*.

Berdasarkan penentuan bobot awal tersebut, terdapat 4 model neural network yaitu NN-1, NN-2, NN-3 dan NN-4 (*Tabel 4*).

Tabel 4. Empat model neural network

Model	Pembobot awal	
	Layer ke-1	Layer ke-2
NN-1	Regresi logistik	regresi linier
NN-2	biner	regresi linier
NN-3		nol
NN-4	acak	acak

Untuk menentukan pembobot awal *layer ke-1*, digunakan *regresi logistik biner* untuk setiap kategori. Caranya adalah dengan memberi nilai 1 pada peubah hasil dari kategori yang sedang dibuat *regresi logistiknya* dan memberi nilai 0 pada peubah hasil kategori lainnya. Dengan demikian akan diperoleh 6 model *regresi logistik biner*, masing-masing untuk tiap kategori.

Metode yang digunakan untuk membuat model *regresi logistik biner* tersebut adalah menggunakan metode *stepwise (Hosmer and Lemeshow, 1989)* dengan nilai *pE* ditentukan sebesar 0.15 dan nilai *pR* 0.20.

Training neural network

Model neural network akan ditraining pada berbagai nilai epoch dan nilai *learning rate* untuk mendapatkan nilai ketepatan prediksi yang tinggi.

Model Regresi Logistik Nominal

Model *regresi logistik nominal* yang dibuat sebanyak 3 model, masing-masing untuk setiap data training. *Regresi logistik nominal* dibuat antara peubah hasil (*terdapat 6 kategori*) dengan 34 variabel bebas. Metode yang digunakan untuk memilih di antara 34 variabel yang dimasukkan ke dalam model adalah metode *stepwise (Hosmer and Lemeshow, 1989)*. Model *regresi logistik nominal* selanjutnya disebut model *RLN*.

Pengujian model neural network

Pengujian (*testing*) dilakukan dengan menggunakan data *testing* yang merupakan pasangan dari data training yang digunakan untuk mencari pembobot model neural network.

Kemudian dilakukan pengukuran terhadap ketepatan model dan nilai *mean square error (mse)* model.

Pengujian model regresi logistik nominal

Pengujian model dilakukan menggunakan data testing. Model yang telah dibangun dengan data training digunakan untuk menduga nilai variabel hasil pada data testing. Selanjutnya dihitung ketepatan dan *mse* model.

Perbandingan antara model neural network dengan model regresi logistik nominal

Perbandingan antara kedua model dilakukan dengan membandingkan ketepatan prediksinya. Ketepatan model merupakan prosentase klasifikasi yang benar.

Selain itu juga akan dianalisis pengaruh penentuan pembobot awal menggunakan regresi logistik biner dan regresi linier terhadap ketepatan model neural network.

HASIL DAN PEMBAHASAN

Model Neural Network

Ketepatan Model Neural Network Sebelum Proses Training

Pada model *NN-1*, *NN-2* dan *NN-4*, pembobot awal layer ke-1 dan layer ke-2 telah ditetapkan sehingga dapat diketahui ketepatan model sebagai hasil pembobotan awal tersebut.

Ketepatan tertinggi model *NN-1*, untuk data training terjadi pada data **bagian 3** sedangkan untuk data testing terjadi pada **bagian 2** (*Tabel 5*). Data **bagian 2** memiliki ketepatan data training paling rendah namun memiliki ketepatan data testing tertinggi.

Tabel 5. Ketepatan model *NN-1* sebelum training

Data	training		testing	
	mse	Ketepatan (%)	mse	Ketepatan (%)
Bagian 1	0,012	93,28	0,014	89,17
Bagian 2	0,011	92,02	0,017	90,00
Bagian 3	0,011	94,17	0,026	83,90

Pola tersebut juga terjadi pada model *NN-2* namun dengan nilai yang lebih kecil (*Tabel 6*).

Tabel 6. Ketepatan model *NN-2* sebelum training

Data	training		testing	
	mse	Ketepatan (%)	mse	Ketepatan (%)
Bagian 1	0,015	89,08	0,021	83,33
Bagian 2	0,013	90,76	0,020	86,67
Bagian 3	0,013	92,50	0,031	76,27

Ketepatan model *NN-1* lebih tinggi dibandingkan dengan model *NN-2* sedangkan *mse* model *NN-1* lebih rendah dibandingkan *NN-2* untuk tiap bagian data. Hal tersebut berlaku baik untuk data *training* maupun data *testing*. Perbedaan tertinggi ketepatan prediksi terdapat pada data testing bagian 3 yaitu untuk *NN-1* 83,90% sedangkan *NN-2* 76,27%.

Ketepatan model *NN-1* yang lebih tinggi dari *NN-2* menunjukkan bahwa pembobot awal layer ke-2 dengan model $y_k = f(h1, h2, h3, h4, h5, h6)$ lebih baik dibandingkan dengan model $y_k = f(h_k)$.

Pada model *NN-4*, pembobot awal yang ditetapkan secara acak menghasilkan nilai ketepatan 0% untuk data training dan testing (*Tabel 7*). Hal tersebut menunjukkan bahwa sebelum proses training, model *NN-4* tidak menghasilkan satu pun output yang benar.

akibat dari pembagian data menjadi 3 bagian untuk proses validasi silang yang dilakukan secara acak.

Tabel 11. Hasil tertinggi model NN-4

Data	training		testing	
	mse	Ketepatan (%)	mse	Ketepatan (%)
Bagian 1	0,079	36,56	0,081	35,83
Bagian 2	0,053	72,27	0,054	72,50
Bagian 3	0,083	56,25	0,093	52,54

Ketepatan tertinggi model NN-4 sebesar 75,50% dicapai pada data bagian 2 dan terendah sebesar 35,83% dicapai pada data bagian 1 (Tabel 11). Hasil tersebut jauh lebih kecil dibandingkan dengan model NN-1, NN-2 dan NN-3.

Hasil tersebut juga sejalan dengan hasil yang dicapai sebelum proses training yang berarti bahwa inisialisasi pembobot awal menggunakan regresi logistik dan regresi linier memberikan hasil yang lebih baik dibandingkan dengan inisialisasi pembobot awal secara acak

Model Regresi Logistik Nominal (RLN)

Ketepatan tertinggi model RLN sebesar 75,00% dicapai pada data bagian 2 dan terendah sebesar 61,86% dicapai pada data bagian 3 (Tabel 12) sama dengan pola yang terdapat pada model NN-1, NN-2 dan NN-3 namun dengan nilai yang jauh lebih kecil.

Tabel 12. Ketepatan model RLN tiap bagian data

Data	training		testing	
	mse	Ketepatan (%)	mse	Ketepatan (%)
Bagian 1	0.259	78.15	0.391	74.17
Bagian 2	0.302	73.95	0.275	75.00
Bagian 3	0.251	80.00	0.465	61.86

Ketepatan Rata-rata Model NN Setelah Training dan Model RLN

Tabel 13 menunjukkan ketepatan rata-rata keempat model neural network dan model RLN yang merupakan rata-rata ke-3 bagian data.

Pada Tabel 13 dapat dilihat bahwa, untuk data training, mse dan ketepatan model untuk ke-3 model dengan pembobotan awal regresi logistik biner, yaitu NN-1, NN-2 dan NN-3, memiliki nilai yang sama sebesar 0,02 dan 99,44%. Demikian juga untuk data testing, ketiga model menunjukkan nilai yang hampir sama. Hal tersebut memperkuat hasil-hasil sebelumnya bahwa pembobotan awal hanya perlu dilakukan pada layer ke-1 dengan regresi logistik biner dan layer ke-2 cukup diberi pembobot nol. Ketepatan model NN-3, pembobot layer ke-2 nol, untuk data testing 91,04% lebih baik dibandingkan model NN-1 (91,03%).

Ketepatan rata-rata model NN-1, NN-2 dan NN-3 lebih tinggi dibandingkan dengan model NN-4 maupun model RLN. Hasil tersebut sejalan dengan hasil yang diperoleh bagi model-model tersebut sebelum dilakukan proses training.

Pada Tabel 13 juga dapat dilihat bahwa ketepatan model RLN lebih tinggi dibandingkan dengan model NN-4 baik untuk data training maupun data testing. Demikian juga dengan mse untuk model RLN lebih kecil dibandingkan dengan mse model NN-4.

Tabel 13. Ketepatan rata-rata model NN dan model RLN

Model	Training		testing	
	mse	Ketepatan (%)	mse	Ketepatan (%)
NN-1	0,002	99,44	0,014	91,03
NN-2	0.002	99.44	0.015	91.04
NN-3	0,002	99,44	0,015	91,04
NN-4	0,072	55,02	0,076	53,63
RLN	0,271	77,37	0,377	70,34

Dengan demikian, dari ke-5 model yang dibuat, model terbaik adalah model neural network dengan pembobot awal layer ke-1 menggunakan regresi logistik biner dan pembobot layer ke-2 ditentukan nilainya nol, yaitu model NN-3. Model NN-2 meskipun memiliki ketepatan sama dengan NN-3 namun masih diperlukan usaha lagi untuk mencari pembobot awal layer ke-2.

Perbandingan dengan Hasil Penelitian Lain

Tabel 14 menunjukkan ketepatan yang diperoleh pada penelitian ini dan hasil-hasil dari beberapa penelitian lain. Ketepatan tertinggi, untuk data testing, pada penelitian ini adalah pada model NN-2 dan NN-3 yaitu 91,04%. Hasil tersebut lebih tinggi dibandingkan dengan algoritma c4.5, kernel-unsupervised dan NNA namun lebih rendah dibandingkan hasil penelitian lainnya. Untuk data training, ketepatan model NN-1, NN-2 dan NN-3 sebesar 99,44% lebih tinggi dibandingkan dengan hasil-hasil penelitian lain.

Tabel 14. Ketepatan yang diperoleh dan hasil-hasil penelitian lain

model	Ketepatan (%)	
	training	testing
NN-1	99,44	91,03
NN-2	99,44	91,04
NN-3	99,44	91,04
NN-4	55,02	53,63
RLN	77,37	70,34
Vfi ¹⁾		99,20
GA ²⁾		95,00
cs gp ³⁾		96,64
c4.5 ³⁾		89,12
kernel-supervised ⁴⁾	95,14	94,47
kernel-unsupervised ⁴⁾	82,66	81,48
Neural network dan apriori (NNA) ⁵⁾	93,00	68,00

Ket.: ¹⁾ Guvenir (2000), ²⁾ Fidelis (2000), ³⁾Bojarczuk (2001), ⁴⁾ Castelano (2003), ⁵⁾ Lay (1999)

Hasil yang diperoleh pada penelitian ini telah meningkatkan ketepatan prediksi model neural network dengan algoritma propagasi balik dari 93,00% menjadi 99,44 % untuk data training dan meningkatkan ketepatan dari 68% menjadi 91,04% untuk data testing.

KESIMPULAN DAN SARAN

Kesimpulan

Penentuan pembobot awal menggunakan regresi logistik biner dan regresi linier menghasilkan ketepatan model yang lebih tinggi dibandingkan penentuan pembobot awal secara acak.

Penentuan pembobot awal layer ke-2 menggunakan regresi linier tidak menghasilkan ketepatan yang lebih tinggi dibandingkan dengan pembobot awal nol.

Model neural network dengan pembobot awal ditentukan melalui regresi logistik biner memiliki ketepatan lebih tinggi dibandingkan dengan model regresi logistik nominal.

Ketepatan model neural network dengan pembobot awal ditentukan dengan regresi logistik biner lebih tinggi dibandingkan model neural network yang dilanjutkan dengan algoritma apriori.

Saran

Untuk meningkatkan ketepatan prediksi model neural network, pada saat membagi data untuk validasi silang perlu diperhatikan keragaman datanya sehingga tiap-tiap bagian data memiliki keragaman yang hampir sama.

DAFTAR PUSTAKA

Bojarczuk, C.C; H.S. Lopes; A.A. Freitas. (2001). *Data Mining with Constarined-Syntax Genetic*

- Programming: Applications in Medical Data Sets*.
http://www.magix.fri.unilj.si/idama_p2001/papers/bojarczuk.pdf
- Castelano, G., C. Castiello, A.M. finelli dan C. Leone. (2003). *Diagnosis of Dermatological Diseases by A Neuro-Fuzzy System*.
<http://www.di.uniba.it/~castella/papers/eusflat2003.pdf>
- Fidelis, M.V.; H.S. Lopes; A.A. Freitas. (2000). *Discovering Comprehensive Classification Rules with a Genetic Algorithm*. Proc. Congress on Evolutionary Computation – 2000 (CEC-2000), 808-820, La Jolla, CA, USA.
- Fu, Limin. (1994). *Neural Networks in Computer Intelligence*. McGraw-Hill Inc, Singapore.
- Güvenir H.A., and N. Emeksiz (2000). *An Expert System for the Differential Diagnosis of Erythematous-Squamous Diseases*. Expert Systems With Applications. Vol. 18, No. 1, (2000), pp. 43-49
- Güvenir H.A., G. Demiröz and N. Ilter (1998). *Learning Differential Diagnosis of Erythematous-Squamous Diseases using Voting Feature Intervals*. Artificial Intelligence in Medicine, Vol. 13, No. 3, (1998), pp. 147-165
- Hosmer D.W. and S. Lemeshow. (1989). *Applied Logistic Regression*. John Wiley & Sons, Inc. New York.
- Lay, C.B., M. Khalid, dan R. Yusuf. (1999). *Intelligent Database by Neural Network and Data Mining*. [http://www.cairo.utm.my/blchua_datamining.pdf]
- Merz, C. and Murphy, P.M. (1996). *UCI Repository of machine learning databases* [<http://www.ics.uci.edu/pub/machine-learning-databases>] University of California, Irvine, Department of Information and Computer Science.
- Schalkoff, Robert J (1992). *Pattern Recognition: Statistical, structural and neural approaches*. John Wiley & Sons. Singapore.
- Schumacher, M., Reinhard Robner, and Warner Vach (1996). *Neural networks and logistic regression : Part I. Computational Statistics and Data Analysis*. 21:661-82.
- Walpole, R.E. (1990). Pengantar Statistika edisi ke-3 (edisi terjemahan). P.T. Gramedia, Jakarta.
- Warner, B. and Manavendra Misra (1996). *Understanding Neural Networks as Statistical Tools*. The American Statistician Journal. volume 50. number 4 :284--.
- Wu, F.L. and F.C. Su. (1998). *Comparison of Genetic Algorithm Neural Network with Logistic Regression for Prediction of Gait Pattern on Ankle Arthrodesis*. <http://asb-biomech.org/onlineabs/NACOB98/162/>