

Evolution Strategies Untuk Optimasi Pembentukan Fungsi Regresi Linier Dalam Menentukan Kebutuhan Volume Air Penyiraman Tanah

Robertus Santoso Aji Putra¹, Nurul Hidayat², Mochammad Ali Fauzi³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹ajiputra001@gmail.com, ²ntayadih@ub.ac.id, ³moch.ali.fauzi@ub.ac.id

Abstrak

Laboratorium Benih Balai Pengkajian Teknologi Pertanian Jawa Timur adalah salah satu unit kerja milik pemerintah provinsi yang ditunjuk sebagai pelaksana teknis untuk melakukan pengkajian daya tumbuh benih. Saat ini di tempat tersebut sedang dikembangkan alat siram otomatis berdasarkan sensor kelembapan tanah, namun alat itu belum bisa memprediksi kebutuhan volume air guna menjaga kelembapan media penumbuhan benih. Dengan bantuan sensor kelembapan pada alat dan pengetahuan pakar, data berupa kumpulan amatan kelembapan tanah terhadap kebutuhan volume air telah didapatkan. Penelitian ini dilakukan untuk menerapkan metode regresi linier agar alat tersebut dapat melakukan prediksi berdasarkan pola data dalam bentuk persamaan. Ketepatan hasil prediksi dengan metode ini diukur dengan koefisien determinasi. Koefisien determinasi dapat menurun akibat adanya amatan pencilan yang timbul karena ketidakakuratan hasil observasi. Dari permasalahan tersebut penelitian ini menggunakan *evolution strategies* dengan kriteria informasi sebagai pembanding untuk mendeteksi amatan pencilan guna dihilangkan. Setelah menghilangkan 6 amatan pencilan yang terdeteksi oleh *evolution strategies* pada penelitian ini, terdapat peningkatan koefisien determinasi dari 0.9673 menjadi 0.9935.

Kata kunci: *regresi linier, amatan pencilan, evolution strategies, kriteria informasi*

Abstract

Seed Laboratory BPTP East Java is one of provincial government work units that was appointed as the technical implementer to conduct a study in seed growth. Currently this facility is developing automatic watering device based on soil humidity sensor, but the device cannot predict the volume of water needed in order to keep the soil moisture. With the help of humidity sensor on device and expert's knowledge, the observations dataset of soil moisture to the waters volume requirement are obtained. This study was conducted to implements linear regression method so that the device can perform predictions based on dataset patterns as an equation. The accuracy of prediction results with this method is measured by the coefficient of determination. The coefficient of determination can be decreasing due to the arising of observation outliers because Inaccuracy of observation results. The solution from this study is using evolution strategies with information criteria as comparison for detecting observation outliers to eliminated. After eliminating 6 observations outliers were detected by evolution strategiess in this study, shows increase in the coefficient of determination from 0.9673 to 0.9935.

Keywords: *linear regression, observation outliers, evolution strategies, information criteria*

1. PENDAHULUAN

Balai Pengkajian Teknologi Pertanian Jawa Timur yang selanjutnya disingkat BPTP Jatim adalah unit kerja milik pemerintah provinsi yang memiliki salah satu visi yaitu menyediakan teknologi yang dapat menunjang komoditas pertanian sesuai dengan kondisi

lingkungan yang dibutuhkan. Salah satu fasilitas yang dimiliki BPTP Jatim adalah Laboratorium Benih, yang ditunjuk sebagai pelaksana teknis untuk melakukan pengkajian kualitas benih. Tugas utama dari laboratorium ini adalah melayani permintaan analisis dari berbagai instansi untuk kepentingan penelitian serta melayani permintaan analisis langsung

dari petani (Balai Pengkajian Teknologi Pertanian Jawa Timur, 2010).

Laboratorium Benih BPTP Jatim telah memiliki beberapa alat serta metode untuk menguji kualitas benih, namun pengujian kualitas benih harus disertai dengan pengujian daya tumbuh benih. Sampai saat ini penyiraman media penumbuhan benih di Laboratorium Benih BPTP Jatim masih menggunakan serangkaian pipa untuk menyalurkan air serta belum bekerja secara otomatis guna menghasilkan kelembapan tanah ideal. Hal ini tentu dapat mempengaruhi hasil pertumbuhan benih, dikarenakan untuk tumbuh secara optimal benih membutuhkan volume penyiraman air yang sesuai guna menjaga kelembapan tanah tetap ideal (FAO Land & Water Division, 2003). Diharapkan dengan adanya alat yang dapat menjaga kelembapan tanah secara otomatis di Laboratorium Benih BPTP Jatim hasil pengujian daya tumbuh benih lebih optimal, dikarenakan penyiraman media penumbuhan benih lebih tepat sesuai kebutuhan. Saat ini di Laboratorium Benih BPTP Jatim sedang dikembangkan alat yang dapat melakukan penyiraman secara otomatis. Alat yang sedang dikembangkan tersebut menggunakan sensor yang berfungsi sebagai penerima masukan tingkat kelembapan tanah. Hal ini bertujuan untuk menentukan volume air yang dibutuhkan dalam penyiraman, namun belum ada penghubung antara masukan dan keluaran pada alat tersebut. Tanpa adanya penghubung antara masukan dan keluaran pada alat yang reliable maka akan memunculkan masalah seperti jumlah penyiraman yang terlalu sering ataupun terlalu sedikit yang akan sangat mempengaruhi kandungan nutrisi tanah dan dapat mempengaruhi pertumbuhan tanaman, maka diperlukanlah metode untuk menghubungkan antara input dan output pada alat.

Berdasarkan permasalahan di atas, peneliti ingin menerapkan suatu persamaan pada pengembangan alat siram otomatis yang sedang dikembangkan, sehingga alat tersebut dapat memprediksi kebutuhan volume air berdasarkan kelembapan tanah. Persamaan tersebut didapat dari metode yang dapat mengenali pola hubungan kebutuhan volume air berdasarkan kelembapan tanah awal hingga mencapai ideal. Dengan bantuan sensor kelembapan pada alat yang sedang dikembangkan dan pengetahuan pakar terhadap kelembapan tanah ideal, data berupa kumpulan amatan kebutuhan volume air

berdasarkan kelembapan tanah dapat dicari. Data yang digunakan pada penelitian ini merupakan kumpulan amatan yang cenderung membentuk pola garis lurus dengan asumsi semakin kecil kelembapan tanah maka semakin besar volume air yang dibutuhkan atau sebaliknya. Metode yang cocok digunakan untuk permasalahan ini adalah regresi linier. Regresi linier mempunyai persamaan yang disebut sebagai persamaan regresi. Persamaan regresi merumuskan hubungan yang lurus antara satu amatan dengan amatan lainnya. Setiap amatan mempresentasikan hubungan variabel terikat dengan satu atau lebih variabel bebas (Levin & Rubin, 1998).

Dalam regresi linier, bisa jadi persamaan yang terbentuk kurang kuat untuk menggambarkan pola hubungan antar amatan. Persamaan regresi yang kurang kuat dapat menyebabkan prediksi menjadi salah. Hal ini dikarenakan adanya amatan pencilan (outliers) di dalam data yang ikut membentuk persamaan regresi. Amatan pencilan merupakan amatan yang memiliki nilai terlalu jauh dari amatan lainnya. Dalam penelitian ini amatan pencilan bisa jadi ada di dalam data karena ketidakakuratan sensor dalam menentukan kelembapan tanah atau kurangnya ketelitian dalam mengamati kebutuhan volume air. Untuk mengantisipasi kesalahan prediksi, perlu adanya pendeteksian amatan pencilan sebelum membentuk persamaan regresi (Johnson & Wichern, 1996). Pendeteksian amatan pencilan bisa melalui beberapa kali proses. Hal ini bisa terjadi karena setelah amatan yang diduga sebagai pencilan dihapus persamaan regresi yang terbentuk akan berubah. Perubahan persamaan regresi membuat amatan yang tadinya bukan pencilan menjadi pencilan baru. Selain dibutuhkan proses yang berulang, satu proses pendeteksian membutuhkan beberapa tahap yang harus dilalui (Hadi & Simonoff, 1993). Banyaknya tahap yang dilalui serta proses yang berulang membuat cara ini kurang efisien sehingga perlu satu cara untuk melakukan deteksi pencilan secara serempak. Pada tahun 1992 Mills dan Prasad mengusulkan metode pendeteksian amatan pencilan secara serempak dengan rumus kriteria informasi. Amatan pencilan ditentukan dengan menghitung nilai kriteria informasi terbaik berdasarkan kombinasi amatan yang dimasukkan dalam rumus. Kelemahan dari metode ini adalah semua kombinasi harus dicoba untuk menghasilkan nilai kriteria

informasi terbaik sehingga mengurangi efisiensi. Solusi yang dapat digunakan untuk menyelesaikan permasalahan ini adalah *evolution strategies*. *Evolution strategies* dapat mencari nilai kriteria informasi terbaik tanpa mencoba semua kombinasi (Alma, Kurt, & Ugur, 2008).

Berdasarkan pemaparan latar belakang di atas, maka peneliti ingin meneliti *Evolution strategies* untuk Optimasi Pembentukan Fungsi Regresi Linier dalam Menentukan Kebutuhan Volume Air Penyiraman Tanah. Hasil dari penelitian ini berupa persamaan yang akan digunakan pada pengembangan alat siram otomatis di Laboratorium Benih BPTP Jatim. Dengan menerapkan hasil penelitian ini pada pengembangan alat siram otomatis tersebut, diharapkan dapat membantu Laboratorium Benih BPTP Jatim dalam menjaga kelembapan media penumbuhan benih..

2. REGRESI LINEAR

Regresi linier adalah metode statistik yang cocok digunakan untuk menghitung pola hubungan variabel terikat dengan satu atau lebih variabel bebas pada masing-masing amatan. Regresi linier juga dapat digunakan untuk melakukan prediksi pada variabel terikat, dengan cara memasukkan variabel bebas yang ingin dicari nilai variabel terikatnya ke dalam persamaan tersebut. Namun yang perlu diingat, saat melakukan prediksi menggunakan persamaan regresi variabel bebas yang dimasukkan harus berada pada rentang data yang digunakan. Misalnya, persamaan regresi diperoleh menggunakan data dengan variabel bebas 1 sampai 100, maka variabel bebas yang kita gunakan untuk melakukan prediksi hanya boleh antara 1 sampai 100. Konsep ini disebut sebagai interpolasi (R Development Core Team, 2008).

Untuk melakukan prediksi menggunakan regresi linier, terlebih dahulu harus dicari persamaan regresinya (Kenney & Keeping, 1962). Rumus persamaan regresi linier untuk satu variabel bebas ditunjukkan pada persamaan berikut:

$$\hat{Y} = a + bX$$

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n(\sum X^2) - (\sum X)^2}$$

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

Dimana \hat{Y} =variabel terikat hasil prediksi

yang didapatkan dengan memasukkan variabel bebas (X) pada persamaan regresi, X =variabel bebas, Y =variabel terikat, a =konstanta, b =koefisien regresi (kemiringan) dan n =jumlah amatan.

3. OPTIMASI REGRESI LINEAR

Optimasi diperlukan untuk memperkuat persamaan regresi dengan cara mendeteksi amatan pencilan guna dihilangkan sebelum membentuk persamaan tersebut (Johnson & Wichern, 1996).

Cara yang biasa dilakukan untuk mendeteksi amatan pencilan yaitu membandingkan sisaan pada setiap amatan, namun cara ini kurang efisien karena bisa jadi membutuhkan proses berulang. Pada penelitian ini amatan pencilan dideteksi menggunakan *evolution strategies* dengan kriteria informasi sebagai fitness

4. EVOLUTION STRATEGIES

Algoritma *Evolution strategies* mulai dikenal pada tahun 1960-an dimana algoritma ES paling sering digunakan sebagai solusi dari permasalahan *black-box* pada ruang pencarian terus menerus (Hansen, Arnold, & Auger, 2014). Algoritma ES lebih mengutamakan representasi bilangan pecahan (*real-vector*) pada representasi kromosomnya.

. Algoritma ES lebih mengutamakan proses mutasi sebagai operator reproduksinya sedangkan proses rekombinasi dijadikan sebagai opsi tambahan. Algoritma ES lebih mengutamakan operator mutasi pada proses reproduksinya sedangkan rekombinasi dijadikan sebagai opsi tambahan, yang dapat digunakan ataupun tidak digunakan pada proses reproduksi.

ES memiliki empat tipe siklus yang akan dijelaskan pada Tabel 1

Tabel 1 Tipe Siklus Algoritma *Evolution strategies*

N	Tipe Siklus Algoritma ES	Deskripsi
1	ES (μ, λ)	Siklus ES (μ, λ) tidak menggunakan Operator rekombinasi pada tahapan reproduksinya. Tahapan seleksi,

		siklus ES (μ, λ) hanya melibatkan <i>offspring</i> sebagai kandidat individu yang akan diseleksi menjadi individu baru pada generasi berikutnya.
2	ES ($\mu/R, \lambda$)	Siklus ES ($\mu/R, \lambda$) menggunakan operator rekombinasi pada tahapan reproduksinya. Pada tahapan seleksi siklus ES ($\mu/R, \lambda$) hanya melibatkan <i>offspring</i> sebagai kandidat individu yang akan diseleksi menjadi individu baru pada generasi berikutnya.
3	ES ($\mu + \lambda$)	Siklus ES ($\mu + \lambda$) tidak menggunakan operator rekombinasi pada tahapan reproduksinya. Pada tahapan seleksi, siklus ES ($\mu + \lambda$) melibatkan <i>offspring</i> dan induk sebagai kandidat individu yang akan diseleksi menjadi individu baru pada generasi berikutnya.
4	ES ($\mu/R + \lambda$)	Siklus ES ($\mu/R + \lambda$) menggunakan operator rekombinasi pada tahapan reproduksinya. Pada tahapan seleksi, siklus ES ($\mu/R + \lambda$) melibatkan <i>offspring</i> dan induk sebagai kandidat individu yang akan diseleksi menjadi individu baru pada generasi berikutnya.

Algoritma ES memiliki siklus yang mirip dengan algoritma evolusi lainnya, yakni algoritma ES juga menerapkan proses inialisasi, reproduksi, evaluasi, dan seleksi.

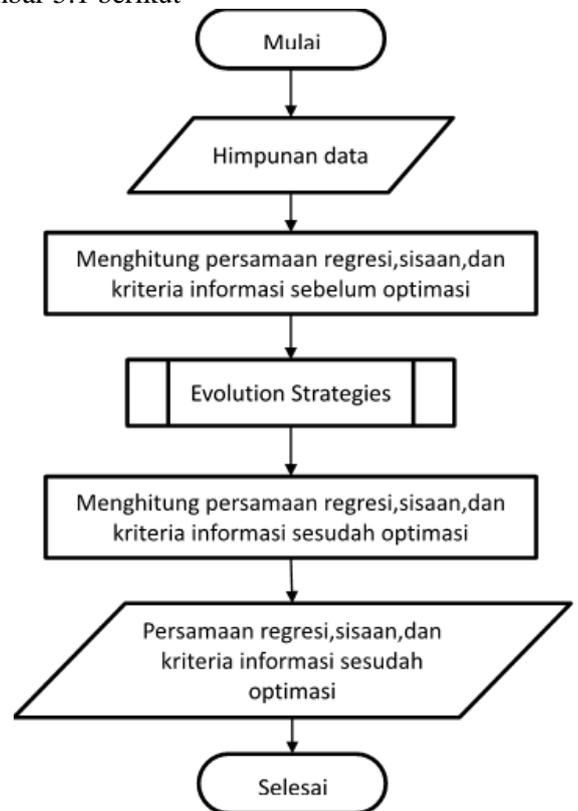
Siklus algoritma ES akan dideskripsikan melalui *pseudocode* pada Gambar 1

Baris	Prosedur ES ($\mu/r, \lambda$)
1.	Begin
2.	$t := 0$
3.	inisialisasi $P(t) := \{(p_m(0), q_m(0)), m=1, \dots, \mu\}$;
4.	while (bukan kondisi berhenti) do
5.	for $i=1$ to λ do begin
6.	$\alpha_i := \text{pilihInduk}(P(t), r)$;
7.	$p_i := \text{rekombinasi}(\alpha_i)$;
8.	$q_i := \text{rekombinasi}(\alpha_i)$;
9.	$p'_i := \text{mutasi}(p_i)$;
10.	$q'_i := \text{mutasi}(q_i)$;
11.	$F_i := \text{anak}(p'_i, q'_i)$;
12.	end
13.	$C(t) := \{(p_l(t), q_l(t)), l=1, \dots, \lambda\}$;
14.	case tipe_seleksi of
15.	case ($\mu/r, \lambda$):
16.	$P(t+1) := \text{seleksi}(C(t), \mu)$;
17.	case ($\mu/r + \lambda$):
18.	$P(t+1) := \text{seleksi}(P(t), C(t), \mu)$;
19.	end
20.	$t := t+1$;
21.	end while
22.	End

Gambar 1 Pseudocode *Evolution strategies*

5. PERANCANGAN

Diagram alir sistem akan dijelaskan pada gambar 5.1 berikut



Gambar 5. 1 Diagram alir

Diagram alir pada gambar 5.1 diatas menjelaskan tentang proses data dari data asli hingga menentukan hasil optimasi dengan menggunakan metode *evolution strategies*.

5.1 Perhitungan Kriteria Informasi

Pada proses perhitungan kriteria informasi informasi dimulai dari membentuk persamaan regresi guna mencari sisaan (*error*) untuk masing-masing amatan. Sisaan ini nantinya akan digunakan untuk menghitung penduga ragam model regresi yang dibutuhkan dalam mencari nilai kriteria informasi. Proses pencarian Kriteria Informasi akan di jelaskan pada persamaan dibawah ini:

1. Hitung nilai konstanta (*a*).

$$a = \frac{(\sum Y)(\sum X^2) - (\sum X)(\sum XY)}{n(\sum X^2) - (\sum X)^2}$$

$$a = \frac{(29.925)(316.3611) - (91.89)(111.427375)}{30(316.3611) - (91.89)^2}$$

$$a = -0.737259477$$
2. Hitung nilai kemiringan (*b*).

$$b = \frac{n(\sum XY) - (\sum X)(\sum Y)}{n(\sum X^2) - (\sum X)^2}$$

$$b = \frac{30(111.427375) - (91.89)(29.925)}{30(316.3611) - (91.89)^2}$$

$$b = 0.566359607$$

Setelah konstanta dan kemiringan didapatkan maka proses selanjutnya adalah menghitung variabel terikat prediksi dan sisaan.

1. Menghitung variabel terikat prediksi.

$$\hat{Y} = a + bX$$

$$\hat{Y} = -0.737259477 + 0.566359607 * 3.75$$

$$\hat{Y} = 1.386589$$
2. Menghitung sisaan.

$$e = Y - \hat{Y}$$

$$e = 1.325 - 1.386589$$

$$e = -0.06159$$

Untuk menghitung penduga ragam model regresi, hanya nilai sisaan dari amatan dengan gen bernilai 1 saja yang digunakan. Ragam model regresi akan digunakan untuk menghitung nilai kriteria informasi, yang mana digunakan sebagai *fitness* pada *evolution strategies*. Proses perhitungan penduga ragam model regresi ditunjukkan pada persamaan dibawah ini:

1. Menghitung penduga ragam model regresi.

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n-m_d} (e_i^2)}{n - m_d}$$

$$\hat{\sigma}^2 = \frac{0.06579}{30 - 6}$$

$$\hat{\sigma}^2 = 0.00274123$$

2. Menghitung kriteria informasi.

$$KI = \log \hat{\sigma}^2 + (1 + p) \frac{\log n}{n} + Km_d \frac{\log n}{n}$$

$$KI = \log 0.00274123 + (1 + 1) \frac{\log 30}{30} + 1.5 * 6 \frac{\log 30}{30}$$

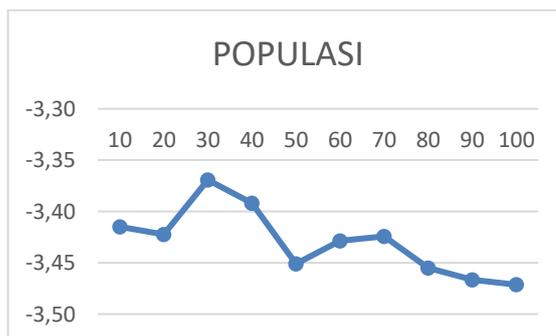
$$KI = -2.020443374$$

Proses di atas merupakan hasil akhir dari perhitungan nilai kriteria informasi. Berdasarkan proses di atas, nilai kriteria informasi akan berbeda untuk setiap susunan gen pada suatu individu meskipun dalam proses pencarian persamaan regresinya tanpa melibatkan susunan gen.

6. PENGUJIAN DAN PEMBAHASAN

6.1 Hasil Pengujian Populasi

Pengujian ini dilakukan pada jumlah populasi 10 hingga 100 dengan kelipatan 10. Parameter lain yang digunakan pada pengujian ini yaitu jumlah generasi 10, jumlah *offspring* sebesar 0,5μ. Pengujian ini dilakukan untuk mencari jumlah populasi yang menghasilkan fitness terbaik. Hasil dari pengujian dapat dilihat pada gambar dibawah ini.



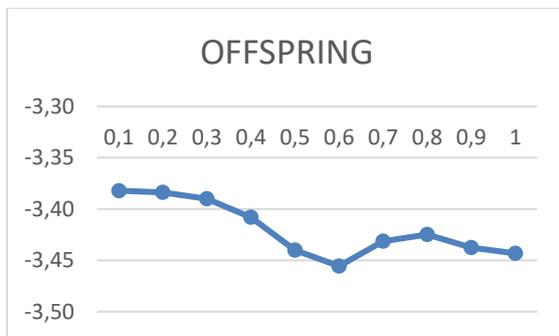
Gambar 6. 1 Hasil pengujian populasi

dari gambar diatas dapat disimpulkan bahwa jumlah populasi yang menghasilkan fitness terbaik adalah sebanyak 30 populasi.

6.2 Hasil Pengujian Offspring

Pengujian ini dilakukan pada jumlah *offspring* 0,1μ hingga 1μ dengan kelipatan 0,1. Parameter lain yang digunakan pada pengujian

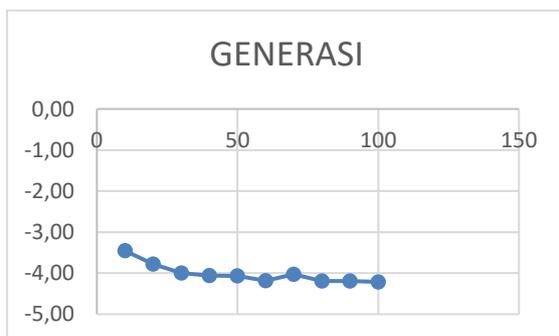
ini yaitu jumlah generasi 10, jumlah populasi sebesar 40. Pengujian ini dilakukan untuk menemukan jumlah offspring yang menghasilkan fitness terbaik. Hasil dari pengujian ini dapat dilihat pada gambar dibawah ini.



Gambar 6. 2 Hasil pengujian offspring

6.3 Hasil Pengujian Generasi

Pengujian ini dilakukan pada jumlah generasi 10 hingga 100 dengan kelipatan 10. Parameter lain yang digunakan pada pengujian ini yaitu jumlah populasi 40, jumlah offspring 0.5μ. Pengujian ini dilakukan untuk mencari jumlah generasi yang menghasilkan fitness terbaik. Hasil dari pengujian ini dapat dilihat pada gambar dibawah ini.



Gambar 6. 3 Hasil pengujian generasi

dari gambar diatas dapat disimpulkan bahwa jumlah generasi yang menghasilkan fitness terbaik adalah 10 generasi.

7. KESIMPULAN

Regresi linier sebagai alat prediksi dapat dioptimasi menggunakan *evolution strategies* dengan nilai kriteria informasi sebagai *fitness* guna menghilangkan amatan pencilan, dengan syarat parameter *evolution strategies* yang digunakan sesuai. Parameter yang sesuai untuk

digunakan yaitu jumlah individu dan jumlah generasi, namun jika terlalu besar tidak akan ada kenaikan *fitness* bahkan bisa terjadi penurunan dengan proses semakin banyak. Pada penelitian ini *fitness* yang baik diperoleh dengan jumlah populasi 30 dan jumlah generasi 10. Parameter lainnya yaitu jumlah *offspring*, parameter ini dapat menghasilkan *fitness* yang baik jika cenderung seimbang. Pada penelitian ini jumlah offspring dengan *fitness* terbaik diperoleh pada 0.1μ

DAFTAR PUSTAKA

- Alma, Ö. G., Kurt, S., & Ugur, A. (2008). Genetic algorithms for outlier detection in multiple regression with different information criteria. *Journal of Statistical Computation and Simulation*, 29–47.
- Balai Pengkajian Teknologi Pertanian Jawa Timur. (2010, 5 10). *Laboratorium Benih*. Retrieved 4 1, 2017, from Balai Pengkajian Teknologi Pertanian Jawa Timur: <http://jatim.litbang.pertanian.go.id/ind/index.php/layanan/laboratorium-benih>
- Draper, N. R., & Smith, H. (1998). *Analisis Regresi Terapan*. Jakarta: PT. Gramedia Pustaka Utama.
- FAO Land & Water Division. (2003, 3 1). *Crop Water Information*. Retrieved 1 4, 2017, from FAO Land & Water Management: <http://www.fao.org/nr/water/cropinfo.htm>
- Gen, M., & Cheng, R. (1997). *Genetic Algorithms and Engineering Design*. New York: John Willey & Sons, Inc.
- Hadi, A. S., & Simonoff, J. S. (1993). Procedures for the Identification of Multiple Outliers in Linear Models. *Journal of the American Statistical Association*, 88.
- Johnson, R. A., & Wichern, D. W. (1996). *Applied Multivariate Statistical*. New Jersey: Prentice Hall of India Private Limited.
- Kenney, J. F., & Keeping, E. S. (1962). *Linear Regression and Correlation*. Princeton, NJ: Van Nostrand.
- Levin, R. I., & Rubin, D. S. (1998). *Statistics for management*. New delhi: Prentice Hall.

- Mahmudy, W. F. (2013). *Algoritma Evolusi*. Malang: Program Teknologi Informasi dan Ilmu Komputer (PTIIK) Universitas Brawijaya.
- Mills, J., & Prasad, K. (1992). A Comparison of Model Selection Criteria. *Econometric Rev 11*, 201–233.
- R Development Core Team. (2008). *A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Schwarz, G. (1978). Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2), 461-464.
- Software and Systems Engineering Vocabulary*. (2008, 7 31). Retrieved 11 2016, 24, from SEVOCAB:
https://pascal.computer.org/sev_display/search.action;jsessionid=5f80746ac7214cf2e2ea49dd3494
- Tolvi, J. (2004). Genetic algorithms for outlier detection and variable selection in linear regression models. *Soft Computing* 8, 527–533.
- Yansari, M., Ratnawati, D. E. & M., 2016. *Optimasi Biaya dan Asupan Gizi Pasien Diet Khusus Dengan Menggunakan Algoritma Evolution strategies*. Malang: Fakultas Ilmu Komputer Universitas Brawijaya