

# Penerapan Algoritme Genetik untuk Seleksi Peubah Regresi Logistik

Dian Ayuningtyas\*, Bagus Sartono\*, Farit Mochamad Afendi\*

\*Departemen Statistika Institut Pertanian Bogor

**Abstrak**—In a study, interaction factors are the potential to have important effects on the response variable. But research involving interaction factors often encounters two problems, namely the excessive number of variables and the difficulty of implementing the heredity principle. The alternative solution is to do variable selection using a metaheuristic optimization method. In this study, the logistic regression variable selection was done using a genetic algorithm. The genetic algorithm is modified so that every independent variable has a different probability to be included in the model. That probability is based on the absolute value of the correlation of the independent variable with the response variable. These modifications have a positive effect on the results of variable selection. To choose significant independent variables, 30 repetitions of the genetic algorithm can be performed using the objective function AIC. Of the 30 repetitions, if a variable appears in all formed models, then the variable is an independent variable that has a significant effect on the response variable. The application of this method to Myopia data can show significant variables well.

**Kata kunci**—algoritme genetik, klasifikasi, regresi logistik, seleksi peubah, simulasi

## I. PENDAHULUAN

### A. Latar Belakang

Penelitian ilmiah seringkali dibuat untuk mengidentifikasi hal-hal yang memengaruhi sesuatu hal yang lain. Dalam banyak kasus, pengaruh suatu peubah bebas terhadap peubah respon dapat dipengaruhi oleh peubah bebas lainnya, hal ini yang sering disebut sebagai faktor interaksi. Dalam penelitian yang melibatkan interaksi, setidaknya berlaku dua prinsip utama yaitu efek hirarki dan efek hereditas. Efek hirarki menyatakan bahwa pengaruh dari faktor yang derajatnya lebih rendah lebih besar dibanding pengaruh faktor yang derajatnya lebih tinggi (Yuan M, 2007). Prinsip efek hereditas menyatakan bahwa interaksi hanya dapat aktif apabila salah satu atau kedua faktor utamanya aktif (Yuan M, 2007).

Penelitian yang melibatkan faktor interaksi secara umum menemui dua permasalahan. Permasalahan pertama adalah mengenai jumlah peubah bebas yang terlalu banyak. Permasalahan kedua adalah penerapan prinsip efek hereditas dalam model yang dihasilkan. Permasalahan pertama yang telah disebutkan sebelumnya dapat diatasi dengan seleksi peubah pada permodelan yang tepat. Model regresi logistik adalah suatu model regresi yang memiliki peubah respon yang berskala biner sehingga merupakan salah satu model klasifikasi. Model ini memiliki kelebihan yaitu interpretasi pengaruh masing-masing peubah bebas dapat dilakukan dengan cukup baik dibandingkan dengan model klasifikasi yang lain.

Banyak metode seleksi peubah yang dapat diterapkan pada analisis regresi logistik seperti *stepwise* (Wang D, 2004), *forward* (Tan H, 2005), dan *backward* (Wang D, 2004). Semua metode seleksi peubah yang telah disebutkan sebelumnya dapat menghasilkan akurasi yang cukup baik pada penelitian sebelumnya. Tetapi metode tersebut belum mampu menjawab permasalahan kedua, yaitu kesulitan penerapan prinsip hereditas. Untuk dapat menerapkan prinsip tersebut, maka metode yang digunakan haruslah suatu metode yang dapat diubah sesuai dengan kebutuhan.

Proses seleksi variabel dapat dilihat sebagai proses mengoptimalkan ukuran kebaikan model untuk mendapatkan peubah yang paling tepat masuk ke dalam model. Metode optimasi yang dipilih adalah algoritma genetik karena metode ini telah terbukti dapat memberikan solusi optimum yang cukup baik dengan waktu komputasi yang cukup cepat. Dengan alasan tersebut, penulis menulis skripsi dengan judul “Penerapan Algoritme Genetik untuk Seleksi Peubah Regresi Logistik”.

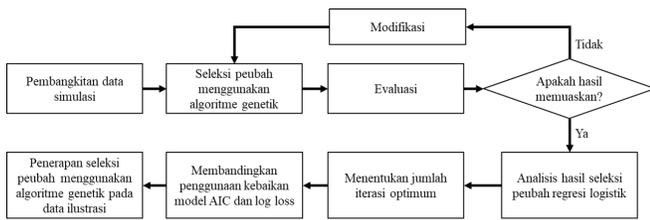
B. Tujuan

Tujuan dari penelitian ini adalah menerapkan algoritma genetik untuk seleksi peubah regresi logistik.

II. METODE

A. Kerangka Pemikiran

Untuk mengatasi permasalahan permodelan data mengandung interaksi, dilakukan upaya seleksi peubah. Seleksi peubah dapat dilakukan dengan banyak cara, salah satunya dengan melibatkan metode optimasi. Dalam penelitian ini, dibahas seleksi peubah menggunakan metode optimasi yaitu algoritme genetik. Kerangka pikiran penelitian ini disajikan pada Gambar 1.



Gambar 1. Kerangka pemikiran penelitian

B. Data Simulasi

Data yang digunakan dalam penelitian ini adalah 100 set data hasil bangkitan untuk setiap ukuran kebaikan. Setiap set data bangkitan terdiri dari 34 peubah bebas numerik dan satu peubah respon biner serta memiliki 500 observasi. Tahapan simulasi data adalah sebagai berikut:

- 1) Penentuan sebaran dan parameter peubah bebas. Peubah bebas adalah 34 peubah saling bebas yang menyebar  $N(0,0.2)$ .
- 2) Penentuan peubah bebas berpengaruh. Dari 34 peubah bebas, ditentukan ada 3 peubah bebas dan 1 interaksi yang berpengaruh signifikan terhadap peubah respon. Peubah-peubah tersebut adalah  $X_4$ ,  $X_{13}$ , dan  $X_{30}$ , sedangkan interaksi yang berpengaruh signifikan adalah  $X_{13} * X_{30}$ . Peubah bebas selain ketiga peubah dan satu interaksi tersebut ditentukan tidak berpengaruh sama sekali terhadap peubah respon.

- 3) Penghitungan nilai  $ln(odds)$  untuk tiap observasi. Perhitungan  $ln(odds)$  menggunakan formula sebagai berikut:

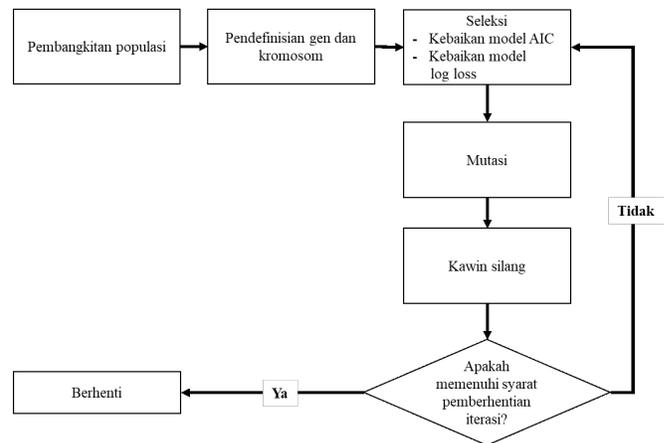
$$ln(odds) = 2 + 4X_4 + 4X_{13} + 4X_{30} + 16X_{13} * X_{30}$$

- 4) Pembangkitan peubah respon. Peubah respon dibangkitkan menyebar binomial dengan besar peluang pada masing-masing observasi adalah:

$$\pi(x) = \frac{1}{1 + e^{-ln(odds)}}$$

C. Seleksi Peubah Regresi Logistik menggunakan Algoritme Genetik

Setiap ukuran kebaikan memiliki 100 set data yang dibangkitkan secara acak. Prosedur penerapan seleksi peubah regresi logistik menggunakan algoritme genetik dilakukan terhadap masing masing set data sebanyak 5 kali. Masing-masing pengulangan akan menghasilkan satu model terbaik sehingga didapatkan 5 model untuk setiap set data. Dari 5 model tersebut, kemudian dipilih 1 model yang terbaik untuk selanjutnya dianalisis. Selanjutnya, tahapan algoritme genetik pada seleksi peubah regresi logistik diilustrasikan pada Gambar 2.



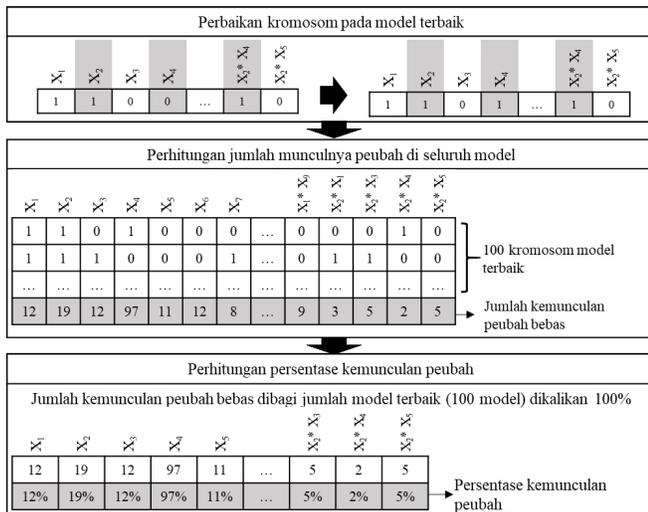
Gambar 2. Diagram alur algoritme genetik

Seleksi peubah regresi logistik menggunakan dua ukuran kebaikan model, yaitu AIC dan log loss, sehingga seluruh proses analisis dilakukan dua kali, dengan masing-masing tahapan adalah sebagai berikut :

- 1) Pendefinisian gen dan kromosom
- 2) Pembangkitan populasi
- 3) Seleksi

- 4) Kawin silang
- 5) Mutasi
- 6) Penghentian iterasi

Setelah seleksi peubah regresi logistik menggunakan algoritme genetik selesai, dilakukan penghitungan persentase kemunculan peubah. Tahapan perhitungan persentase kemunculan peubah disajikan pada Gambar 3.



Gambar 3. Tahapan perhitungan persentase kemunculan peubah

**D. Penentuan Jumlah Ulangan yang Efisien untuk Seleksi Peubah Menggunakan Algoritme Genetik**

Tahapan yang dilakukan untuk menentukan jumlah iterasi optimum pada seleksi peubah menggunakan algoritme genetik adalah sebagai berikut:

- 1) Pembangkitan 1 set data
- 2) Penerapan algoritme genetik sebanyak 100 kali
- 3) Pemilihan model secara acak
- 4) Penghitungan jumlah kuadrat selisih persentase kemunculan dengan formula  $\sum_{i=1}^{59} (p_{i,n} - p_{i,100})^2$
- 5) Penentuan jumlah iterasi optimum

**E. Data Ilustrasi**

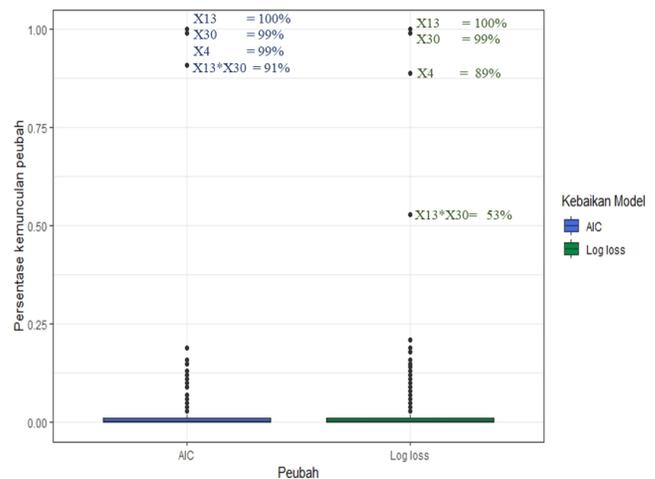
Data ilustrasi yang digunakan adalah data Myopia yang diambil dari buku *Applied Logistic Regression Third Edition* karya Hosmer dan Lemeshow pada tahun 2013. Data ini merupakan bagian dari data dalam penelitian *Orinda Longitudinal Study of Myopia* (OLSM) yang mengkaji

tentang perkembangan dan resiko Myopia pada anak. Data Myopia yang digunakan dalam penelitian ini memiliki 618 observasi yang berasal dari anak-anak yang tidak memiliki permasalahan miopi pada awal penelitian. Data ini memiliki 16 peubah bebas dan 1 peubah respon.

**III. HASIL DAN PEMBAHASAN**

**A. Hasil Seleksi Peubah menggunakan Algoritme Genetik dengan Generasi Pertama Dibangkitkan Acak**

Dari pemrosesan seleksi peubah menggunakan algoritme genetik, didapatkan dua hasil secara umum, yaitu peubah-peubah penyusun model dan nilai kebaikan model terbaik. Kebaikan model yang digunakan sebagai fungsi objektif pada optimasi algoritme genetik ada dua, yaitu AIC dan log loss. Gambar 4 menunjukkan hasil algoritme genetik dengan populasi awal acak.



Gambar 4. Boxplot persentase kemunculan peubah hasil algoritme genetik dengan generasi pertama dibangkitkan acak

Dalam optimasi yang dilakukan, kebaikan model yang digunakan sebagai fungsi objektif adalah AIC dan log loss. Pada pembangkitan data simulasi, peubah bebas yang diatur berpengaruh signifikan adalah faktor utama  $X_4$ ,  $X_{13}$ ,  $X_{30}$ , dan interaksi antara peubah  $X_{13}$  dan  $X_{30}$ . Hasil seleksi peubah pada Gambar 4 menunjukkan keempat peubah tersebut memiliki persentase kemunculan yang tertinggi baik pada kebaikan model AIC maupun log loss. Selain keempat peubah yang diatur berpengaruh,

peubah lain memiliki persentase kemunculan sangat kecil bahkan bernilai nol pada kedua kebaikan model yang digunakan. Meskipun begitu, beberapa diantaranya masih memiliki persentase kemunculan yang relatif besar yaitu mencapai 21% pada log loss, dan 19% pada AIC. Angka ini cukup besar mengingat peubah peubah tersebut tidak diatur berpengaruh signifikan terhadap peubah respon sejak awal.

Tabel I  
RINGKASAN HASIL SELEKSI PEUBAH MENGGUNAKAN ALGORITME GENETIK DENGAN GENERASI PERTAMA DIBANGKITKAN ACAK

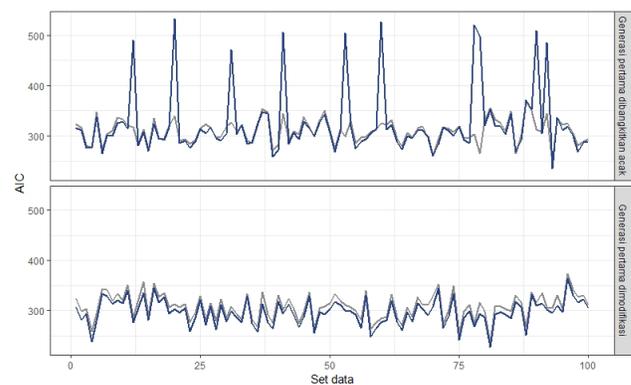
		AIC	Log loss
<i>False Negative</i>	$X_4$	1%	11%
	$X_{13}$	0%	0%
	$X_{30}$	1%	1%
	$X_{13} * X_{30}$	9%	47%
<i>False Positive</i>	Maksimum	19%	21%
	Minimum	0%	0%
	Rata-rata	0.97%	1.11%
	Jumlah peubah	30.63%	33.84%

Terlihat bahwa baik *false negative* ataupun *false positive* bernilai relatif besar. Hal ini terutama terjadi ketika kebaikan model yang digunakan adalah log loss. Besarnya kesalahan menggunakan log loss lebih besar dibandingkan menggunakan AIC. *False negative* hanya terjadi pada empat peubah yang sesungguhnya berpengaruh signifikan, tetapi tidak masuk dalam model terbaik. Nilai *false negative* berkisar antara 0% hingga 47%. Nilai 47% terjadi pada interaksi antara peubah  $X_{13}$  dan  $X_{30}$  ketika menggunakan kebaikan model log loss. Dari seluruh uraian di atas, terlihat bahwa algoritme genetik dengan generasi pertama acak belum mampu membedakan dengan tegas peubah yang berpengaruh terhadap peubah respon. Selain itu, metode ini belum mampu menghasilkan nilai AIC dan log loss yang baik. Jadi, perlu dilakukan modifikasi terhadap tahapan-tahapan dalam algoritme genetik.

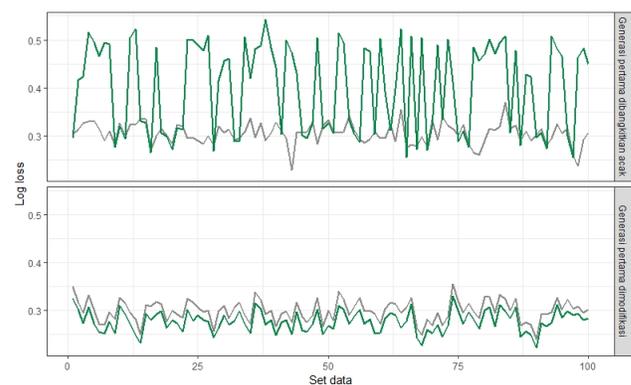
### B. Hasil Seleksi Peubah menggunakan Algoritme Genetik dengan Modifikasi Generasi Pertama

Untuk mengatasi hasil yang kurang baik, dilakukan modifikasi terhadap peluang kemunculan angka 1 di tiap gen. Modifikasi dilakukan pada tahap pembangkitan generasi pertama. Hal

ini menyebabkan besarnya peluang ditentukan berdasarkan nilai mutlak korelasi antara peubah bebas dengan peubah respon. Dalam penelitian ini, korelasi digunakan sebagai indikator, sehingga tidak ada masalah dalam penggunaannya meskipun peubah respon berskala biner. Gen dengan peubah bebas yang memiliki nilai mutlak korelasi yang lebih besar akan memiliki peluang munculnya angka 1 lebih besar pula. Perbedaan nilai peluang ini membuat peubah-peubah yang memiliki nilai mutlak korelasi terlalu kecil, akan memiliki peluang kemunculan angka 1 pada gen padanannya sebesar 0, sehingga peubah yang dianggap tidak memiliki pengaruh sama sekali terhadap peubah respon akan terseleksi di generasi pertama.



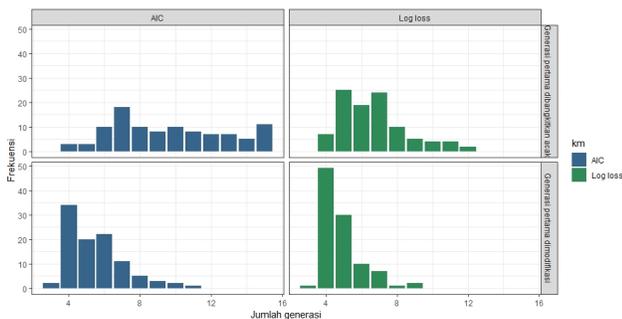
Gambar 5. Perbandingan AIC model sesungguhnya dan AIC hasil seleksi peubah regresi logistik menggunakan algoritme genetik tanpa modifikasi dan termodifikasi



Gambar 6. Perbandingan log loss model sesungguhnya dan log loss hasil seleksi peubah regresi logistik menggunakan algoritme genetik tanpa modifikasi dan termodifikasi

1) *Nilai AIC dan Log Loss Model*: Dari 100 model terbaik yang terbentuk dari algoritme genetik

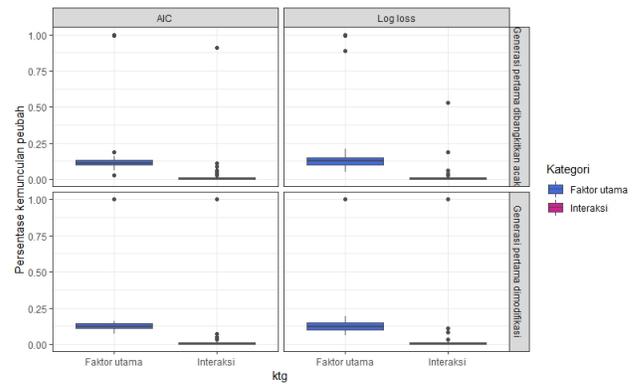
termodifikasi, semuanya memiliki nilai AIC yang lebih baik dari AIC sesungguhnya. Hal ini juga terjadi pada kebaikan model log loss. Model GA standar menghasilkan model yang tidak lebih baik dari model sesungguhnya seperti yang telah dijelaskan pada hasil algoritme genetik dengan generasi pertama dibangkitkan acak. Kesimpulan yang dapat diambil dari uraian di atas adalah, modifikasi pada pembangkitan generasi pertama algoritme genetik mampu meningkatkan kebaikan model regresi logistik yang dihasilkan.



Gambar 7. Diagram batang jumlah generasi yang diperlukan hingga didapatkan model terbaik pada seleksi peubah regresi logistik menggunakan algoritme genetik

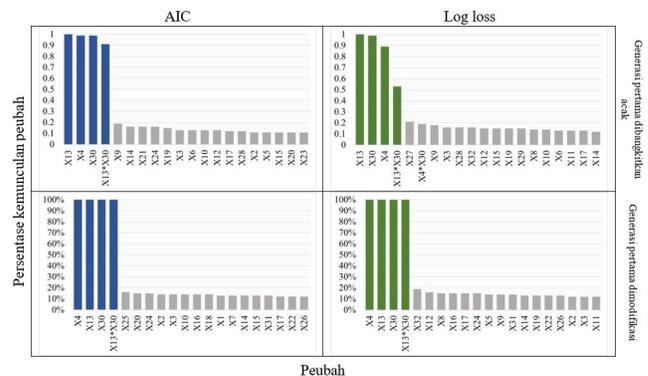
2) *Jumlah Generasi:* Dari Gambar 7 terlihat bahwa pada algoritme genetik dengan generasi pertama dibangkitkan acak, jumlah generasi relatif lebih besar dibanding ketika algoritme genetik telah dimodifikasi. Jumlah generasi yang lebih banyak tentunya memengaruhi waktu komputasi. Semakin banyak generasi, maka waktu komputasi akan semakin lama dan pemrosesan semakin tidak efektif. Sebelum modifikasi dilakukan, jumlah generasi pada kebaikan model AIC mencapai jumlah maksimum yaitu 15. Hal ini bisa jadi mengindikasikan bahwa metode belum mendapatkan model yang terbaik, tetapi harus berhenti karena mencapai jumlah iterasi maksimum. Dari uraian tersebut, dapat disimpulkan bahwa modifikasi meningkatkan efisiensi pemrosesan karena mampu mengurangi jumlah iterasi yang perlu dilakukan baik pada kebaikan model AIC ataupun log loss.

3) *Model Hasil Seleksi Peubah:* Dari Gambar 8 terlihat bahwa ada empat peubah yang memiliki persentase kemunculan tertinggi. Peubah bebas tersebut adalah  $X_4$ ,  $X_{13}$ ,  $X_{30}$  dan interaksi antara  $X_{13}$  dan  $X_{30}$ . Jika mengabaikan faktor interaksi



Gambar 8. Diagram perbandingan AIC model awal dan AIC hasil seleksi peubah regresi logistik menggunakan algoritme genetik

antara peubah bebas  $X_{13}$  dan  $X_{30}$ , maka persentase kemunculan peubah bebas interaksi terbesar adalah 7% pada kebaikan model AIC dan 11% pada kebaikan model log loss. Angka ini masih lebih kecil dibandingkan dengan persentase kemunculan faktor utama, baik pada algoritme genetik termodifikasi maupun tidak. Hal ini mengindikasikan bahwa bagaimanapun, faktor utama memiliki pengaruh yang lebih besar dibandingkan dengan peubah bebas interaksi. Hal ini sesuai dengan prinsip efek hirarki yang menyatakan bahwa pengaruh dari faktor yang derajatnya lebih rendah lebih besar dibanding pengaruh faktor yang derajatnya lebih tinggi (Yuan M, 2007).



Gambar 9. Diagram batang 20 peubah bebas yang memiliki persentase kemunculan tertinggi

4) *Penentuan Peubah yang Berpengaruh terhadap Peubah Respon:* Untuk dapat membedakan antara peubah bebas yang berpengaruh dan tidak berpengaruh pada data riil, hasil algoritme genetik termodifikasi dapat digunakan sebagai

acuan. Peubah bebas yang muncul dalam semua model terbaik yang terbentuk adalah peubah bebas yang berpengaruh terhadap peubah respon. Misal, ulangan yang dilakukan adalah 100 kali, maka peubah bebas harus muncul dalam 100 model yang terbentuk untuk dapat dianggap sebagai peubah bebas yang berpengaruh signifikan terhadap peubah respon.

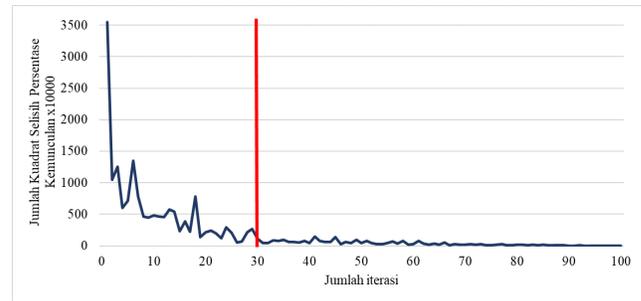
Tabel II  
RINGKASAN HASIL SELEKSI PEUBAH MENGGUNAKAN ALGORITME GENETIK DENGAN MODIFIKASI PADA PEMBANGKITAN GENERASI PERTAMA

		AIC	Log loss
<i>False Negative</i>	$X_4$	0%	0%
	$X_{13}$	0%	0%
	$X_{30}$	0%	0%
	$X_{13} * X_{30}$	0%	0%
<i>False Positive</i>	Maksimum	16%	19%
	Minimum	0%	0%
	Rata-rata	1.02%	1.02%
	Jumlah peubah	37.23%	32.83%

5) *Ringkasan:* Hasil seleksi peubah menggunakan kebaikan model AIC dan log loss memberikan hasil yang relatif sama. Keduanya mampu menghasilkan *false negative* sebesar 0%, dan memiliki rataan *false positive* yang sama. Meskipun begitu, nilai *false positive* maksimum pada AIC lebih kecil dibandingkan dengan log loss dan jumlah peubah tak berpengaruh yang masuk dalam model menggunakan AIC lebih besar pula dibandingkan dengan log loss. Dengan mempertimbangkan hal tersebut, maka AIC lebih dipilih dibanding dengan log loss. Hal ini dikarenakan nilai *false positive* maksimum akan lebih berpengaruh terhadap kesimpulan hasil seleksi peubah dibandingkan dengan jumlah peubah tak berpengaruh yang masuk dalam model. Nilai *false positive* yang tinggi akan membuat suatu peubah yang tidak berpengaruh disimpulkan berpengaruh terhadap peubah respon. Nilai *false positive* yang relatif lebih kecil akan sulit memengaruhi kesimpulan akhir meskipun terjadi pada banyak peubah bebas.

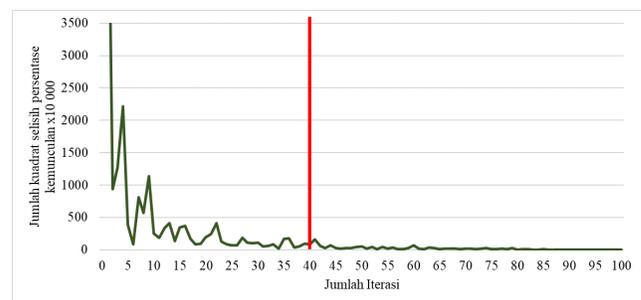
### C. Iterasi Optimum pada Seleksi Peubah Regresi Logistik Menggunakan Algoritme Genetik

Gambar 10 menunjukkan diagram garis antara jumlah kuadrat selisih persentase kemunculan den-



Gambar 10. Diagram jumlah iterasi optimum pada kebaikan model AIC

gan jumlah iterasi. Terlihat bahwa jumlah kuadrat selisih memiliki nilai mendekati 0 ketika jumlah iterasi lebih dari 30 iterasi. Sebagian besar nilai jumlah kuadrat selisih persentase kemunculan bernilai bahkan kurang dari 100 ketika dikalikan dengan 10 000. Hal ini berarti ketika jumlah iterasi lebih dari 30, persentase kemunculan peubah yang dihasilkan tidak jauh berbeda dengan persentase kemunculan peubah yang dihasilkan ketika iterasi 100 kali. Seangkan ketika jumlah iterasi kurang dari 30 kali, jumlah kuadrat selisih persentase terlihat tidak stabil. Nilai tersebut juga naik terus menerus ketika jumlah iterasi terus dikurangi. Oleh sebab itu, untuk mendapatkan efisiensi yang terbaik, iterasi optimum untuk seleksi peubah regresi logistik menggunakan algoritme genetik dengan AIC adalah 30 iterasi.

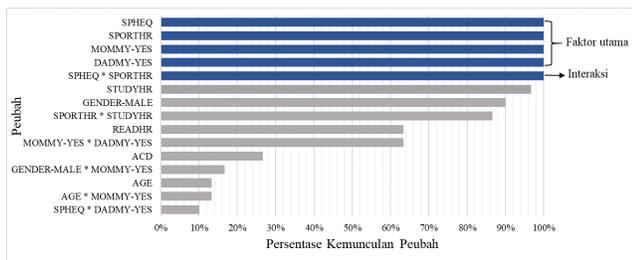


Gambar 11. Diagram jumlah iterasi optimum pada kebaikan model log loss

Garis hijau pada Gambar 11 menggambarkan nilai jumlah kuadrat selisih persentase kemunculan dikalikan dengan 1000 pada masing-masing  $n$  (jumlah iterasi) pada pemrosesan menggunakan kebaikan model log loss. Secara umum, ketika iterasi berjumlah lebih dari 40, jumlah kuadrat selisih persentase kemunculan bernilai kurang dari 100. Nilai

ini relatif kecil, sehingga dapat dikatakan bahwa pada jumlah iterasi lebih dari 40, vektor persentase kemunculan relatif sama dengan ketika iterasi berjumlah 100. Apabila iterasi diturunkan hingga kurang dari 40, jumlah kuadrat selisih persentase kemunculan mulai tidak stabil dan menjadi lebih dari 100. Dari uraian tersebut, maka dapat disimpulkan bahwa jumlah iterasi algoritme genetik optimum pada kebaikan model log loss adalah 40 iterasi.

D. Seleksi Peubah pada Data Ilustrasi



Gambar 12. Diagram batang persentase kemunculan peubah bebas hasil pemrosesan data simulasi

Gambar 12 menyajikan hasil seleksi peubah regresi logistik menggunakan algoritme genetik pada data MYOPIA. Peubah yang berpengaruh terhadap peubah respon adalah peubah yang muncul dalam seluruh model terbaik yang didapatkan atau memiliki persentase kemunculan peubah 100%. Dari semua peubah faktor utama dan interaksi yang ada, terdapat 5 peubah bebas yang berpengaruh terhadap peubah respon. Peubah tersebut adalah SPHEQ (*Spherical equivalent refraction*), SPORTHR (waktu yang dihabiskan anak untuk berolahraga atau bermain di luar rumah di luar jam sekolah), MOMMY-YES (apakah ibu menderita miopi), DADDY-YES (apakah ayah menderita miopi) dan interaksi antara peubah SPHEQ dan SPORTHR.

Dari hasil di atas terlihat bahwa 4 peubah bebas terbukti berpengaruh signifikan terhadap peubah respon pada taraf nyata 0.05. Peubah yang berpengaruh adalah SPHEQ (*textitSpherical equivalent refraction*), MOMMY-YES (apakah ibu menderita miopi), DADDY-YES (apakah ayah menderita miopi) dan interaksi antara peubah SPHEQ dan SPORTHR. Satu peubah yang lain tidak berpengaruh terhadap peubah respon, peubah tersebut adalah SPORTHR (waktu yang dihabiskan anak

Tabel III  
HASIL UJI WALD UNTUK PEUBAH BEBAS YANG TERPILIH

	Estimate	Std. Error	Z value	P-value	
(intercept)	-0.832	0.479	-1.739	0.082	.
SPORTHR	-0.004	0.028	-0.156	0.876	
SPHEQ	-2.419	0.704	-3.434	0.001	***
MOMMY.YES	0.700	0.308	2.274	0.0230	*
DADDY.YES	0.873	0.306	2.848	0.004	**
SPORTHR*	-0.147	0.066	-2.237	0.025	*
SPHEQ					

untuk berolahraga atau bermain di luar rumah di luar jam sekolah).

IV. SIMPULAN DAN SARAN

A. Simpulan

Algoritma genetik dapat digunakan sebagai pendekatan untuk melakukan seleksi peubah regresi logistik. Modifikasi generasi pertama regresi logistik memberikan efek positif terhadap hasil yang diberikan. Modifikasi berupa pemberian peluang kemunculan yang berbeda-beda pada tiap peubah bebas berdasarkan nilai mutlak korelasi terhadap peubah respon. Untuk memilih peubah bebas yang berpengaruh signifikan terhadap peubah respon, dapat dilakukan iterasi 30 menggunakan fungsi objektif AIC. Dari 30 kali iterasi, apabila suatu peubah muncul dalam semua model yang terbentuk, maka peubah tersebut adalah peubah bebas yang berpengaruh signifikan terhadap peubah respon. Dengan menerapkan algoritme genetik pada seleksi peubah set data MYOPIA, disimpulkan bahwa peubah bebas yang berpengaruh signifikan terhadap peubah respon MYOPIC adalah SPHEQ (*textitSpherical equivalent refraction*), SPORTHR (waktu yang dihabiskan anak untuk berolahraga atau bermain di luar rumah di luar jam sekolah), MOMMY-YES (apakah ibu menderita miopi), DADDY-YES (apakah ayah menderita miopi) dan interaksi antara peubah SPHEQ dan SPORTHR. Hasil seleksi peubah regresi logistik menggunakan algoritme genetik telah memenuhi prinsip hereditas dan hirarkhi.

B. Saran

Hasil seleksi peubah belum cukup baik membedakan peubah yang berpengaruh terhadap peubah respon. Modifikasi lebih lanjut dapat ditambahkan

terhadap algoritme genetik. Penelitian selanjutnya dapat melakukan perhitungan *False Discovery Rate* (FDR) yang didapatkan dari metode ini.

#### DAFTAR PUSTAKA

- Tan H, Lin SJ, L. B. S. S. (2005). Comparing exhaustive chaid classification tree and forward stepwise logistic regression (lr) in explaining the prescribing of antidepressants. *Value in Health* 8(3), 1524–4733.
- Wang D, Zhang W, B. A. (2004). Comparison of bayesian model averaging and stepwise methods for model selection in logistic regression. *Statistics in Medicine* 23(22), 34513467.
- Yuan M, Joseph VR, L. Y. (2007). An efficient variable selection approach for analyzing designed experiments. *Technometrics* 49(4), 430–439.