

PEMODELAN REGRESI NON LINEAR MENGGUNAKAN ALGORITMA GENETIKA UNTUK PREDIKSI KEBUTUHAN AIR PDAM KOTA MALANG

Vitara Nindya Putri Hasan¹, Wayan Firdaus Mahmudy²

¹Mahasiswa Program Studi Teknik Informatika/Ilmu Komputer Universitas Brawijaya

²Dosen Program Studi Teknik Informatika/Ilmu Komputer Universitas Brawijaya

Program Studi Teknik Informatika/Ilmu Komputer

Fakultas Ilmu Komputer

Universitas Brawijaya

Jalan Veteran Malang 65145, Indonesia

Email: ¹vitaranindyaph@gmail.com, ²wayanfm@ub.ac.id

ABSTRAK

Kota Malang adalah kota yang selalu mengalami peningkatan jumlah penduduk setiap tahunnya karena banyaknya penduduk yang memilih transmigrasi dan menetap tinggal di Malang. Hal ini disebabkan oleh lokasi Kota Malang yang berada di dataran tinggi dan dikelilingi oleh pegunungan sehingga memiliki iklim yang sejuk, selain itu Kota Malang dijuluki sebagai kota pendidikan. Seiring dengan meningkatnya jumlah penduduk di Kota Malang maka meningkat pula kebutuhan konsumen air bersih dari PDAM. Perubahan pemakaian air tersebut jika tidak diolah dengan baik maka akan menyebabkan beberapa persoalan diantaranya apabila PDAM terlalu banyak mendistribusikan air bersih ke konsumen maka akan berakibat pemborosan air dan sebaliknya apabila distribusi air bersih PDAM kurang maka konsumen akan kekurangan air bersih. Oleh karena itu dibutuhkan suatu estimasi untuk memperkirakan dengan tepat seberapa besar volume air yang diperlukan di tahun-tahun berikutnya. Permasalahan tersebut akan dimodelkan dengan persamaan regresi non linear yang terdiri dari variabel bebas (X) dan variabel terikat (Y). Solusi yang dapat diberikan adalah dengan memberikan koefisien regresi terbaik menggunakan konsep Algoritma Genetika. Proses reproduksi menggunakan *one-point-crossover* dan *random mutation*, untuk proses seleksinya menggunakan model *elitism selection*. Dari Algoritma Genetika tersebut didapatkan parameter terbaik yaitu ukuran populasi sebanyak 225, generasi terbaik sebanyak 1750 generasi, kombinasi *cr* : *mr* adalah 0,6 : 0,4 dengan nilai *fitness* tertinggi yaitu 107.997. Hasil akhir berupa variabel yang menjadi model regresi.

Kata kunci: Regresi Non Linear, Algoritma Genetika, Prediksi, Pemakaian air.

ABSTRACT

Malang is a city that the number of people every year is always increasing, because many people choose transmigration and stay in Malang. This is caused by the location of Malang that is located in the highlands and surrounded by mountains so it has a cool climate, furthermore Malang known as the city of education. Along with the increasing population in Malang it also increase the consumer water consumption from PDAM. The change of water consumption if it is not treated properly, it will cause some problems when the PDAM has too many of water to distribute to consumers it will result in wastage of water and otherwise if the distribution of water less than normal, then the consumer will have a shortage of water. Therefore it is necessary to estimate to predict exactly how much the water volume needed in subsequent years. This problem will be modeled with non linear regression that consist of the independent variable (X) and the dependent variable (Y). Solution that can be given is to provide the best regression coefficient using the concept of Genetic Algorithm. In the reproduction process using one-point-crossover and random mutation, for the selection process using a elitism selection models. Of the genetic algorithm parameters obtained the best population size is 225, the best generation as much as 1750 generation, combination of *cr* : *mr* is 0,6 : 0,4 with the highest fitness value is 107.997. The final result is a variable that changed into the regression model.

Keywords : Non Linear Regression, Genetic Algorithm, Predict, Water Consumption

1. PENDAHULUAN

Pendahuluan akan membahas mengenai latar belakang, rumusan masalah dan batasan masalah.

1.1 Latar Belakang

Kota Malang adalah kota yang selalu mengalami peningkatan jumlah penduduk setiap tahunnya, hal ini disebabkan oleh beberapa faktor diantaranya lokasi Kota Malang yang berada di dataran tinggi dan dikelilingi oleh pegunungan sehingga memiliki iklim

yang sejuk, selain itu Malang memiliki fasilitas dan kualitas pendidikan yang memadai. Seiring dengan meningkatnya jumlah penduduk di Kota Malang maka meningkat pula kebutuhan konsumen akan air bersih dari PDAM (Syaiful & Achmad, 2014). Perubahan pemakaian air tersebut jika tidak diolah dengan baik maka akan menyebabkan beberapa persoalan diantaranya apabila PDAM terlalu banyak mendistribusikan air bersih ke konsumen maka akan berakibat pemborosan air dan sebaliknya apabila distribusi air bersih PDAM kurang maka konsumen akan kekurangan air bersih. Oleh karena itu dibutuhkan suatu estimasi untuk memperkirakan dengan tepat seberapa besar volume air PDAM Kota Malang yang diperlukan untuk melayani kebutuhan air para konsumen PDAM Kota Malang di tahun-tahun berikutnya.

Untuk mengatasi permasalahan ini akan dimodelkan dengan model regresi non *linear* yang terdiri dari variabel terikat (dependen, Y) dan variabel bebas (independen, X) (Yusnandar, 2004). Pada pemodelan model regresi dapat ditentukan apakah terdapat hubungan antara variabel x dan y, lalu bagaimana hubungan antara variabel x dan y, serta mengapa terbentuk hubungan antara variabel x dan y. Pada permasalahan serta solusi yang digunakan untuk menyelesaikan permasalahan tersebut dirasa kurang begitu optimal.

Maka dalam penelitian ini akan membahas bagaimana membangun model regresi non linear dengan algoritma genetika dalam memprediksi kebutuhan air PDAM Kota Malang. Dengan adanya model ini diharapkan dapat memberikan perkiraan kebutuhan air PDAM di Kota Malang dengan baik serta dapat dijadikan sebagai bahan pertimbangan untuk pengambilan keputusan yang akan dilakukan oleh PDAM di masa mendatang.

1.2 Rumusan Masalah

Berdasarkan uraian latar belakang diatas dapat dirumuskan permasalahan sebagai berikut :

1. Bagaimana menentukan nilai parameter algoritma genetika yang tepat untuk model regresi *non linear*?
2. Bagaimana perbandingan variabel yang terlibat pada model regresi dan model regresi yang dibentuk menggunakan algoritma genetika?

1.3 Batasan Masalah

Batasan masalah yang akan dijadikan sebagai pedoman dalam pelaksanaan penelitian ini adalah sebagai berikut :

1. Pemodelan regresi yang digunakan dalam permasalahan ini adalah persamaan regresi *non linear*.
2. Metode yang dipakai adalah Algoritma Genetika yang dimulai dari proses inialisasi, reproduksi, evaluasi dan seleksi.
3. Obyek yang digunakan adalah pemakaian air oleh konsumen PDAM.
4. Data yang di buat acuan pemakaian air oleh konsumen PDAM 6 tahun terakhir, Januari 2008-Februari 2014.

2. LANDASAN KEPUSTAKAAN

2.1 Air

Menurut peraturan menteri kesehatan republik Indonesia tentang syarat-syarat dan pengawasan kualitas air pengertian air adalah air minum, air bersih, air kolam renang, dan air pemandian umum. Air minum adalah air yang kualitasnya memenuhi syarat kesehatan dan dapat langsung diminum. Air bersih adalah air yang digunakan untuk keperluan sehari-hari yang kualitasnya memenuhi syarat kesehatan dan dapat diminum apabila telah dimasak. Kebutuhan air merupakan jumlah air yang diperlukan untuk memenuhi kebutuhan makhluk hidup diantaranya keperluan rumah tangga, pengelolaan kota, industri dan lain-lainnya (Depkes, 1990).

2.2 Regresi Non Linear

Dalam mencari persamaan garis yang paling tepat, pendugaan parameter regresi dapat digunakan untuk memilih model yang menghasilkan galat yang kecil. Analisis regresi digunakan untuk mengetahui bentuk hubungan. Sedangkan untuk keeratn hubungan dapat diketahui dengan analisis korelasi. Regresi non linear merupakan suatu metode statistika yang digunakan untuk membentuk suatu model hubungan antara variabel terikat (dependen, Y) dengan satu atau lebih variabel bebas (independen, X) (Sugiarto, 1992) yang bertujuan untuk mengestimasi serta memprediksi nilai rata-rata variabel dependen berdasarkan nilai variabel independen yang diketahui. Hasil analisis regresi berupa koefisien pada masing-masing variabel X (independen). Koefisien tadi didapat dengan cara memprediksi nilai variabel Y (dependen) dengan suatu persamaan. Perhitungan koefisien regresi bertujuan untuk memperkecil penyimpangan antara nilai aktual dan nilai estimasi variabel Y (dependen) berdasarkan data yang ada (Prajitno, 1981). Analisis regresi ini terdiri dari dua model persamaan, yaitu (Sugiarto, 1992):

1. Analisis Regresi Sederhana
$$Y = a + bX \dots\dots\dots(2-1)$$
2. Analisis Regresi Berganda

$$b_0 n + b_1 \sum X_1 + b_2 \sum X_2 = \sum Y \dots\dots(2-2)$$

$$b_0 \sum X_1 + b_1 \sum X_1^2 + b_2 \sum X_1 X_2 = \sum X_1 Y \dots\dots(2-3)$$

$$b_0 \sum X_2 + b_1 \sum X_2 X_1 + b_2 \sum X_2^2 = \sum X_2 Y \dots\dots(2-4)$$

Pada Algoritma Genetika ini terdapat beberapa tahapan-tahapan yang harus dilakukan, berikut adalah flowchart Algoritma Genetika yang ditunjukkan pada Gambar 2.1.

2.3 Algoritma Genetika

Prinsip Algoritma Genetik diambil dari teori evolusi Darwin yaitu “yang kuat adalah yang menang” yang artinya dalam proses seleksi alam hanya individu yang memiliki tingkat kebugaran tinggi yang akan bertahan hidup. Konsep dalam teori Darwin ini kemudian diadopsi menjadi algoritma komputasi untuk mencari solusi suatu permasalahan yang lebih alamiah (Zukhri, 2014).

Algoritma Genetika adalah algoritma pencarian hasil terbaik yang berdasarkan pada seleksi alam dan genetika alam dengan melakukan proses reproduksi atau perkawinan. Pada Algoritma Genetika setiap individu dalam populasi dinamakan kromosom, setiap kromosom mewakili sebuah solusi untuk masalah yang akan dihadapi. Variabel solusi dikodekan ke dalam string yang memetakan barisan gen, yang akan menjadikan karakteristik dari sebuah solusi dari suatu permasalahan (Mahmudy, 2013).

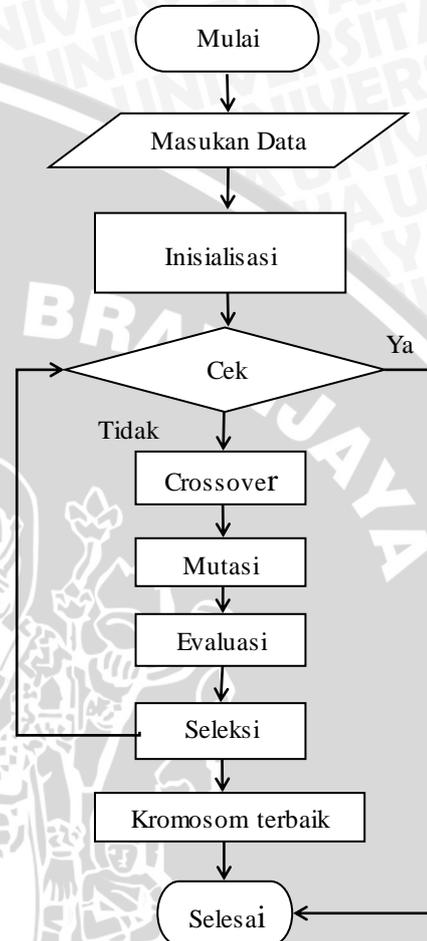
Kromosom-kromosom yang terpilih nantinya akan melakukan proses reproduksi. Pada proses reproduksi kombinasi individu-individu dilakukan secara acak. Proses reproduksi akan menghasilkan struktur gen yang nantinya akan menghasilkan suatu gen yang baru untuk proses selanjutnya yaitu proses seleksi (Mahmudy, 2013).

Pada proses reproduksi menghasilkan gen induk yang terbaik untuk setiap generasi. Dengan mengambil gen induk yang terbaik harapannya akan diperoleh gen *offspring* yang lebih baik, tetapi pada kenyataannya tidak selalu mendapatkan hasil keturunan yang lebih baik dari induknya. Kemungkinan hasil keturunannya bias lebih baik, sama baiknya, atau lebih buruk (Mahmudy, 2013)

Berikut adalah *pseudocode* proses pada Algoritma Genetika yang diawali dari inisialisasi kromosom sampai proses seleksi (Mahmudy, 2013).

```

Procedure Algoritmagenetika
begin
j = 0
inisialisasi P(j)
while (bukan kondisi berhenti) do
reproduksi I(j) dari P(j)
evaluasi P(j) dan I(j)
seleksi P(j+1) dari P(j) dan I(j)
j = j + 1
end while
end
    
```



Gambar 2.1 Tahapan Penelitian Algoritma Genetika

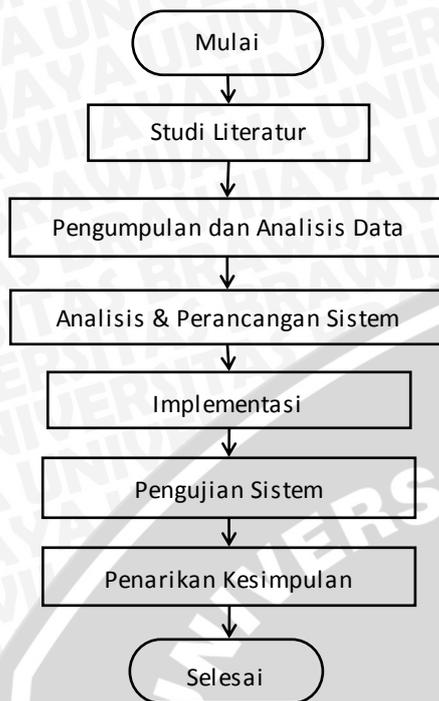
3. METODOLOGI

Pada bab ini akan dijelaskan tahapan susunan kegiatan penelitian membangun sebuah system Algoritma Genetika secara terstruktur serta perancangan system yang baik.

3.1 Tahapan Penelitian

Langkah-langkah yang dilakukan dalam penelitian adalah sebagai berikut :





Gambar 3.1 Tahapan Penelitian Algoritma Genetika

Berdasarkan bagan pada Gambar 3.1, tahapan penelitian skripsi dapat dijelaskan sebagai berikut:

1. Mengumpulkan literatur yang terkait dan mempelajari metode-metode yang akan digunakan dalam penelitian skripsi ini.
2. Melakukan analisa terhadap kebutuhan system.
3. Melakukan perancangan system.
4. Melakukan implementasi berdasarkan dari hasil analisa dan perancangan yang telah dilakukan sebelumnya.
5. Melakukan pengujian terhadap sistem, apakah sistem tersebut sudah baik atau masih membutuhkan perbaikan.

3.2 Teknik Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah:

1. Data yang digunakan adalah data distribusi dan pemakaian air PDAM Januari 2008- Februari 2014. Data tersebut didapat dari penelitain sebelumnya tentang peramalan kebutuhan air PDAM Kota Malang menggunakan *backpropagation* (Irawan, et al., 2015). Data distribusi dan pemakaian air PDAM yang diolah adalah data pemakaian air oleh pelanggan setiap bulan.
2. Parameter Algoritma Genetika yang diperlukan berupa jumlah populasi (popSize), jumlah generasi, crossover rate (cr), dan mutation rate (mr).

3.3 Algoritma yang digunakan

Pada penelitian ini menggunakan Algoritma Genetika yang seperti telah diuraikan pada kajian pustaka telah terbukti efektif digunakan untuk memberikan nilai estimasi yang lebih akurat dengan rata-rata error yang lebih kecil dibandingkan estimasi menggunakan linear biasa serta memberikan solusi optimal dalam menyelesaikan permasalahan yang memiliki banyak kemungkinan solusi.

3.4 Kebutuhan Sistem

Analisa kebutuhan berguna untuk mendapatkan apa saja yang diperlukan dalam pembuatan sistem ini. Di mana kebutuhan-kebutuhan tersebut antara lain:

1. Kebutuhan Hardware, yaitu:
 - Komputer PC Laptop Acer Aspire E14
 - RAM 2,00 GB
 - Monitor 14"
2. Kebutuhan Software, yaitu:
 - Sistem Operasi Microsoft Windows 8
 - Aplikasi Netbeans 7.3.1
 - Sistem menggunakan bahasa pemrograman Java
3. Kebutuhan data, meliputi:
 - Data distribusi dan pemakaian air PDAM Januari 2008-Februari 2014.

3.5 Pengujian Algoritma

Pengujian dilakukan dengan membandingkan hasil dari aplikasi yang dibangun menggunakan algen dengan hasil dari regresi non linear. sehingga akan ditemukan model yang baik untuk prediksi kebutuhan air PDAM.

4. PERANCANGAN

4.1 Formulasi Permasalahan

Pada penelitian prediksi pemakaian air dengan pemodelan regresi non linear menggunakan algoritma genetika yaitu berfokus untuk mencari model variabel apa saja yang digunakan. Untuk perhitungan koefisien regresi non linear menggunakan library commons-math 3.3.6.

Dimisalkan PDAM ingin memprediksi pemakaian volume air PDAM di Kota Malang dengan berdasarkan data 6 tahun terakhir. Maka untuk proses perhitungan regresi adalah sebagai berikut :

1. Pemakaian air berdasarkan periode per-tahun. Pembentukan data historis pemakaian air ditentukan berdasarkan periodenya (X). Dengan periode tertentu (X) data satu bulan membutuhkan pemakaian air sejumlah banyaknya periode lebih 1(Y). Sehingga dengan data pemakaian air 4 periode (4X), data pemakaian pada bulan ini diprediksi berdasarkan data 1 bulan, 2 bulan, 3 bulan, sampai 4 bulan kebelakang. Permasalahan tersebut akan diformulasikan pada Tabel 1.

Tabel 4.1 Data Pemakaian Air Berdasarkan Pemakaian 4 Bulan Kebelakang

No	Bln-Thn	Y	X1	X2	X3	X4
1	Mar-13	2068653	1987364	2127658	2374281	2240356
2	Apr-13	2118364	2068653	1987364	2127658	2374281
3	May-13	2092653	2118364	2068653	1987364	2127658
4	Jun-13	1936538	2092653	2118364	2068653	1987364
5	Jul-13	2287385	1936538	2092653	2118364	2068653
6	Aug-13	2178435	2287385	1936538	2092653	2118364
7	Sep-13	2278648	2178435	2287385	1936538	2092653
8	Oct-13	2317624	2278648	2178435	2287385	1936538
9	Nov-13	2144479	2317624	2278648	2178435	2287385
10	Dec-13	2198034	2144479	2317624	2278648	2178435

Pada data yang ditunjukkan pada Tabel 4.1 akan dilakukan pembentukan fungsi regresi menggunakan persamaan (2-2), (2-3), dan (2-4) yang dijabarkan pada bab sebelumnya, maka pembentukan regresi yang terbentuk dari Tabel 4.1 ditunjukkan pada persamaan (4-1).

$$Y' = 101129951.1 - 27.97564522 X1 - 33.00710196 X2 - 14.39848558 X3 - 16.60370498 X4 + 6.46142E - 06 X1^2 + 7.80195E - 06 X2^2 + 3.29401E - 06 X3^2 + 3.74523E - 06 X4^2 \dots\dots\dots(4-1)$$

4.2 Siklus Penyelesaian Pemodelan Regresi Non Linear menggunakan Algoritma Genetika

Pada subbab ini akan dibahas pemodelan regresi non linear pada pemakaian air dengan menggunakan Algoritma Genetika. Proses pemodelan regresi dengan Algoritma Genetika akan dideskripsikan pada *pseudocode* Gambar 4.1.

```

Procedure Pemodelan Regresi Linear dengan Algoritma Genetika
begin
    t = data history air
    inialisasi parameter
    pemodelan regresi
    inialisasi P(t)
    while (bukan kondisi stop) do
        reproduksi C(t) dari P(t)
            crossover
            mutasi
        evaluasi P(t) dan C(t)
        seleksi P(t+1) dari P(t) dan C(t)
        t = t + 1
    end while
    prediksi regresi non linear baru
end
    
```

Gambar 4.1 Pseudocode sistem
Sumber : (Mahmudy, 2013)

4.3 Siklus Penyelesaian Masalah Menggunakan Algoritma Genetika

Permasalahan dalam memprediksi pemakaian air menggunakan model regresi non linear dengan algoritma genetika dapat dijabarkan dengan perhitungan manual yang seperti telah dijelaskan pada Gambar 4.1. Sebagai contoh dalam perhitungan ini akan menggunakan contoh data pada Tabel 4.3. Berikut inialisasi parameter awalnya :

- Jumlah Generasi = 1
- Ukuran Populasi (popSize) = 5
- Crossover rate (cr) = 0,6
- Mutation rate (mr) = 0,4

4.3.1 Representasi Kromosom, Mencari Error, dan Perhitungan Fitness

Kromosom tersusun atas bilangan biner acak yang menyatakan orde periode data yang berpengaruh pada prediksi pemakaian air dengan model regresi non linear. Panjang kromosom dalam satu individu adalah dua kali jumlah periode data dimana x1, x2, x3, x4 adalah data aktual pemakaian air dan x1², x2², x3², x4² adalah hasil perhitungan pangkat dari data aktual x1,x2,x3,x4. Sehingga dalam menentukan model regresi non linear pada Tabel 4.2 indeks ke-1 hingga ke-8 menyatakan periode data yang akan digunakan apabila nilai biner bernilai 1 dan apabila biner bernilai 0 maka tidak digunakan. Detail pembentukan kromosom akan ditampilkan pada Tabel 4.2 dengan didapatkan model regresi non linear (x3 dan x4).

Tabel 4.2 Pembentukan Kromosom

Kromosom								Model Rregresi
1	2	3	4	5	6	7	8	
x1	X2	X3	x4	x1 ²	x2 ²	x3 ²	x4 ²	[x3],[x4]
0	0	1	1	0	0	0	0	

Lalu untuk menghitung tingkat error menggunakan persamaan :

$$\varepsilon = \frac{1}{n} \sum_{i=1}^n (Y_i - Y'_i)^2$$

Setelah didapatkan nilai error selanjutnya mencari fitness dengan persamaan :

$$fitness = \frac{10000000}{error} + \frac{1}{1 + (\alpha \times n)}$$

Sehingga diperoleh representasi kromosom seperti pada Tabel 4.3.

Tabel 4.3 Hasil Perhitungan Fitness

Kromosom								model regresi	Fitness
1	2	3	4	5	6	7	8		
x1	X2	X3	x4	x1 ²	x2 ²	x3 ²	x4 ²		
0	0	1	1	0	0	0	0	[x3],[x4]	0.930121905

4.2.2 Inialisasi Populasi Awal

Inisialisasi populasi awal merupakan pembentukan kromosom-kromosom sebanyak jumlah populasi yang telah ditemukan pada awal pembuatan parameter. Ukuran populasi atau sering disebut dengan popSize pada permasalahan kali ini dibuat sebanyak 5 populasi yang dibangkitkan secara acak yaitu [1] dan [0] yang disebut dengan insialisasi biner (Mahmudy, 2013).

2.2.3 Crossover

Metode crossover yang dipilih adalah metode *one-cut-point crossover*, yaitu dengan memilih dua induk secara acak dari populasi yang telah dibentuk, setelah itu menentukan titik sebagai cut point, gen yang berada di titik potong akan dipindah silang, dan hasilnya akan didapat anak (offspring) dari kombinasi kedua induk tersebut (Mahmudy, 2013). Pada penelitian ini populasi awal sebanyak 5 individu dengan crossover rate (cr) sebanyak 0,6 maka akan menghasilkan offspring $0,6 \times 5 = 3$. Berikut akan ditunjukkan proses *one cut point cross* pada Tabel 4.4.

Tabel 4.4 Crossover

P1	0	0	1	1	0	0	0	0
P3	0	1	1	1	1	0	0	0
C1	0	0	1	1	1	0	0	0
C2	0	1	1	1	0	0	0	0

P2	1	1	0	0	1	0	0	1
P5	1	0	0	0	0	0	0	1
C3	1	1	0	0	0	0	0	1

2.2.4 Mutasi

Proses mutasi yang digunakan adalah random mutation yaitu dengan memilih satu induk secara acak dari populasi, kemudian menentukan titik mutasi. Setelah titik itu terpilih maka dilakukan mutasi terhadap gen yang apabila nilai awal bernilai 1 maka berubah menjadi 0, begitu juga sebaliknya (Mahmudy, 2013). Dengan mutation rate sebesar 0.4 maka banyaknya offspring yang dihasilkan adalah $0,4 \times 5 = 2$. Proses mutasi akan ditunjukkan pada Tabel 4.5.

Tabel 4.5 Mutasi

P4	0	0	0	0	1	1	0	0
C4	0	0	0	1	1	1	0	0

P5	1	0	0	0	0	0	0	1
C5	1	0	0	1	0	0	0	1

2.2.5 Proses Evaluasi

Proses evaluasi merupakan proses untuk menghitung nilai fitness dari seluruh kromosom yang ada baik itu parent maupun offspring. Pada proses evaluasi kromosom yang terpilih untuk dijadikan solusi selanjutnya adalah kromosom yang memiliki nilai fitness lebih baik. Proses menghitung nilai fitness yaitu dengan menghitung nilai error terkecil dari

kemungkinan model yang terbentuk dari kromosom (Mahmudy, 2013).

2.2.6 Proses Seleksi

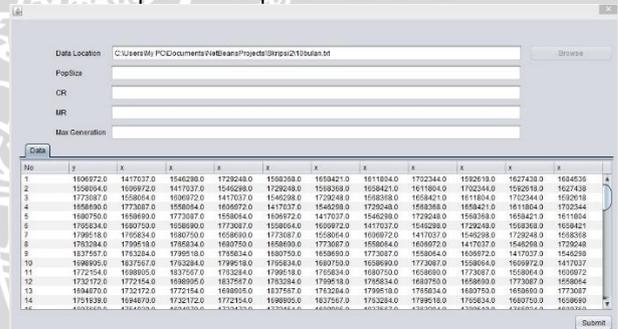
Proses seleksi adalah mengeliminasi kromosom yang memiliki fitness rendah. Tujuannya adalah untuk mempertahankan individu yang mempunyai nilai fitness tinggi agar dapat bertahan hidup pada generasi selanjutnya. Pada penelitian ini metode seleksi yang digunakan adalah elitism selection, yaitu dengan mengumpulkan semua kromosom parent dan offspring dalam satu penampungan. Individu dengan nilai fitness yang lebih besar dibanding dengan individu yang lain merupakan kromosom terbaik dalam populasi yang mampu bertahan hidup pada generasi selanjutnya. Jumlah Individu terbaik yang bertahan hidup diambil sesuai dengan popSize yang telah ditentukan sebelumnya. Metode elitism selection akan menjamin menghasilkan individu yang terbaik dan akan selalu lolos dalam seleksi (Mahmudy, 2013).

5. IMPLEMENTASI

Berikut merupakan antarmuka hasil implemetasi.

5.1 Antarmuka Data

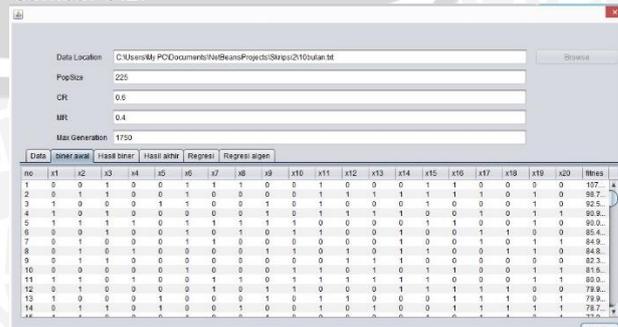
Pada tab data ini, akan ditampilkan data pemakai air pada periode yang diinputkan dan inputan dari parameter-parameter algoritma genetika. Antarmuka berada dan dapat dilihat pada Gambar 5.1.



Gambar 5. 1 Antarmuka Data

5.2 Antarmuka Biner Awal

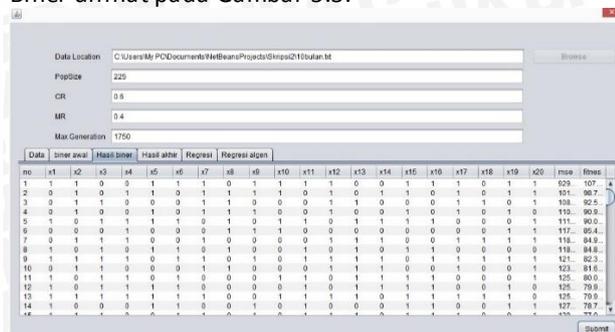
Pada tab biner awal ini, pengguna dapat melihat populasi awal dalam biner kromosm seperti pada Gambar 5.2.



Gambar 5. 2 Antarmuka Biner Awal

5.3 Antarmuka Hasil Biner

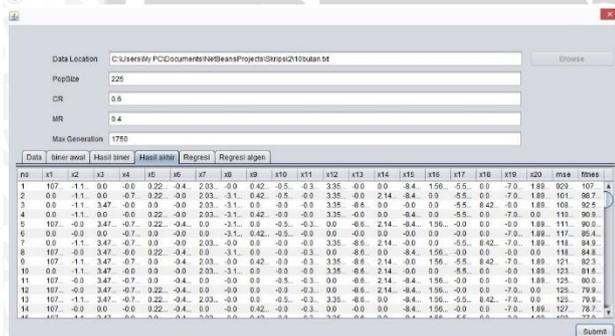
Pada tab hasil biner, akan ditampilkan biner kromosom dalam generasi terakhir. Antarmuka Hasil Biner dilihat pada Gambar 5.3.



Gambar 5. 3 Antar Muka Hasil Biner

5.4 Antarmuka Hasil Akhir

Pada tab hasil akhir, akan ditampilkan MSE dan fitness pada populasi biner kromosom dalam generasi terakhir. Antarmuka hasil akhir dilihat pada Gambar 5.4.



Gambar 5. 4 Antar Muka Hasil Akhir

5.5 Antarmuka Regresi

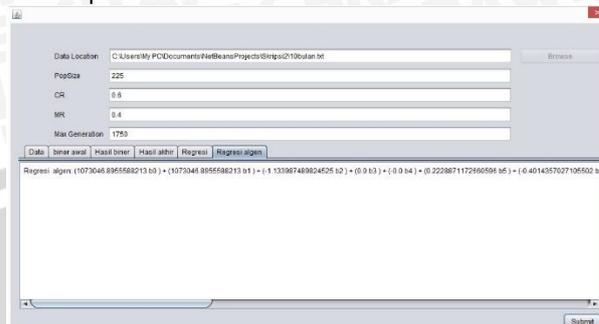
Pada tab regresi, akan ditampilkan model regresi non linear dari data periode yang diinputkan. Antarmuka regresi dilihat pada Gambar 5.5.



Gambar 5. 5 Antar Muka Regresi

5.6 Antarmuka Regresi Algen

Pada tab regresi algen, akan ditampilkan model regresi algoritma genetika. Antarmuka regresi algen dilihat pada Gambar 5.6.

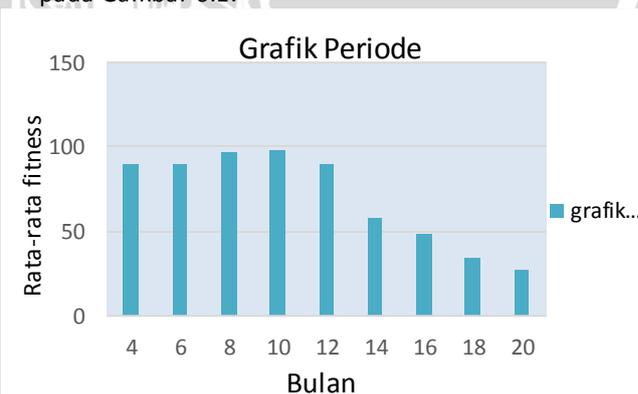


Gambar 5. 6 Antar Muka Regresi Algen

6. PENGUJIAN

6.1 Pengujian dan Analisis Banyaknya Periode Pemakaian Air (Dalam Bulan)

Pengujian periode data ini dilakukan untuk mengetahui banyaknya periode pemakaian air (dalam bulan) yang optimal berdasarkan dari hasil rata-rata fitness terbaik. Data periode pemakaian air yang digunakan dimulai dari 4 bulan kebelakang sampai 20 bulan kebelakang. Dengan menggunakan ukuran populasi yaitu 60 populasi, jumlah generasi yaitu 10 generasi, dan kombinasi cr dan mr yaitu 0.6:0.4. Untuk data yang di uji menggunakan data pemakaian air PDAM dari tahun 2008 sampai 2013. Setiap periode yang diuji akan dilakukan 10 kali percobaan dan akan diperoleh nilai rata-rata nilai fitness optimal terhadap percobaan periode pemakaian air. Hasil pengujian akan direpresentasikan pada Grafik yang ditunjukkan pada Gambar 6.1.



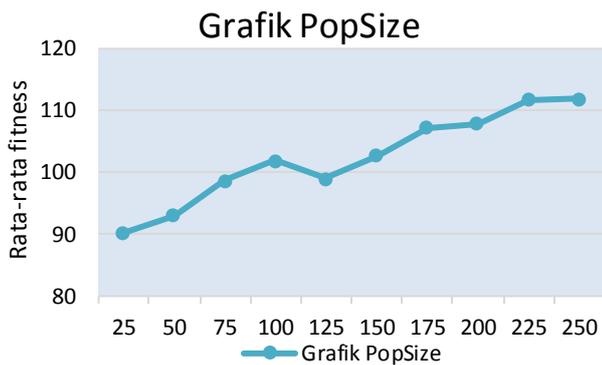
Gambar 6.1 Grafik Periode

Pada gambar 6.1 grafik uji coba banyaknya periode (dalam bulan) menghasilkan jumlah periode data terbaik adalah sebanyak 10 dengan nilai rata-rata fitness sebesar 97.739. Sedangkan nilai rata-rata fitness paling rendah ada pada periode 20 bulan dengan nilai rata-rata fitness sebesar 27.800. Hasil uji coba untuk ukuran populasi mengalami peningkatan

dimulai dari periode 4 bulan kebelakang hingga 10 bulan kebelakang dan mengalami penurunan dari periode 10 bulan hingga 20 bulan kebelakang.

6.2 Pengujian dan Analisis Ukuran Populasi

Pengujian ukuran populasi ini bertujuan untuk mengetahui ukuran populasi yang optimal dengan mengacu dari hasil rata-rata fitness terbaik pada permasalahan prediksi kebutuhan air. Data yang digunakan dalam pengujian ukuran populasi ini sebanyak 340 data pemakaian air PDAM untuk prediksi 4 bulan kebelakang mulai tahun 2008 sampai 2013. Jumlah generasi yang dipakai adalah 2000 dengan banyak populasi kelipatan 25, mulai dari 25 popSize sampai 250 popSize. Nilai *crossover rate* yang digunakan adalah 0,6 dan *mutation rate* yang digunakan adalah 0,4. Setiap ukuran populasi yang diuji akan dilakukan 10 kali percobaan dan akan diperoleh nilai rata-rata nilai fitness optimal terhadap percobaan ukuran populasi terbaik. Hasil pengujian akan direpresentasikan pada Grafik yang ditunjukkan pada Gambar 6.2.



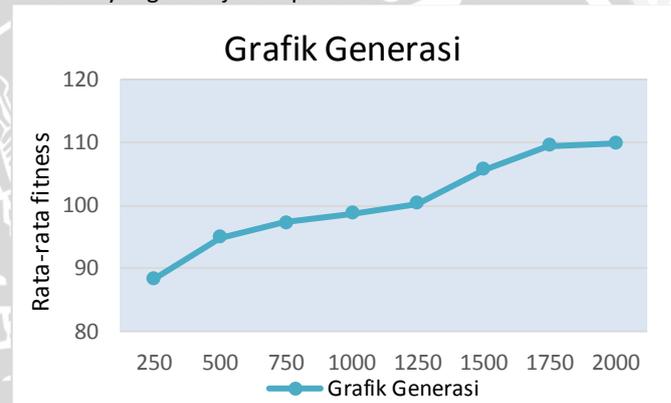
Gambar 6.2 Grafik PopSize

Pada Gambar 6.2 hasil uji coba populasi didapatkan jumlah populasi terbaik sebanyak 225 dengan rata-rata nilai fitness yaitu 111.808. Pada percobaan dengan ukuran populasi sebanyak 125 mengalami penurunan nilai rata-rata fitness meski tidak signifikan, tetapi percobaan selanjutnya menunjukkan nilai rata-rata *fitness* mengalami kenaikan dikarenakan semakin banyak populasi yang dihasilkan maka peluang untuk menghasilkan generasi dengan solusi terbaik semakin besar (Mahmudy, 2013). Pada penelitian kali ini saat batas tertentu ukuran populasi 225 dapat dilihat bahwa nilai fitness tidak mengalami kenaikan yang signifikan dan cenderung stabil. Sehingga percobaan dihentikan pada batas populasi 250 dan uji coba populasi sebanyak 225 yang dianggap sebagai solusi paling optimal. Pembatasan ukuran populasi perlu dilakukan karena semakin banyak ukuran populasi maka waktu yang dibutuhkan untuk

proses Algoritma Genetika juga semakin besar (Permatasari & Mahmudy, 2014)

6.3 Pengujian dan Analisis Ukuran Generasi

Pengujian banyaknya generasi ini dilakukan untuk mengetahui ukuran generasi yang optimal berdasarkan dari hasil rata-rata fitness terbaik. Banyak generasi yang akan dilakukan dengan kelipatan 250 mulai dari 250 generasi sampai 2500 generasi. Menggunakan hasil uji coba ukuran populasi terbaik pada pengujian sebelumnya yaitu 225 populasi, nilai *crossover rate* dan *mutation rate* yang digunakan yaitu 0.6:0.4. Untuk data yang digunakan dalam pengujian ukuran populasi ini sebanyak 340 data pemakaian kebutuhan air PDAM untuk prediksi 4 bulan kebelakang mulai tahun 2008 sampai 2013. Setiap ukuran generasi akan dilakukan 10 kali percobaan dan akan diperoleh nilai rata-rata nilai fitness optimal terhadap percobaan ukuran generasi terbaik. Hasil pengujian akan direpresentasikan pada Grafik yang ditunjukkan pada Gambar 6.3.



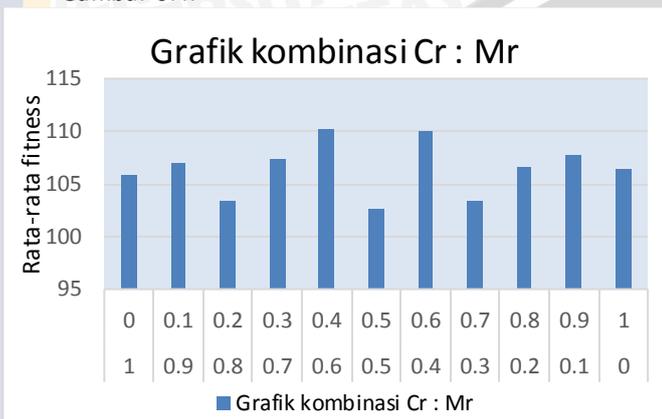
Gambar 6.3 Grafik Generasi

Pada Gambar 6.3 uji coba Generasi didapat hasil penggunaan ukuran generasi terbaik adalah dengan generasi sebesar 1750 dengan rata-rata nilai fitness 109.482. Pada saat generasi sebesar 250 dengan rata-rata fitness 88.228 menjadi generasi terendah dikarenakan algoritma genetika belum melakukan proses secara optimal (Permatasari & Mahmudy, 2014). Dalam grafik pengujian ini bahwa sesudah generasi ke 225 rata-rata fitness menjadi stabil, hal ini dikarenakan semakin banyak generasi nilai *fitness* yang dihasilkan akan cenderung sama sehingga terjadi konvergensi, apabila terus dilakukan penambahan generasi tidak akan menghasilkan solusi yang lebih baik dan cenderung akan membuang waktu (Mahmudy, 2013).

6.4 Pengujian dan Analisis Kombinasi Cr dan Mr

Pengujian kombinasi *crossover rate* (cr) dan *mutation rate* (mr) ini dilakukan untuk mengetahui kombinasi cr dan mr yang optimal berdasarkan hasil rata-rata fitness terbaik. Pada percobaan ini Kombinasi cr dan mr berjumlah 1. Dengan menggunakan ukuran

populasi terbaik yang telah diuji sebelumnya yaitu 225 populasi dan jumlah generasi terbaik sesuai dengan percobaan sebelumnya yaitu 1750 generasi. Data yang digunakan dalam pengujian kombinasi cr dan mr ini sebanyak 340 data pemakaian air PDAM untuk prediksi 10 bulan kebelakang mulai tahun 2008 sampai 2013. Setiap kombinasi cr dan mr yang diuji akan dilakukan 10 kali percobaan dan akan diperoleh nilai rata-rata nilai fitness optimal terhadap percobaan kombinasi cr dan mr terbaik. Hasil pengujian akan direpresentasikan pada Grafik yang ditunjukkan pada Gambar 6.4.



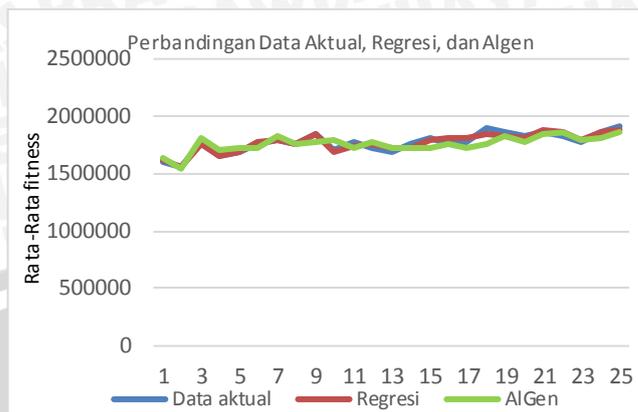
Gambar 6.4 Grafik Kombinasi Cr dan Mr

Pada gambar 6.4 grafik hasil uji coba Crossover Rate (cr) dan Mutation rate (mr) menunjukkan bahwa kombinasi cr dan mr berpengaruh terhadap rata-rata nilai fitness. Rata-rata fitness dengan hasil terbaik adalah ketika cr dan mr berbanding 0.6 : 0.4 dengan rata-rata fitness sebesar 110.2524, sedangkan rata-rata nilai fitness terendah ketika cr 1 dan mr 0 dengan rata-rata fitness sebesar 105.867.

6.5 Hasil dan Analisis Perbandingan Regesi Non Linear dan Algoritma Genetika

Pada percobaan dalam membentuk regresi non linear dengan algoritma genetika solusi terbaik yang diberikan untuk algoritma genetika yaitu pada ukuran popSize sebesar 225, generasi terbaik sebanyak 1750 generasi, kombinasi nilai crossover rate dan mutation rate terbaik yaitu 0.6 : 0.4.

Persamaan dari model regresi non linear yang terbentuk dengan algoritma genetika ini akan dibandingkan dengan persamaan regresi non linear yang juga dapat melakukan prediksi di bulan selanjutnya. Sampel data yang diambil untuk testing perbandingan menggunakan pemakaian pada periode 10 bulan. Hasil pengujian perbandingan antara model regresi non linear yang dibentuk dengan algoritma genetika dengan regresi non linear dapat dilihat pada Grafik yang ditunjukkan pada Gambar 6.5.



Gambar 6.5 Hasil dan Analisis Perbandingan Data Aktual, Regresi, dan AlGen

Pada Tabel 6.5 didapatkan nilai error MSE untuk model regresi non linear yang terbentuk dengan algoritma genetika sebesar 10379.50166. Sedangkan nilai error untuk regresi non linear sebesar 5098.945774 yang tanpa melibatkan algoritma genetika, meskipun nilai error MSE yang dihasilkan dibentuk dengan algoritma genetika lebih besar namun variabel yang terlibat lebih sedikit. Hal ini dipengaruhi oleh jumlah alfa yang digunakan pada proses perhitungan fitness. Apabila nilai alfa yang digunakan adalah 0 maka variabel yang dihasilkan oleh model regresi akan terlibat semua dan jika nilai alfa diperbesar maka jumlah variabel yang dihasilkan oleh model regresi akan semakin sedikit. Secara prinsip lebih sedikit variabel yang terlibat maka lebih baik (Stonjavic, et al., 2013). Oleh karena itu model regresi non linear yang dibentuk dengan algoritma genetika lebih baik dibandingkan dengan regresi non linear.

7. KESIMPULAN DAN SARAN

7.1 Kesimpulan

Kesimpulan yang dihasilkan dari hasil uji coba yang telah dilakukan dalam penerapan Algoritma Genetika untuk pemodelan regresi non linear dalam pemakaian air PDAM adalah sebagai berikut :

1. Parameter Algoritma Genetika berpengaruh terhadap hasil solusi yang diberikan. Nilai parameter yang besar belum tentu menghasilkan solusi yang terbaik. Dari hasil uji coba sistem pemodelan regresi untuk pemakaian air PDAM dengan menggunakan algoritma genetika, ukuran populasi terbaik adalah sebanyak 225 popSize dengan rata-rata nilai fitness 111.808, generasi yang paling mendekati solusi terbaik adalah sebanyak 1750 generasi dengan rata-rata nilai fitness 109.482, serta nilai kombinasi tertinggi

adalah *crossover rate* sama dengan 0,6 dan *mutation rate* sama dengan 0,4 dengan rata-rata nilai *fitness* 110.252.

- Jumlah variabel yang dihasilkan algoritma genetika lebih sedikit dibanding dengan jumlah variabel model regresi, namun nilai error yang dihasilkan algoritma genetika lebih besar dibanding dengan model regresi, hal ini dikarenakan jumlah nilai alfa yang digunakan. Apabila nilai alfa yang digunakan adalah 0 maka semua variabel pada model regresi akan terlibat dan jika nilai alfa diperbesar maka jumlah variabel yang terlibat akan semakin sedikit. Secara prinsip lebih sedikit variabel yang terlibat maka lebih baik. Oleh karena itu model regresi *non linear* yang dibentuk dengan algoritma genetika lebih baik dibandingkan dengan regresi *non linear*.

7.2 Saran

Berdasarkan kesimpulan diatas, ada beberapa point yang perlu diperhatikan jika pembaca ingin melakukan penelitian lebih lanjut.

- Titik optimal pada uji coba populasi, uji coba generasi, dan uji coba kombinasi *cr* dan *mr* terjebak dalam kisaran yang sama. Untuk memperluas area pencarian algoritma genetika dalam kasus prediksi pemakaian air PDAM dengan pemodelan regresi *non linear* dapat diterapkan metode *crossover*, mutasi, dan seleksi yang lain.
- Semakin besar periode yang digunakan, prediksi pemakaian air akan menjauhi data konsumsi pemakaian air aktual. Agar mendekati data aktual bisa dilakukan pengujian periode menggunakan data campuran, harian, ataupun mingguan.
- Saran untuk penelitian selanjutnya agar nilai *error* yang dihasilkan oleh sistem mendekati minimum bisa dilakukan pengujian perbandingan terhadap variabel-variabel yang terlibat di dalam model regresi.

Daftar Pustaka

- Depkes, R., 1990. *Peraturan Menteri Kesehatan RI No 416/Menkes/Per/IX/1990 Tentang Syarat-syarat dan Pengawasan Kualitas Air*. Jakarta: Jakarta.
- Irawan, Dewi, C. & holissodin, I., 2015. Implementasi Metode Jaringan Syaraf Tiruan Backpropagation Pada Peramalan Kebutuhan Air PDAM Kota Malang. *DORO repository jurnal mahasiswa PTIIK*, Volume 6, p. 11.
- Mahmudy, W. F., 2013. *Algoritma Evolusi*. Malang: Universitas Brawijaya.
- Permatasari, A. I. & Mahmudy, W. F., 2014. Pemodelan regresi linear dalam konsumsi Kwh listrik di Kota

Batu menggunakan algoritma genetika.. *DORO : Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya*, Volume 4, pp. 6,18.

- Prajitno, D., 1981. *Analisis Regresi dan Kolerasi untuk Penelitian Pertanian*. 1 ed. Yogyakarta: Liberty Yogyakarta.
- Stonjavic, B., Milivojevic, M. & Ivanovic, M., 2013. Adaptive System For Dam Behavior Modeling Based In Linear Regression And Genetic Algorithms. *IEEE*, Volume 65, pp. 182-190..
- Sugiarto, 1992. *Tahap Awal + Aplikasi Analisis Regresi*. Yogtakarta: Andi Offset.
- Sugiharto, 1987. *Dasar-dasar Pengelolaan Air Limbah*. Jakarta: Universitas Indonesia.
- Syaiful & Achmad, 2014. *Segudang Prestasi dan Layanan PDAM*. s.l.:mediacenter.malangkota.go.id/2012/12.
- Yusnandar, M., 2004. Aplikasi Analisis Regresi Non Linear Model Kuadratik Terhadap Produksi Susu Kambing Peranakan Etawah (PE) selama 90 Hari Pertama Laktasi. *Informatika Pertanian*, Volume 13, pp. 736-743.
- Zukhri, Z., 2014. *Algoritma Genetika Metode Komputasi Evolusioner untuk Menyelesaikan Masalah Optimasi*. 1 ed. Yogyakarta: Andi.