

PERBANDINGAN MODEL *LOGISTIC REGRESSION* DAN *NEURAL NETWORK* PADA KELAHIRAN BAYI DENGAN BERAT BADAN RENDAH

M. HASRUL H.

Teknik Informatika Fakultas Sains dan Teknologi
Universitas Islam Negeri Alauddin Makassar
E-mail: muhammad.hasrul@uin-alauddin.ac.id

ABSTRAK

Evaluasi risiko kelahiran bayi dengan berat badan lahir rendah disingkat BBLR menjadi suatu persoalan yang menarik untuk dibahas. Sejumlah penelitian bidang kesehatan khususnya persalinan bertujuan meminimalisir risiko BBLR telah banyak dilakukan dalam usaha mengelak bayi lahir secara abnormal. Data dikelompokkan menjadi dua kelas, bayi dengan berat badan rendah dan bayi dengan berat normal. Sehingga teknik klasifikasi *data mining* menjadi tepat untuk diterapkan. Penelitian ini mengenai perbandingan algoritma *Logistic Regression* dan *Neural Network* yang diterapkan pada data kelahiran bayi berat badan rendah (*low birth height*). Hasil eksperimen menunjukkan bahwa *Neural Network* lebih baik dalam melakukan klasifikasi risiko BBLR 97,37% dibanding *Logistic Regression* dengan akurasi sebesar 94,71. Penerapan klasifikasi data mining menggunakan *Logistic Regression* dan *Neural Network* dapat menghasilkan nilai AUC yang baik (*excellence classification*) sehingga dapat diimplementasikan untuk menentukan risiko BBLR.

Kata Kunci: *Bayi berat badan lahir rendah, Klasifikasi, Logistic Regression, Neural Network;*

I. PENDAHULUAN

Kemajuan teknologi informasi yang dinamis sekarang ini, manusia sering berupaya dalam memenuhi keperluan akan mudahnya dalam pengolahan data penyajian dan pencariannya. Di segala lini memerlukan kemudahan untuk penanganan informasi yang mereka miliki, tidak terkecuali pada studi persalinan. Peningkatan akan kualitas sumber daya manusia sebaiknya dilaksanakan sejak dini supaya bayi yang dilahirkan memiliki potensi yang tinggi untuk meraih tingkat produktivitas yang maksimal. Dalam hal ini yang dimaksudkan yakni agar sejak masih dalam kandungan ibu, keadaan kesehatan dan terawat serta gizi janin harus

dalam kondisi baik. Kualitas bayi yang dilahirkan oleh ibu bergantung dengan kondisi gizi sebelum mengandung dan selama mengandung yang dimana pertumbuhan saraf terjadi saat janin masih berusia dini. Berat badan bayi saat lahir rendah menjadi salah satu acuan untuk mengukur kualitas bayi atau generasi penerus. Pada umumnya, bayi akan kurang dapat meredam adanya tekanan lingkungan yang baru dimana dapat berdampak pada terhambatnya tumbuh kembang hingga mengganggu kelangsungan hidup dari bayi serta meningkatnya risiko kesehatan (morbiditas) dan kematian (mortalitas) bayi karena mudah terkena penyakit terhadap gangguan belajar, masalah perilaku, keadaan infeksi saluran pernafasan bagian bawah dan lain-lain. Oleh sebab itu BBLR menjadi salah satu indikator tingkat kesehatan pada suatu masyarakat (Saraswati & Sumarno, 1998).

Pada masa perang dunia I di Eropa, kurangnya pangan terhadap kegagalan kehamilan termasuk kelahiran bayi BBLR saling terkait. Namun, para ibu yang mengalami kekurangan protein yang ringan dan energi masih dapat melahirkan bayi secara normal dan sehat. Hingga saat ini, yang masih belum diketahui ialah pada tingkat berapa kekurangan protein dan energi yang dapat mengakibatkan pada kelahiran bayi dengan bobot badan lahir kurang dari 2,5 kilogram. Diduga bahwa ibu hamil yang mengalami kekurangan energi kronik (KEK) lebih dominan dan berpotensi BBLR. Dengan demikian, pada saat sebelum sampai kehamilan berlangsung terjadi kekurangan protein dan energi terjadi. Anemia yang berat waktu kehamilan juga dapat menjadi penyebab bayi dengan BBLR. Secara logis, ini berarti ibu hamil yang mengalami KEK sekaligus menderita anemia mempunyai risiko yang lebih besar untuk melahirkan bayi dengan BBLR (Illsley, 1984). Evaluasi risiko melahirkan bayi BBLR menjadi salah satu persoalan yang menarik untuk dibahas. Sejumlah penelitian bidang kesehatan khususnya persalinan untuk meminimalisir risiko bayi dengan BBLR telah banyak diteliti dalam usaha menjauhi bayi lahir secara abnormal. Dalam penyelesaian masalah analisis risiko bayi dengan BBLR, data akan dikategorikan menjadi dua kelas yakni bayi dengan berat badan normal dan bayi dengan berat badan rendah. Dengan demikian klasifikasi *data mining* tepat untuk diterapkan. *Data mining* merupakan sebuah

runtunan dimana data yang sudah ada di dalam basis data akan ditemukan polanya (*pattern*) secara otomatis atau semi otomatis yang ditujukan untuk menyelesaikan dan menjadi solusi suatu masalah (Witten, Frank, & Hell, 2011).

Terdapat beberapa teknik yang dimiliki oleh *Data mining*, yakni klasifikasi dan *clustering*. Klasifikasi bekerja memecah objek yang ditugaskan hanya ke satu nomor kategori yang biasa disebut kelas. Klasifikasi merupakan salah satu teknik pembelajaran yang difungsikan untuk mengestimasi nilai dari atribut kategori target. Teknik *clustering* bekerja mengkategorikan data atau objek berdasar pada kemiripan antar data, dengan demikian anggota pada satu kelompok mempunyai banyak kemiripan jika dibandingkan dengan kelompok lain. (Bramer, 2013).

Dalam penelitian ini, akan dibahas mengenai penerapan algoritma *Logistic Regression* dan *Neural Network* untuk penentuan risiko kelahiran bayi berat badan rendah yang diharapkan dapat menjadi solusi menurunkan kelahiran bayi berat badan rendah.

II. METODE PENELITIAN

A. KLASIFIKASI

Metode klasifikasi adalah proses penemuan fungsi atau model yang mendiskripsikan dan membandingkan kelas data atau konsep yang digunakan dengan tujuan agar mampu memprediksi kelas dari objek yang label kelasnya tidak diketahui (Han, Pei, & Kamber, 2011). Algoritma klasifikasi yang sering digunakan yakni *K-Nearest Neighbor (K-NN)*, *Naïve Bayes/ Bayesian classifiers*, *Decision trees*, Jaringan Saraf Tiruan (*Neural Networks*), *Support Vector Machines (SVM)*, *Algoritma Genetika*, *Metode Rule Based*, *Analisa Statistik*, *Rough sets* dan *Memory based reasoning*.

Ada 2 proses yang tersusun dalam klasifikasi data. Pertama adalah fase pelatihan atau pembelajaran (*learning*), dimana algoritma klasifikasi dibuat dengan tujuan melakukan analisis data pelatihan selanjutnya dihadirkan dalam bentuk aturan klasifikasi. Kedua adalah dimana data pengujian dipakai untuk estimasi nilai

akurasi dari aturan klasifikasi. Ada empat komponen yang menjadi dasar Proses klasifikasi (Gorunescu, 2011):

a. Kelas

Kelas merupakan variabel terikat bertipe kategorikal yang direpresentasikan dalam bentuk ‘label’ yang terdapat pada suatu objek. Contohnya: *credit risk*, risiko penyakit pada pasien, gempa bumi, *customer loyalty* dan sebagainya.

b. *Predictor*

Predictor merupakan variabel bebas yang direpresentasikan oleh atribut data atau karakter. Contohnya: penghasilan, jumlah simpanan atau tabungan, merokok, mengkonsumsi minuman beralkohol dan sebagainya.

c. Dataset Pelatihan (*Training Dataset*)

Dataset Pelatihan terdiri dari satu set data yang mengandung nilai dari dua komponen yang di atas dan difungsikan untuk menentukan kelas yang sesuai berdasarkan *predictor*.

d. *Testing Dataset*

Dataset Pengujian berisi data baru yang selanjutnya dikategorikan terhadap model atau fungsi yang telah dibuat dan nilai akurasi yang diamati.

B. LOGISTIC REGRESSION

Logistic Regression atau regresi logistik adalah model atau fungsi matematis untuk memperkirakan *dependent variabel* (dikotomi) dari beberapa variabel independen yang bersifat kontinyu dan kategorik. Regresi logistik tidak membutuhkan asumsi kenormalan baik pada sebaran *error* maupun pada peubah responnya dan relasi antara peubah respon dengan peubah bebas tidak perlu diasumsikan linier. Tetapi terdapat asumsi yang diberlakukan yakni ketiadaan multikolinieritas antar peubah bebas (Hosmer, Lemeshow, & Sturdivant, 2013). Adapun model regresi logistik adalah sebagai berikut:

$$\ln \left[\frac{p}{1-p} \right] = a + b_1x_1 + b_2x_2 + \dots + b_kx_k$$

Dimana:

$\ln(p/(1-p))$ merupakan logodds (logit). Logaritma alami dari odds. Odds merupakan perbandingan peluang satu peristiwa untuk tidak terjadi dan peluang satu peristiwa untuk terjadi. a merupakan konstanta (intersep); b_1, b_2, \dots, b_k merupakan koefisien regresi variabel prediktor; x_1, x_2, \dots, x_k merupakan variabel prediktor / bebas yang pengaruhnya akan diteliti; p merupakan peluang untuk terjadinya “peristiwa” dari variabel terikat yang dikotomus.

Berikut ini langkah-langkah melakukan analisis regresi logistik.

1. Melakukan analisis regresi logistik univariat sederhana pada setiap variabel terikat dengan variabel bebas.
2. Mencari potensial kovariat menggunakan regresi logistik univariat, yaitu variabel yang secara substansi berpengaruh terhadap variabel dependen dan signifikan ($p\text{-value} < 0,25$) sesuai yang direkomendasikan oleh Mickey dan Greenland.
3. Melakukan deteksi *collinearities*, untuk mengecek apakah antar variabel independen terdapat korelasi. Dikatakan ada gejala kolinieritas jika nilai $r > 0,8$.
4. Melakukan analisis regresi logistik multivariat dengan semua potensial kovariat
5. Melakukan eliminasi variabel independen secara satu per satu menggunakan uji statistik Wald. Eliminasi diawali dari variabel yang nilai p nya paling besar, sampai semua variabel yang masuk model signifikan semua ($p < 0,05$).
6. Melakukan uji variabel interaksi yaitu melakukan regresi logistik multivariat dengan semua variabel hasil langkah 5 ditambah dengan variabel interaksi yang mungkin (didasari logika substantif) secara bersama. Kemudian mengeluarkan variabel interaksi dimulai dari yang paling tidak signifikan sampai didapat variabel yang signifikan semua.

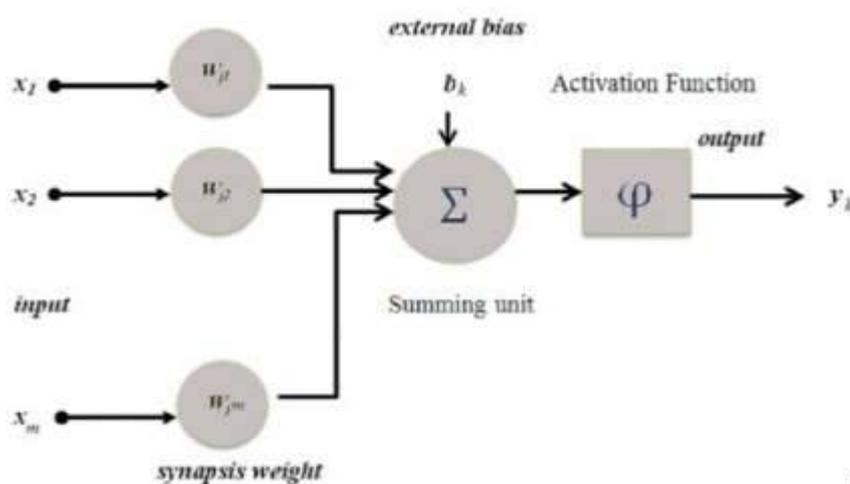
C. NEURAL NETWORK

Jaringan Saraf Tiruan (*Artificial Neural Network*) diadopsi dari saraf makhluk hidup. *Neural Network* tersusun atas bagian pemroses yang menyerupai sejumlah tautan terboboti di antara bagian-bagian tersebut dan neuron (Fu, 2003). *Neural network* mempunyai arsitektur paralel-terdistribusi dengan sejumlah besar node dan tautan. Setiap tautan menghubungkan satu node dengan node yang lain dan mempunyai pembobot tertentu.

Topologi *Neural network* ditentukan oleh struktur dan skema interkoneksi. Struktur sering dinyatakan sebagai jumlah layer dan jumlah node per layer. Keunggulan *Neural Network* adalah kehandalan *robust* pada *missing data*, memprediksi dan kecepatan waktu (Santosa, 2007).

Unit pemroses informasi dalam jaringan saraf tiruan disebut Neuron terdiri atas:

1. Penghubung atau Set Sinapsis ditandai dengan adanya massa (*weight*).
2. Unit Penjumlah, untuk menambahkan sinyal masukan yang diberi massa
3. Fungsi aktivasi, berfungsi merestriksi besarnya keluaran dari sebuah neuran.



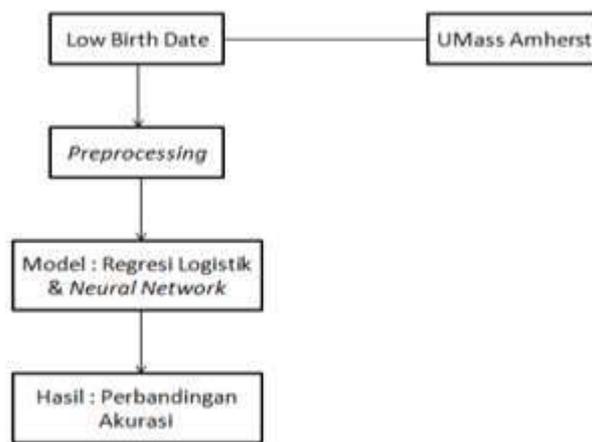
Gambar 1. Model Neuron

III.HASIL DAN PEMBAHASAN

A.DESAIN EKSPERIMEN

Penelitian ini menggunakan data yang berasal dari *University of Massachusetts Amherst* dengan judul *Low Birth Weight Data*. Dataset ini dikumpulkan dari 189 *record* observasi kelahiran ibu dan terdiri dari 11 atribut.

Percobaan dilaksanakan dengan melakukan *testing* terhadap dataset dilanjut dengan klasifikasi berat badan pada kelahiran bayi dengan menggunakan model Logistik Regresi serta membandingkannya dengan model *Neural Network* dan mengaplikasikan dalam *framework* RapidMiner. Dari hasil percobaan akan diperoleh tingkat atau nilai akurasi. Bagan alir dari percobaan ini ditunjukkan gambar 2 di bawah.



Gambar 2. Bagan Alir Desain Eksperimen

D. PEMBAHASAN

Setelah melakukan metode klasifikasi terhadap dataset *Low Birth Weight* menggunakan algoritma *Logistic Regression* dan *Neural Network*, maka

didapatkan hasil perbandingan untuk tingkat akurasi tertera pada tabel 1 di bawah.

Tabel 1 Perbandingan Akurasi

<u>Algoritma</u>	<u>Akurasi</u>	<u>AUC</u>
<i>Logistic Regression</i>	94.71%	0.965
<i>Neural Network</i>	97.37%	0.999

Berdasarkan pada Tabel 2 yang ditunjukkan di bawah ini dapat diketahui bahwa Algoritma *Neural Network* memiliki nilai akurasi tertinggi sebesar **97.37%** sedangkan nilai akurasi sebesar **94.71%** dicapai dengan menggunakan algoritma *Logistic Regression*.

Tabel 2 Klasifikasi AUC

<u>Performance</u>	<u>Klasifikasi</u>
0.90 – 1.00	<u>Paling baik</u>
0.80 – 0.90	<u>Baik</u>
0.70 – 0.80	<u>Adil atau sama</u>
0.60 – 0.70	<u>Rendah</u>
0.50 – 0.60	<u>Gagal</u>

Dari hasil perhitungan AUC dari masing-masing algoritma, dapat diketahui bahwa dalam penelitian ini algoritma *logistic regression* (**0.965**) dan *neural network* (**0.999**) sehingga kedua algoritma tersebut termasuk dalam klasifikasi paling baik.

IV.KESIMPULAN

Berdasarkan perbandingan kedua algoritma klasifikasi *data mining* penelitian ini yakni *logistic regression* dan *neural network* untuk prediksi kelahiran bayi dengan berat badan rendah (*low birth weight*) dengan menggunakan 189 sampel data kelahiran pada ibu hasil pengumpulan *University of Massachusetts Amherst*, maka dapat dianalisa hasilnya bahwa model *neural network* merupakan algoritma yang paling akurat jika dibandingkan dengan algoritma *logistic regression*. Sedangkan untuk nilai area di bawah kurva, *logistic regression* dan *neural network* termasuk dalam kategori klasifikasi paling baik.

DAFTAR PUSTAKA

- Bramer, M. (2013). *Principles of Data Mining*. London: Springer.
- Fu, L. M. (2003). *Neural Networks in Computer Intelligence*. Tata McGraw-Hill Education.
- Gorunescu, F. (2011). *Data Mining: Concepts, Models and Techniques*. Springer.
- Han, J., Pei, J., & Kamber, M. (2011). *Data Mining: Concepts, Models, and Techniques*. Verlag Berlin Heidelberg: Springer.
- Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied Logistic Regression*. New York: John Willey and Sons.
- Illsley, R. (1984). *Low birth weight: a medical, psychological, and social study*. John Wiley & Sons Inc.
- Santosa, B. (2007). *Data Mining Teknik Pemanfaatan Data Untuk Keperluan Bisnis*. Yogyakarta: Graha Ilmu.
- Saraswati, E., & Sumarno, I. (1998). Risiko Ibu Hamil Kurang Energi Kronis dan Anemia Untuk Melahirkan Bayi dengan Berat Badan Lahir Rendah. *PGM*, 41-49.
- Statistical Software Information*. (2015, January 15). Retrieved from University of Massachusetts Amherst:
<https://www.umass.edu/statdata/statdata/data/lowbwt.txt>
- Witten, I. H., Frank, E., & Hall, M. A. (2011). Data mining: practical machine learning tools and technique. *ACM SIGSOFT Software Engineering Notes*, 76-77.