

## Seleksi Peubah dengan Menggunakan Algoritme Genetika pada Data Rancangan Faktorial Pecahan Lewat Jenuh Dua Taraf\*

Ani Safitri<sup>1</sup>, Rahma Anisa<sup>2‡</sup>, and Bagus Sartono<sup>3</sup>

<sup>1,2,3</sup>Department of Statistics, IPB University, Indonesia  
‡corresponding author: [r.rahma.anisa@gmail.com](mailto:r.rahma.anisa@gmail.com)

Copyright © 2020 Ani Safitri, Rahma Anisa, and Bagus Sartono. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

### Abstract

In certain fields, experiments involve many factors and are constrained by costs. Reducing runs is one of the solutions to reduce experiment costs. But that can cause the number of runs to become less than the number of factors. This case of experimental design also is known as a supersaturated design. The important factors in this design are generally estimated by involving variable selection such as forward selection, stepwise regression, and penalized regression. Genetic algorithm is one of the methods that can be used for variable selection, especially for high dimensional data or supersaturated design. This study aims to use a genetic algorithm for variable selection in the supersaturated design and compare the genetic algorithm results with a stepwise regression which is generally used for a simple design. This study also involved fractional factorial design principles. The result showed that the main factors and interactions of the genetic algorithm and stepwise regression were quite different. But the principle was the same because the variables correlated. The genetic algorithm model had a smaller AIC and BIC and all of the main factors and interactions which had chosen were significant on the 0.1%. Therefore genetic algorithm model was chosen although computation time was much longer than stepwise regression.

**Keywords:** genetic algorithms, fractional factorial, supersaturated, stepwise regression, variable selection.

---

\* Received: Jul 2019; Reviewed: Des 2020; Published: Jan 2021

## 1. Pendahuluan

Pada berbagai bidang tertentu seperti farmasi, kromatografi, dan sosial, percobaan dilakukan dengan melibatkan banyak faktor sehingga membutuhkan banyak unit percobaan dan terkendala biaya. Salah satu solusi masalah tersebut yaitu dengan mengurangi unit percobaan. Berkurangnya unit percobaan menyebabkan percobaan hanya dapat menggunakan sebagian perlakuan sehingga rancangan percobaan yang digunakan yaitu rancangan faktorial pecahan.

Rancangan jenuh (*saturated design*) dan rancangan lewat jenuh (*supersaturated design*) merupakan rancangan faktorial pecahan. Pengaruh faktor pada rancangan jenuh masih dapat diduga menggunakan metode kuadrat terkecil (MKT) sedangkan pengaruh faktor pada rancangan lewat jenuh tidak dapat diduga menggunakan MKT. Hal ini karena derajat bebas pada rancangan lewat jenuh tidak cukup untuk menduga pengaruh faktor (Salawu et al., 2015).

Pengaruh faktor pada rancangan lewat jenuh umumnya diduga dengan menggunakan metode yang melibatkan seleksi peubah seperti *forward regression*, regresi bertatar (*stepwise regression*), dan regresi berkendala (*penalized least squares*). Selain metode-metode tersebut yang dapat menyeleksi peubah, algoritme genetika juga dapat digunakan untuk seleksi peubah, khususnya pada data berdimensi tinggi seperti rancangan lewat jenuh.

Paterlini & Minerva (2010) menyimpulkan bahwa pemilihan peubah dengan menggunakan algoritme genetika pada model regresi menghasilkan nilai AIC, BIC, dan SIC lebih kecil dibandingkan regresi bertatar. Penelitian ini juga menggunakan algoritme genetika dan regresi bertatar untuk seleksi peubah namun data yang digunakan berbeda sehingga kesimpulannya juga dapat berbeda dari penelitian Paterlini & Minerva (2010).

Penelitian ini menggunakan algoritme genetika untuk seleksi peubah pada data rancangan faktorial pecahan lewat jenuh dua taraf dan membandingkan hasilnya dengan regresi bertatar yang umumnya digunakan pada rancangan sederhana. Penelitian ini juga melibatkan prinsip-prinsip faktorial pecahan khususnya prinsip *effect heredity* sehingga penelitian ini juga melibatkan interaksi antar pengaruh utamanya. Penelitian ini bertujuan menerapkan algoritme genetika untuk seleksi peubah pada data rancangan faktorial pecahan lewat jenuh dua taraf dan membandingkan hasilnya dengan regresi bertatar.

## 2. Metodologi

### 2.1 Bahan dan Data

Penelitian ini menggunakan dua data sekunder, yaitu:

#### a. Data Percobaan Pembuatan Amida Lemak Sulfat

Data percobaan kimia yang melibatkan minyak zaitun sebagai bahan awal pembuatan amida lemak sulfat. Amida lemak sulfat digunakan sebagai dispersan sabun jeruk nipis. Data ini terdiri atas 18 unit percobaan dan 31 faktor. Faktor-faktornya yaitu  $X_1$ - $X_{31}$  dan peubah responnya merupakan persentase hasil reaksi pembuatan amida lemak sulfat. Percobaan kimia ini berusaha memaksimalkan hasil reaksi dan mengurangi sisa diethanolamine dan sulfat yang berbahaya bagi kulit (Rais et al., 2009).

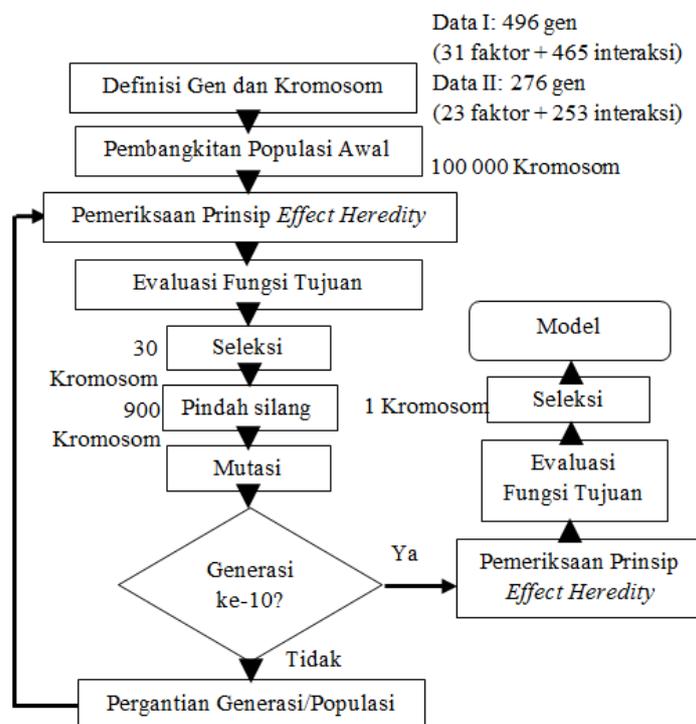
**b. Data Percobaan Pembuatan Karet**

Penelitian ini menggunakan data penelitian Abraham *et al.* (1999) yang juga digunakan oleh Dejaegher *et al.* (2007) yang ingin mengetahui faktor-faktor yang paling berpengaruh terhadap usia karet pada proses pembuatan karet. Data ini terdiri dari 23 faktor dan 14 unit percobaan. Peubah responnya merupakan usia karet. Data penelitian Abraham *et al.* (1999) ini merupakan modifikasi data penelitian Lin (1993) dengan menghilangkan faktor ke-16 karena faktor ke-13 dan faktor ke-16 serupa nilainya. Penulisan nama faktornya tidak dibedakan dari data sebelumnya sehingga nama faktor-faktornya yaitu  $X_1-X_{15}$  dan  $X_{17}-X_{24}$ . Data yang digunakan Lin (1993) merupakan data rancangan lewat jenuh yang dirancang dari rancangan sederhana pada penelitian Williams (1968). Data ini terdiri dari 24 faktor dan 28 unit percobaan. Nama interaksi di kedua yaitu  $X_iX_j$ . Pada data pertama  $i, j = 1, 2, \dots, 31$  dan  $i \neq j$ . Pada data kedua  $i, j = 1, 2, \dots, 24$  dan  $i \neq j \neq 16$ .

**2.2 Prosedur Analisis Data**

Analisis data dilakukan dengan menggunakan *Software R 3.5.1*, Prosedur analisis data dilakukan dengan ketentuan sebagai berikut:

1. Regresi bertatar menggunakan kriteria kebaikan AIC yang tersedia di *package RcmdrMisc* yang diulang sebanyak sepuluh kali lalu dipilih satu model terbaik dengan nilai AIC dan BIC terkecil. Regresi bertatar juga diatur agar interaksi yang masuk ke dalam model juga mengikutsertakan pengaruh utama yang membentuk interaksi tersebut.
2. Tahapan algoritme genetika ditunjukkan pada Gambar 1 dengan ketentuan sebagai berikut:



Gambar 1: Tahapan algoritme genetika.

a. Definisi gen dan kromosom

Setiap kromosom memiliki gen sebanyak pengaruh utama dan interaksi. Interaksi yang digunakan merupakan interaksi antar pengaruh utama. Nilai interaksi merupakan perkalian nilai antar dua pengaruh utama sehingga pengaruh utama dan interaksi bernilai -1 atau +1.

Gen bernilai 1 artinya pengaruh utama atau interaksi yang berpadanan dengan gen tersebut masuk ke dalam model dugaan sedangkan gen bernilai 0 berarti sebaliknya. Data pertama terdiri dari 31 pengaruh utama dan 465 interaksi ( $C_2^{31}$ ) sehingga kromosomnya memiliki 496 gen. Data kedua terdiri dari 23 pengaruh utama dan 253 interaksi ( $C_2^{23}$ ) sehingga kromosomnya memiliki 276 gen.

b. Pembangkitan populasi awal

Populasi awal terdiri dari 100 000 kromosom sehingga matriks populasi data pertama berukuran  $100\ 000 \times 496$  yang berisi 49 600 000 gen dan matriks populasi data kedua berukuran  $100\ 000 \times 276$  yang berisi 27 600 000 gen.

Populasi awal dibangkitkan dari bilangan acak Bernoulli. Pembangkitan bilangan acak dilakukan di setiap kolom matriks populasi dengan peluang yang berbeda. Peluangnya bergantung pada korelasi pengaruh utama dan interaksi yang bersesuaian dengan kolom-kolom tersebut terhadap respon. Pengaruh utama dan interaksi yang memiliki nilai korelasi tinggi terhadap respon memiliki peluang bilangan acak Bernoulli lebih tinggi. Apabila nilai peluang-peluang tersebut dijumlahkan, nilainya mendekati angka 0.032 untuk data pertama dan angka 0.043 untuk data kedua. Pada data percobaan pembuatan amida lemak sulfat peubah bebas yang ingin dipilih sebanyak 16 sedangkan pada data percobaan pembuatan karet peubah bebas yang ingin dipilih sebanyak 12. Peubah bebas yang ingin dipilih merupakan perpaduan pengaruh utama dan interaksi. Angka 0.032 dan 0.043 merupakan proporsi peubah bebas yang ingin dipilih dalam model terhadap banyaknya gen atau banyaknya pengaruh utama dan interaksi yang ada.

c. Pemeriksaan prinsip *effect heredity*

Pada data pertama terdapat 31 faktor maka gen ke-1 sampai gen ke-31 mewakili pengaruh utama dan gen ke-32 sampai gen ke-496 mewakili interaksi dua pengaruh utama. Gen ke-32 mewakili interaksi pengaruh utama ke-1 dan ke-2, gen ke-33 mewakili interaksi pengaruh utama ke-1 dan ke-3, dan begitu seterusnya. Pada data kedua terdapat 23 faktor maka gen ke-1 sampai gen ke-23 mewakili pengaruh utama dan gen ke-24 sampai gen ke-276 mewakili interaksi dua pengaruh utama. Gen ke-24 mewakili interaksi pengaruh utama ke-1 dan ke-2, gen ke-25 mewakili interaksi pengaruh utama ke-1 dan ke-3, dan begitu seterusnya.

Algoritme genetika juga diatur agar interaksi yang terpilih juga mengikutsertakan pengaruh utama yang membentuk interaksi tersebut masuk ke dalam model. Misalnya pada data pertama, gen ke-32 bernilai 1, gen ke-1 bernilai 1, dan gen ke-2 bernilai 0, maka gen ke-2 perlu diubah nilainya menjadi 1. Hal ini karena gen ke-32 merupakan gen yang mewakili interaksi bagi pengaruh utama ke-1 dan pengaruh utama ke-2. Pada tahapan ini ada dua gugus kromosom yaitu gugus kromosom yang sudah diatur agar mengikuti aturan tersebut dan gugus kromosom sebelumnya yang belum diatur mengikuti aturan tersebut. Hal ini karena tahapan selain tahapan c, d, dan e, gugus kromosom yang digunakan yaitu gugus kromosom yang tidak mengikuti aturan tersebut. Hal ini bertujuan agar algoritme genetika tidak mudah konvergen terhadap solusi tertentu.

d. Evaluasi fungsi tujuan

Fungsi tujuan yang digunakan yaitu meminimumkan nilai AIC. Fungsi tujuan dihitung sebanyak dua kali karena terdapat dua gugus kromosom hasil tahapan poin c. AIC bernilai 10 000 000 untuk kromosom yang memiliki gen bernilai 1 sebanyak lebih dari 16 gen untuk data pertama dan 12 gen untuk data kedua, sedangkan kurang dari itu nilai AIC sesuai nilai aslinya.

e. Seleksi

Seleksi menggunakan teknik *rank selection* yang terlebih dahulu mengurutkan secara *ascending* nilai AIC untuk gugus kromosom yang sudah mengikuti aturan poin c lalu mengurutkan secara *ascending* nilai AIC untuk gugus kromosom yang sebelumnya tidak mengikuti aturan poin c. Kromosom yang terpilih merupakan kromosom yang berada di urutan ke-1 sampai 30. Ukuran matriks populasi data pertama menjadi  $30 \times 496$ , sedangkan data kedua  $30 \times 276$ .

f. Pindah silang

Pindah silang menggunakan teknik *Single-point crossover* dengan menggunakan induk yang berasal dari hasil tahapan seleksi. Induk yang terpilih ada sebanyak 30 kromosom sehingga individu baru yang terbentuk sebanyak 870 kromosom ( $2 \times C_2^{30}$ ). Ukuran matriks populasi data pertama menjadi  $900 \times 496$ , sedangkan data kedua  $900 \times 276$ .

g. Mutasi

Ukuran matriks mutasi sama dengan ukuran matriks populasi. Nilai sel pada matriks mutasi dibangkitkan dari bilangan acak Bernoulli dengan peluang sebesar  $4.49 \times 10^{-7}$  untuk data pertama dan  $6.02 \times 10^{-7}$  untuk data kedua. Sel matriks mutasi data pertama sebanyak 446 400 sel ( $900 \times 496$ ), sedangkan matriks mutasi data kedua sebanyak 248 400 sel ( $900 \times 276$ ). Sel-sel matriks mutasi berpadanan dengan gen-gen matriks populasi sehingga sel yang bernilai satu pada matriks mutasi akan mengubah nilai gen yang berpadanan pada matriks populasi.

h. Pergantian generasi atau populasi

Setelah generasi mengalami mutasi, terbentuklah generasi baru. Selanjutnya generasi baru tersebut mengalami tahapan di poin c sampai poin g yang

disebut sebagai iterasi. Iterasi terus berlangsung sampai generasi ke-10 mengalami mutasi.

i. Tahapan setelah generasi ke-10 mengalami mutasi

Apabila generasi ke-10 sudah mengalami mutasi, proses iterasi dihentikan lalu dipilihlah satu kromosom terbaik dari generasi tersebut dengan cara seperti tahapan poin c sampai poin g.

j. Pengulangan algoritme genetika

Keseluruhan tahapan poin b sampai i merupakan satu kesatuan algoritme genetika. Pada penelitian ini algoritme genetika diulang sebanyak sepuluh kali sehingga terdapat sepuluh model algoritme genetika. Selanjutnya, dipilihlah model terbaik yang memiliki nilai AIC dan BIC terkecil dari sepuluh model tersebut.

3. Model terbaik yang dihasilkan regresi bertatar dan algoritme genetika dibandingkan dengan cara membandingkan nilai AIC dan BIC yang dijadikan sebagai prioritas utama pemilihan model, membandingkan pengaruh utama dan interaksi yang dijadikan sebagai prioritas kedua pemilihan model, dan membandingkan waktu komputasi kedua model yang dijadikan sebagai prioritas terakhir pemilihan model. Pengaruh utama dan interaksi diduga menggunakan MKT dan diuji signifikansinya menggunakan uji t pada taraf nyata maksimal 5%. Hipotesis pada uji t yaitu:

H0 :  $\beta_i = 0$  pengaruh utama atau interaksi tidak berpengaruh nyata

H1 :  $\beta_i \neq 0$  pengaruh utama atau interaksi berpengaruh nyata.

$$t_{hitung} = \frac{b_i - \beta_i}{SE(b_i)}$$

Nilai  $b_i$  merupakan dugaan parameter regresi ke-i,  $\beta_i$  merupakan parameter regresi ke-i berdasarkan hipotesis, dan  $SE(b_i)$  merupakan simpangan baku dugaan parameter regresi ke-i. Hipotesis nol ditolak apabila  $|t_{hitung}| \geq t_{tabel}$  atau  $p\text{-value} \leq \alpha$ . Nilai  $t_{tabel}$  dapat dilihat pada tabel  $t$  dengan memanfaatkan informasi  $\alpha$  dan derajat bebas ( $df = n - k - 1$ ) dengan  $k$  merupakan banyaknya peubah bebas pada model (Draper & Smith, 1992).

### 3. Hasil dan Pembahasan

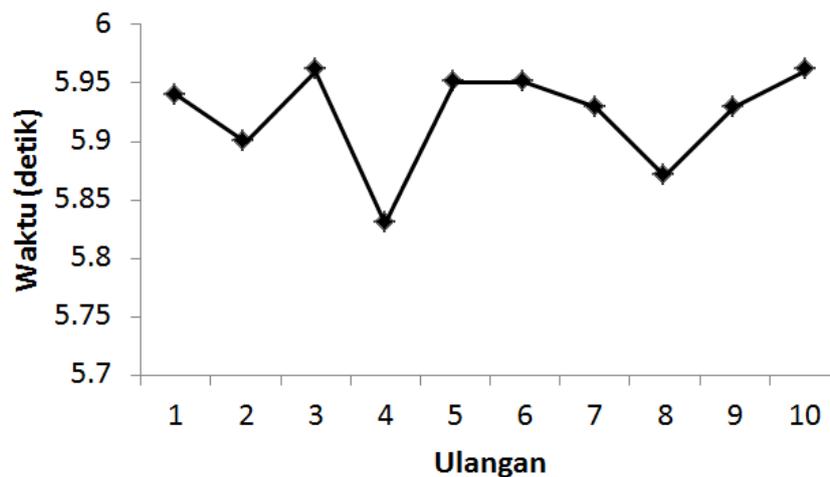
#### 3.1 Analisis Data Percobaan Pembuatan Amida Lemak Sulfat

##### Hasil Seleksi Peubah oleh Rais et al. (2009) (Model L1)

Rais et al. (2009) menggabungkan regresi gulud, regresi bertatar, dan *all-subset regression* untuk menyeleksi faktor dan menduga pengaruh faktor data ini. Model yang dihasilkan didefinisikan oleh penulis sebagai model L1. Faktor yang terkandung didalam model ini yaitu  $X_{13}$ ,  $X_{18}$ ,  $X_{19}$ ,  $X_{20}$ ,  $X_{24}$ ,  $X_{27}$ ,  $X_{28}$ ,  $X_{29}$ , dan  $X_{30}$ . Faktor  $X_{13}$ ,  $X_{20}$ , dan  $X_{30}$  berpengaruh nyata pada taraf nyata 1%, sedangkan faktor-faktor lainnya berpengaruh nyata pada taraf nyata 0.1%.

### Hasil Seleksi Peubah dengan Regresi Bertatar (Model RB1)

Gambar 2 menunjukkan bahwa waktu komputasi regresi bertatar di setiap ulangan cenderung sama. Rata-rata waktu komputasi untuk sepuluh ulangan regresi bertatar yaitu 5.922 detik. Setiap ulangan regresi bertatar menghasilkan model yang sama pada data ini. Model tersebut didefinisikan oleh penulis sebagai model RB1. Pengaruh utama dan interaksi yang terkandung didalam model ini yaitu  $X_2$ ,  $X_4$ ,  $X_{16}$ ,  $X_{18}$ ,  $X_{19}$ ,  $X_{21}$ ,  $X_{24}$ ,  $X_{27}$ ,  $X_{28}$ ,  $X_{29}$ ,  $X_2X_{28}$ ,  $X_4X_{16}$ ,  $X_4X_{29}$ ,  $X_{16}X_{29}$ ,  $X_{18}X_{21}$ , dan  $X_{24}X_{28}$ . Semua pengaruh utama dan interaksi tersebut berpengaruh nyata pada taraf nyata 0.1%. Model RB1 mengikuti prinsip *strong heredity* karena interaksi berpengaruh nyata ketika semua pengaruh utama yang membentuk interaksi tersebut berpengaruh nyata.



Gambar 2: Waktu komputasi regresi bertatar untuk data minyak zaitun.

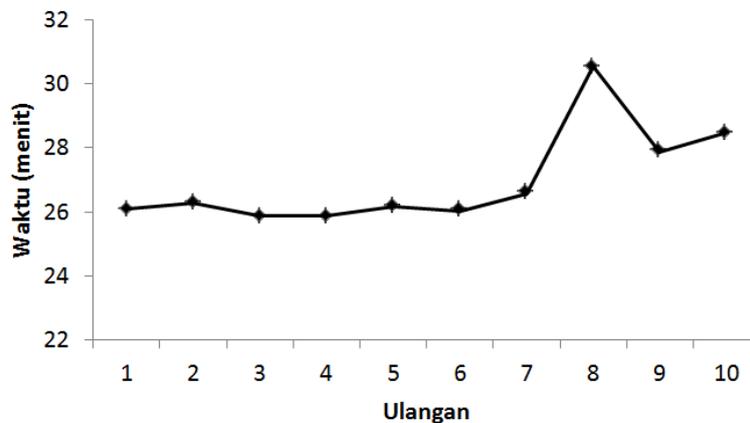
Terdapat perubahan signifikansi pengaruh utama dan interaksi apabila salah satu interaksi dikeluarkan dari model RB1. Interaksi  $X_4X_{29}$  dan  $X_{24}X_{28}$  merupakan bagian terpenting dari model RB1 karena semua pengaruh utama dan interaksi menjadi tidak berpengaruh apabila salah satu dari interaksi tersebut dikeluarkan dari model RB1.

Model RB1 terdiri dari sepuluh pengaruh utama dan enam interaksi. Dua pengaruh utama diantaranya tidak memiliki interaksi yaitu  $X_{27}$  dan  $X_{19}$ . Faktor  $X_{16}$  bergantung pada interaksi  $X_4X_{16}$ , faktor  $X_4$  dan  $X_{29}$  bergantung pada interaksi  $X_4X_{29}$ , dan faktor  $X_{24}$  dan  $X_{28}$  bergantung pada interaksi  $X_{24}X_{28}$ . Ketergantungan tersebut bermakna bahwa suatu pengaruh utama menjadi tidak berpengaruh setelah interaksi yang melibatkan pengaruh utama tersebut dikeluarkan dari model RB1.

### Hasil Seleksi Peubah dengan Algoritme Genetika (Model GA1)

Gambar 3 menunjukkan bahwa waktu komputasi algoritme genetika di setiap ulangan berbeda. Rata-rata waktu komputasi untuk sepuluh ulangan algoritme genetika yaitu 26.98 menit. Terdapat model yang sama dan berbeda di setiap ulangan algoritme genetika. Model yang dipilih yaitu model algoritme genetika yang memiliki nilai AIC dan BIC terkecil dan didefinisikan oleh penulis sebagai model GA1. Pengaruh utama dan interaksi yang terkandung didalam model ini yaitu  $X_1$ ,  $X_3$ ,  $X_6$ ,  $X_8$ ,  $X_9$ ,  $X_{10}$ ,  $X_{13}$ ,  $X_{14}$ ,  $X_{17}$ ,  $X_{23}$ ,  $X_1X_9$ ,  $X_3X_9$ ,  $X_6X_{14}$ ,  $X_8X_{10}$ ,  $X_{10}X_{17}$ , dan  $X_{13}X_{23}$ . Semua pengaruh utama dan interaksi tersebut berpengaruh nyata pada taraf nyata 0.1%. Model GA1 mengikuti prinsip

*strong heredity* karena interaksi berpengaruh nyata ketika semua pengaruh utama yang membentuk interaksi tersebut berpengaruh nyata.



Gambar 3: Waktu komputasi algoritme genetika untuk data minyak zaitun.

Terdapat perubahan signifikansi pengaruh utama dan interaksi apabila salah satu interaksi dikeluarkan dari model GA1. Interaksi  $X_1X_9$  dan  $X_{13}X_{23}$  merupakan bagian terpenting dari model GA1 karena semua pengaruh utama dan interaksi menjadi tidak berpengaruh apabila salah satu dari interaksi tersebut dikeluarkan dari model GA1.

Model GA1 terdiri dari sepuluh pengaruh utama dan enam interaksi. Semua pengaruh utama pada model GA1 memiliki interaksi. Semua pengaruh utama bergantung pada interaksi yang melibatkan pengaruh utama tersebut kecuali faktor  $X_{14}$  yang tidak bergantung pada interaksi yang melibatkannya dan faktor  $X_9$  yang hanya bergantung pada interaksi  $X_1X_9$ . Ketergantungan tersebut bermakna bahwa suatu pengaruh utama menjadi tidak berpengaruh setelah interaksi yang melibatkan pengaruh utama tersebut dikeluarkan dari model GA1.

### Perbandingan Hasil Seleksi Peubah Ketiga Pendekatan

Pada model L1 dan RB1 terdapat enam pengaruh utama yang sama. Pada model L1 dan GA1 terdapat satu pengaruh utama yang sama. Pada model RB1 dan GA1 tidak ada pengaruh utama atau interaksi yang sama. Model RB1 lebih banyak mengandung pengaruh utama yang sama dengan model L1. Hal ini dapat terjadi karena Rais *et al.* (2009) juga menggunakan regresi bertatar di salah satu tahapan seleksi faktor pada data ini.

Sebanyak tiga pengaruh utama yang ada di model L1 tidak ada di model RB1 yaitu  $X_{13}$ ,  $X_{20}$ , dan  $X_{30}$ . Sebanyak satu pengaruh utama yang ada di model L1 tidak ada di model GA1 yaitu selain faktor  $X_{13}$ . Semua pengaruh utama dan interaksi pada model RB1 tidak ada di model GA1 begitupula sebaliknya. Tidak adanya pengaruh utama atau interaksi di suatu model bisa saja perannya digantikan oleh pengaruh utama atau interaksi di dalam model tersebut.

Tabel 1 menunjukkan terdapat pengaruh utama dan interaksi pada model GA1 dan L1 yang saling berkorelasi sehingga perannya dapat saling menggantikan. Contohnya faktor  $X_{19}$  pada model L1 yang perannya digantikan oleh interaksi  $X_1X_9$  pada model GA1.

Tabel 1: Korelasi pengaruh utama dan interaksi model GA1 dan L1.

Model algoritme genetika (GA1)		Model yang dihasilkan oleh Rais <i>et al.</i> (2009) (L1)		
		X <sub>19</sub>	X <sub>24</sub>	X <sub>29</sub>
X <sub>1</sub> X <sub>9</sub>	koefisien korelasi	-0.671	-	-
	<i>p-value</i>	0.002	-	-
X <sub>3</sub> X <sub>9</sub>	koefisien korelasi	-	-0.671	-0.471
	<i>p-value</i>	-	0.002	0.048
X <sub>13</sub> X <sub>23</sub>	koefisien korelasi	-	-0.671	-
	<i>p-value</i>	-	0.002	-

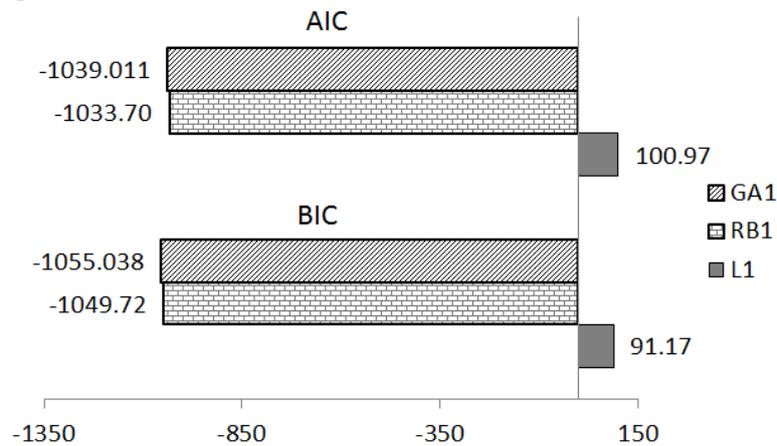
Tabel 2 menunjukkan terdapat pengaruh utama dan interaksi pada model RB1 dan GA1 yang saling berkorelasi sehingga perannya dapat saling menggantikan. Contohnya interaksi X<sub>18</sub>X<sub>21</sub> pada model RB1 yang perannya digantikan oleh pengaruh utama X<sub>8</sub> dan X<sub>23</sub> pada model GA1. Korelasi yang kuat tersebut mengindikasikan adanya pembauran sebagian antara pengaruh utama dan interaksi. Hal ini karena terdapat pengaruh utama yang berkorelasi kuat dengan interaksi namun korelasinya tidak penuh artinya tidak bernilai -1 atau +1.

Tabel 2: Korelasi faktor dan interaksi model RB1 dan GA1.

Model regresi bertatar (RB1)		Model algoritme genetika (GA1)						
		X <sub>8</sub>	X <sub>23</sub>	X <sub>1</sub> X <sub>9</sub>	X <sub>3</sub> X <sub>9</sub>	X <sub>8</sub> X <sub>10</sub>	X <sub>10</sub> X <sub>17</sub>	X <sub>13</sub> X <sub>23</sub>
X <sub>4</sub>	koefisien korelasi	-	-	-	-	-0.671	-	-
	<i>p-value</i>	-	-	-	-	0.002	-	-
X <sub>19</sub>	koefisien korelasi	-	-	-0.671	-	-	-	-
	<i>p-value</i>	-	-	0.002	-	-	-	-
X <sub>24</sub>	koefisien korelasi	-	-	-	-0.471	-	-	-0.671
	<i>p-value</i>	-	-	-	0.048	-	-	0.002
X <sub>29</sub>	koefisien korelasi	-	-	-	-0.471	-	-	-
	<i>p-value</i>	-	-	-	0.048	-	-	-
X <sub>4</sub> X <sub>16</sub>	koefisien korelasi	-	-	-	-	-	0.575	0.55
	<i>p-value</i>	-	-	-	-	-	0.013	0.018
X <sub>16</sub> X <sub>29</sub>	koefisien korelasi	-	-	0.55	-	-	-	-
	<i>p-value</i>	-	-	0.018	-	-	-	-
X <sub>18</sub> X <sub>21</sub>	koefisien korelasi	-0.707	-0.707	-	-	-	-	-
	<i>p-value</i>	0.001	0.001	-	-	-	-	-
X <sub>24</sub> X <sub>28</sub>	koefisien korelasi	-	-	0.553	-	-	-	-
	<i>p-value</i>	-	-	0.017	-	-	-	-

Pembauran selalu terjadi pada rancangan lewat jenuh baik pembauran penuh maupun pembauran sebagian. Pembauran tersebut menyebabkan ketidakbebasan pendugaan pengaruh utama dan interaksi. Pada model RB1 dan GA1 terdapat interaksi yang menyebabkan sebagian besar pengaruh utama menjadi tidak berpengaruh nyata dan terdapat perubahan signifikansi pengaruh utama setelah interaksi yang melibatkan pengaruh utama tersebut dikeluarkan dari model. Hal ini

juga menunjukkan adanya pembauran antara pengaruh utama dan interaksi dua pengaruh utama karena pengaruhnya berubah atau terjadi ketidakbebasan pendugaan pengaruh.



Gambar 4: Nilai AIC dan BIC model L1, RB1, dan GA1.

Gambar 4 menunjukkan bahwa model GA1 memiliki nilai AIC dan BIC paling kecil dan model RB1 memiliki nilai AIC dan BIC lebih kecil dibandingkan model L1. Dengan demikian model GA1 paling baik dan model RB1 lebih baik dibandingkan model L1.

Model hasil algoritme genetika memiliki nilai AIC dan BIC terkecil sebagai prioritas utama pemilihan model terbaik. Semua pengaruh utama dan interaksi berpengaruh nyata pada taraf nyata 0.1% baik model hasil algoritme genetika maupun model hasil regresi bertatar sedangkan model hasil penelitian sebelumnya yaitu penelitian Rais *et al.* (2009) hanya enam pengaruh utama yang berpengaruh nyata pada taraf nyata 0.1% dan tiga pengaruh utama lain berpengaruh nyata pada taraf nyata 1%. Signifikansi pengaruh utama dan interaksi tersebut merupakan prioritas kedua pemilihan model terbaik. Model dengan waktu komputasi tercepat sebagai prioritas terakhir pemilihan model terbaik yaitu model hasil regresi bertatar dengan rata-rata waktu komputasi 5.922 detik, sedangkan rata-rata komputasi model hasil algoritme genetika selama 26.98 menit. Dengan demikian urutan model terbaik penelitian ini yaitu urutan pertama model hasil algoritme genetika (model GA1), urutan kedua model hasil regresi bertatar (model RB1), dan urutan ketiga model hasil penelitian Rais *et al.* (2009) (model L1).

### 3.2 Analisis Data Percobaan Pembuatan Karet

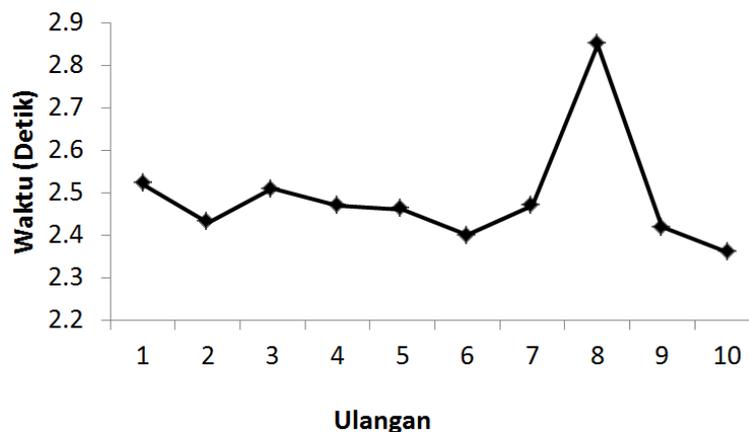
#### Hasil Seleksi Peubah oleh Dejaegher *et al.* (2007) (Model L2)

Dejaegher *et al.* (2007) menduga pengaruh faktor data karet menggunakan metode FEAR (*Fixing Effects and Adding Rows*). Metode ini memungkinkan semua faktor dapat diduga pengaruhnya. Cara kerja metode ini yaitu menambah dugaan observasi dengan menggunakan algoritme tertentu sehingga banyaknya baris data menjadi lebih banyak dua satuan dibandingkan faktor yang digunakan.

Faktor-faktor yang memiliki nilai mutlak koefisien tertinggi urutan ke 1—12 yaitu  $X_2$ ,  $X_6$ ,  $X_8$ ,  $X_9$ ,  $X_{11}$ ,  $X_{12}$ ,  $X_{13}$ ,  $X_{15}$ ,  $X_{17}$ ,  $X_{19}$ ,  $X_{20}$ , dan  $X_{23}$ . Dejaegher *et al.* (2007) menyatakan bahwa hanya faktor  $X_{15}$  yang berpengaruh nyata.

### Hasil Seleksi Peubah dengan Regresi Bertatar (Model RB2)

Gambar 5 menunjukkan bahwa waktu komputasi regresi bertatar di setiap ulangan cenderung sama. Rata-rata waktu komputasi untuk sepuluh ulangan regresi bertatar yaitu 2.489 detik. Setiap ulangan regresi bertatar menghasilkan model yang sama pada data ini. Model tersebut didefinisikan oleh penulis sebagai model RB2. Pengaruh utama dan interaksi yang terkandung didalam model ini yaitu  $X_3$ ,  $X_4$ ,  $X_9$ ,  $X_{10}$ ,  $X_{12}$ ,  $X_{15}$ ,  $X_{20}$ ,  $X_3X_{20}$ ,  $X_4X_{12}$ ,  $X_4X_{15}$ ,  $X_{10}X_{15}$ , dan  $X_{12}X_{15}$ . Semua pengaruh utama dan interaksi tersebut berpengaruh nyata pada taraf nyata 0.1%. Model RB2 mengikuti prinsip *strong heredity* karena interaksi berpengaruh nyata ketika semua pengaruh utama yang membentuk interaksi tersebut berpengaruh nyata.



Gambar 5: Waktu komputasi regresi bertatar untuk data karet.

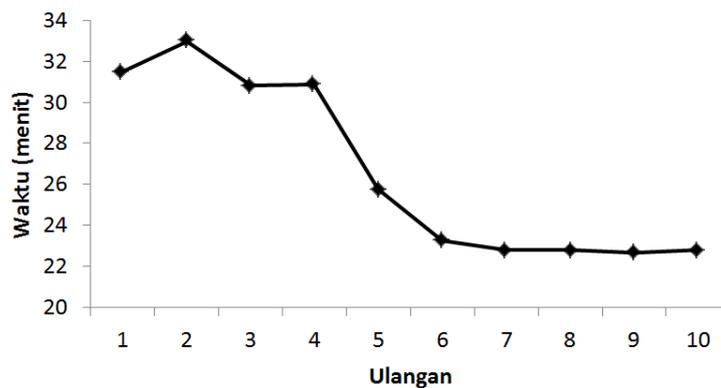
Terdapat perubahan signifikansi pengaruh utama dan interaksi apabila salah satu interaksi dikeluarkan dari model RB2. Interaksi  $X_4X_{12}$  merupakan bagian terpenting dari model RB2 karena hanya faktor  $X_{15}$  yang masih berpengaruh nyata apabila interaksi tersebut dikeluarkan dari model RB2. Faktor  $X_{15}$  merupakan pengaruh utama yang tetap berpengaruh walaupun berbagai interaksi dikeluarkan dari model RB2.

Model RB2 terdiri atas tujuh pengaruh utama dan lima interaksi. Satu pengaruh utama diantaranya tidak memiliki interaksi yaitu faktor  $X_9$ . Faktor  $X_3$  dan  $X_{10}$  bergantung pada interaksi yang melibatkan faktor-faktor tersebut, dan faktor  $X_4$  dan  $X_{12}$  hanya bergantung pada interaksi  $X_4X_{12}$ . Ketergantungan tersebut bermakna bahwa suatu faktor menjadi tidak berpengaruh setelah interaksi yang melibatkan faktor tersebut dikeluarkan dari model RB2.

### Hasil Seleksi Peubah dengan Algoritme Genetika (Model GA2)

Gambar 6 menunjukkan bahwa waktu komputasi algoritme genetika di setiap ulangan berbeda. Rata-rata waktu komputasi untuk sepuluh ulangan algoritme genetika yaitu 26.26 menit. Terdapat model yang sama dan berbeda di setiap ulangan algoritme genetika. Model yang dipilih yaitu model algoritme genetika yang memiliki nilai AIC dan BIC terkecil dan didefinisikan oleh penulis sebagai model GA2. Pengaruh utama dan interaksi yang terkandung didalam model ini yaitu  $X_3$ ,  $X_5$ ,  $X_7$ ,  $X_{10}$ ,  $X_{14}$ ,  $X_{15}$ ,  $X_{20}$ ,  $X_{24}$ ,  $X_3X_5$ ,  $X_3X_{24}$ ,  $X_7X_{20}$ , dan  $X_{10}X_{14}$ . Semua pengaruh utama dan interaksi tersebut berpengaruh nyata pada taraf nyata 0.1%. Model GA2 mengikuti prinsip *strong*

*heredity* karena interaksi berpengaruh nyata ketika semua pengaruh utama yang membentuk interaksi tersebut berpengaruh nyata.



Gambar 6: Waktu komputasi algoritme genetika untuk data karet.

Terdapat perubahan signifikansi pengaruh utama dan interaksi apabila salah satu interaksi dikeluarkan dari model GA2. Interaksi  $X_7X_{20}$  dan  $X_{10}X_{14}$  merupakan bagian terpenting dari model GA2 karena semua pengaruh utama dan interaksi menjadi tidak berpengaruh apabila salah satu dari interaksi tersebut dikeluarkan dari model GA2.

Model GA2 terdiri atas delapan pengaruh utama dan empat interaksi. Satu pengaruh utama diantaranya tidak memiliki interaksi yaitu faktor  $X_{15}$ . Semua pengaruh utama yang memiliki interaksi bergantung pada interaksi yang melibatkan pengaruh utama tersebut. Ketergantungan tersebut bermakna bahwa suatu pengaruh utama menjadi tidak berpengaruh setelah interaksi yang melibatkan pengaruh utama tersebut dikeluarkan dari model GA2.

### Perbandingan Hasil Seleksi Peubah Ketiga Pendekatan

Pada model RB2 dan GA2 terdapat empat faktor yang sama yaitu  $X_3$ ,  $X_{10}$ ,  $X_{15}$ , dan  $X_{20}$ . Keempat faktor tersebut berpengaruh nyata pada taraf nyata 0.1% di kedua model. Faktor  $X_{15}$  juga dinyatakan berpengaruh nyata oleh Dejaegher *et al.* (2007).

Sebanyak tiga pengaruh utama dan lima interaksi yang ada di model di model RB2 tidak ada di model GA2 yaitu  $X_4$ ,  $X_9$ ,  $X_{12}$ ,  $X_3X_{20}$ ,  $X_4X_{12}$ ,  $X_4X_{15}$ ,  $X_{10}X_{15}$ , dan  $X_{12}X_{15}$ . Sebanyak empat pengaruh utama dan empat interaksi yang ada di model GA2 tidak ada di model RB2 yaitu  $X_5$ ,  $X_7$ ,  $X_{14}$ ,  $X_{24}$ ,  $X_3X_5$ ,  $X_3X_{24}$ ,  $X_7X_{20}$ , dan  $X_{10}X_{14}$ . Tidak adanya pengaruh utama atau interaksi di suatu model bisa saja perannya digantikan oleh pengaruh utama atau interaksi di dalam model tersebut.

Tabel 3: Korelasi pengaruh utama dan interaksi model GA2 dan RB2.

Model GA2		Model RB2
		$X_{10}X_{15}$
$X_{24}$	Koefisien korelasi Spearman	-0.577
	<i>p-value</i>	0.031

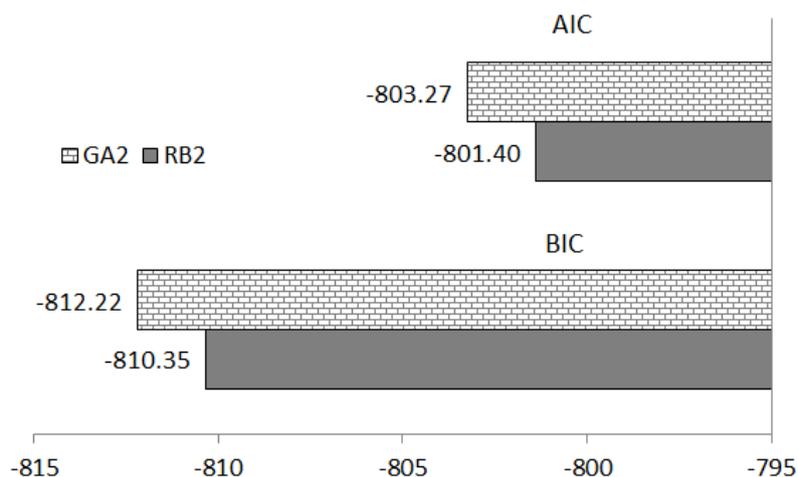
Tabel 3 menunjukkan terdapat pengaruh utama pada model GA2 dan interaksi pada model RB2 yang saling berkorelasi sehingga perannya dapat saling

menggantikan. Faktor  $X_{24}$  pada model GA2 perannya digantikan oleh interaksi  $X_{10}X_{15}$  pada model RB2.

Korelasi yang kuat tersebut mengindikasikan adanya pembauran sebagian antara pengaruh utama dan interaksi. Hal ini karena terdapat pengaruh utama yang berkorelasi kuat dengan interaksi namun korelasinya tidak penuh artinya tidak bernilai -1 atau +1.

Pembauran selalu terjadi pada rancangan lewat jenuh baik pembauran penuh maupun pembauran sebagian. Pembauran tersebut menyebabkan ketidakbebasan pendugaan pengaruh utama dan interaksi. Pada model RB2 dan GA2 terdapat interaksi yang menyebabkan sebagian besar pengaruh utama menjadi tidak berpengaruh nyata dan terdapat perubahan signifikansi pengaruh utama setelah interaksi yang melibatkan pengaruh utama tersebut dikeluarkan dari model. Hal ini juga menunjukkan adanya pembauran antara pengaruh utama dan interaksi dua pengaruh utama karena pengaruhnya berubah atau terjadi ketidakbebasan pendugaan pengaruh.

Gambar 7 menunjukkan bahwa model GA2 memiliki nilai AIC dan BIC lebih kecil dibandingkan model RB2. Dengan demikian model GA2 lebih baik dibandingkan model RB2.



Gambar 7: Nilai AIC dan BIC model RB2 dan GA2.

Model hasil algoritme genetika memiliki nilai AIC dan BIC terkecil sebagai prioritas utama pemilihan model terbaik. Semua pengaruh utama dan interaksi berpengaruh nyata pada taraf nyata 0.1% baik model hasil algoritme genetika maupun model hasil regresi bertatar sedangkan model hasil penelitian sebelumnya yaitu penelitian Dejaegher *et al.* (2007) hanya satu pengaruh utama yang berpengaruh nyata. Signifikansi pengaruh utama dan interaksi tersebut merupakan prioritas kedua pemilihan model terbaik. Model dengan waktu komputasi tercepat sebagai prioritas terakhir pemilihan model terbaik yaitu model hasil regresi bertatar dengan rata-rata waktu komputasi 2.489 detik, sedangkan rata-rata waktu komputasi model hasil algoritme genetika selama 26.26 menit. Dengan demikian urutan model terbaik penelitian ini yaitu urutan pertama model hasil algoritme genetika (model GA2), urutan kedua model hasil regresi bertatar (model RB2), dan urutan ketiga model hasil penelitian Dejaegher *et al.* (2007) (model L2).

### 3.3 Perbandingan Hasil Seleksi Peubah Algoritme Genetika dan Regresi Bertatar di Kedua Data

Pengaruh utama dan interaksi yang terpilih pada model hasil penelitian sebelumnya, algoritme genetika, dan regresi bertatar tidak banyak yang sama. Hal ini dapat terjadi karena algoritme pemilihan peubah di ketiga metode tersebut berbeda. Selain itu, adanya pembauran pada rancangan faktorial pecahan merupakan salah satu penyebab tidak bebasnya pendugaan pengaruh utama dan interaksi sehingga peran suatu pengaruh utama didalam suatu model dapat digantikan oleh interaksi dan sebaliknya. Hal ini dapat diindikasikan dengan adanya korelasi yang kuat antara pengaruh utama dengan pengaruh utama lain atau antara pengaruh utama dengan interaksi dan adanya pengaruh utama yang menjadi tidak signifikan apabila interaksi yang melibatkan pengaruh utama tersebut dikeluarkan dari model. Walaupun demikian, ketiga model masih dapat digunakan untuk memilih faktor-faktor penting.

Model hasil regresi bertatar tidak berbeda di setiap ulangan dan modelnya juga memiliki nilai AIC dan BIC yang sama yaitu lebih besar dibandingkan model hasil algoritme genetika. Pada data percobaan pembuatan amida lemak sulfat, nilai AIC model hasil regresi bertatar lebih besar 5.311 poin dan 5.318 poin untuk nilai BIC. Pada data percobaan pembuatan karet, nilai AIC dan BIC model hasil regresi bertatar lebih besar 1.869 poin. Hal ini mengindikasikan bahwa regresi bertatar terjebak pada solusi optimum lokal sedangkan algoritme genetika tidak demikian. Algoritme genetika masih mungkin menemukan model yang memiliki nilai AIC dan BIC yang lebih kecil sehingga solusinya lebih mendekati solusi optimum global dengan cara yang lebih singkat walaupun tidak selalu menemukan solusi terbaik (Sartono *et al.*, 2015). Regresi bertatar merupakan *greedy algorithm* sehingga lebih mudah konvergen karena kurangnya penjelajahan ruang solusi sehingga terjebak pada solusi optimum lokal (Blum & Roli, 2003).

Waktu komputasi untuk mendapatkan model hasil algoritme genetika jauh lebih lama dibandingkan model hasil regresi bertatar di kedua data. Hal ini dapat terjadi karena regresi bertatar sudah lebih lama dikembangkan dalam pemrograman R oleh para ahli dibidangnya dan sudah tersedia dalam bentuk *package* sedangkan pemrograman algoritme genetika pada penelitian ini masih dalam proses pengembangan dan belum tersedia dalam bentuk *package*.

## 4. Simpulan

Algoritme genetika dapat digunakan sebagai metode seleksi peubah pada data rancangan faktorial pecahan lewat jenuh dua taraf. Penelitian ini menunjukkan bahwa algoritme genetika merupakan metode terbaik untuk memilih peubah di kedua data dibandingkan regresi bertatar dan penelitian sebelumnya. Model hasil algoritme genetika memiliki nilai AIC dan BIC terkecil, serta semua pengaruh utama dan interaksi berpengaruh nyata pada taraf nyata 0.1%. Waktu komputasi algoritme genetika jauh lebih lama dibandingkan regresi bertatar di kedua data. Lamanya waktu komputasi dapat disebabkan oleh berbagai faktor seperti perangkat keras atau lunak yang digunakan dan pemrograman algoritme genetika yang masih dalam tahapan pengembangan sehingga masih perlu diperbaiki.

## Daftar Pustaka

- Abraham, B., Chipman, H., & Vijayan, K. (1999). Some risks in the construction and analysis of supersaturated designs. *Technometrics*, 41(2): 135–141.
- Blum, C., & Roli, A. (2003). Metaheuristics in combinatorial optimization: Overview and conceptual comparison. *ACM Computing Surveys (CSUR)*, 35(3): 268–308.
- Dejaegher, B., Capron, X., & Vander Heyden, Y. (2007). Generalized FEAR method to estimate factor effects in two-level supersaturated designs. *Journal of Chemometrics: A Journal of the Chemometrics Society*, 21(7–9): 303–323.
- Draper, N. R., & Smith, H. (1992). *Analisis regresi terapan*. Jakarta (ID): PT Gramedia Pustaka Utama.
- Lin, D. K. (1993). A new class of supersaturated designs. *Technometrics*, 35(1): 28–31.
- Paterlini, S., & Minerva, T. (2010). Regression model selection using genetic algorithms. *Proceedings of the 11th WSEAS International Conference on Neural Networks and 11th WSEAS International Conference on Evolutionary Computing and 11th WSEAS International Conference on Fuzzy Systems*, 19–27. World Scientific and Engineering Academy and Society (WSEAS).
- Rais, F., Kamoun, A., Chaabouni, M., Claeys-Bruno, M., Phan-Tan-Luu, R., & Sergent, M. (2009). Supersaturated design for screening factors influencing the preparation of sulfated amides of olive pomace oil fatty acids. *Chemometrics and Intelligent Laboratory Systems*, 99(1): 71–78.
- Salawu, I. S., Adeleke, B. L., & Oyeyemi, G. (2015). Review of classical methods in supersaturated designs (SSD) for factor screening. *Journal of Mathematical Theory and Modeling*, 5(5): 38–44.
- Sartono, B., Rahardianto, S., & Hidayatuloh, A. (2015). *Penerapan algoritma genetik dalam analisis statistika*. Bogor (ID): Herya Media.
- Williams, K. (1968). Designed experiments. *Rubber Age*, 100: 65–71.