

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

Vitara Nindya Putri Hasan¹, Wayan Firdaus Mahmudy², Mohammad Zoqi Sarwani³

^{1,2,2}Fakultas Ilmu Komputer/Informatika Universitas Brawijaya, Malang
Email: ¹vitaranindyaph@gmail.com, ²wayanfm@ub.ac.id, ³zaki_swr@yahoo.com

(Naskah masuk: 18 Februari 2016, diterima untuk diterbitkan: 17 Maret 2016)

Abstrak

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 dimodelkan dengan persamaan regresi non linear yang terdiri dari variabel bebas (X) dan variabel terikat (Y). Algoritma Genetika digunakan untuk memilih variabel mana saja yang perlu dilibatkan dalam persamaan regresi. Proses reproduksi menggunakan one-point-crossover dan random mutation, untuk proses seleksinya menggunakan model elitism selection. Dari uji coba 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 model regresi dengan melibatkan sesedikit mungkin variable bebas dan *mean square error* (MSE) terkecil..

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

Abstract

Along with the increasing population in Malang the consumer water consumption from PDAM also increase. 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 get a shortage of water. Therefore it is necessary to estimate for 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). Genetic Algorithm is applied to determine which variables are involved in the regression model. The reproduction process uses one-point-crossover and random mutation, for the selection process uses a elitism selection models. The numerical experiment obtains the best population size is 225, the best generation as much as 1750 generation, combination of cr : mr is 06 : 0.4 with the highest fitness value is 107.997. The final result is a regression model that involves less independent variabls with minimum mean square error (MSE).

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

1. PENDAHULUAN

Semua makhluk hidup pasti memerlukan air. Air digunakan untuk berbagai macam keperluan mulai dari makan, minum, mandi, cuci, bekerja, dan berbagai kebutuhan lainnya. Air bersih adalah air yang telah memenuhi syarat kesehatan dan dapat digunakan untuk keperluan sehari-hari dalam melakukan berbagai aktivitas (Kesehatan, 1990). Badan yang menangani air bersih di Indonesia adalah Perusahaan Daerah Air Minum (PDAM). PDAM didirikan di setiap kota di Indonesia sebagai badan yang melayani kebutuhan masyarakat akan air bersih, salah satunya yang ada di kota Malang. Kota Malang adalah kota yang selalu mengalami peningkatan jumlah penduduk setiap tahunnya (Pemerintahan,

2015). Hal ini disebabkan oleh beberapa faktor diantaranya lokasi Kota Malang yang berada di daratan tinggi sehingga memiliki iklim sejuk, selain itu Malang memiliki fasilitas dan kualitas pendidikan yang memadai sehingga Malang dijuluki sebagai kota pendidikan (Indonesia, 2015). Faktor inilah yang mengakibatkan banyak yang memilih transmigrasi dan menetap tinggal di Malang. Seiring dengan meningkatnya jumlah penduduk di Kota Malang maka meningkat pula kebutuhan konsumen akan air bersih dari PDAM (Syaiful and Achmad, 2012).

Berdasarkan uraian ini maka PDAM harus meningkatkan kebutuhan air bersih setiap tahunnya guna memenuhi kebutuhan air bersih di Kota Malang. Tetapi 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 berikutnya agar tidak terjadi pemborosan atau kekurangan persediaan air untuk masyarakat.

Untuk mengatasi permasalahan ini diperlukan suatu analisis kebutuhan air konsumen PDAM dengan melakukan suatu pemodelan dengan metode regresi non linear menggunakan algoritma genetika. Regresi non linear adalah hubungan antara variable terikat (dependen, Y) dan variabel bebas (independent, X)(Yusnandar, 2004). Pada penelitian yang pernah dilakukan oleh Stonjavic, et al., (2013) yang bertujuan untuk menentukan model regresi linear (MLR) menggunakan algoritma genetika dalam sistem adaptif untuk bendungan Bocac di sungai Vrbas, Republic of Srpska. Hasil analisis menunjukkan bahwa, dalam periode waktu dimana satu atau lebih sensor yang aktif menunjukkan prediksi lebih baik daripada MLR asli yang mengasumsikan bahwa semua sensor selalu aktif. Berdasarkan penjelasan di atas bahwa model regresi yang dihasilkan melalui pendekatan algoritma genetika memberikan hasil yang lebih baik dibanding dengan model regresi. Sehingga penulis memilih model regresi non linear dengan menggunakan algoritma genetika.

2. PENELITIAN TERKAIT

Penelitian yang menggunakan metode regresi non linear pernah dilakukan oleh Yusnandar (2014) untuk mengetahui sejauh mana aplikasi analisis regresi non linear model kuadratik terhadap produksi susu kambing PE selama 90 hari pertama laktasi. Dari hasil pengujian diperoleh persamaan regresi non linear model polynomial kuadratik pada analisis data produksi susu kambing PE selama 90 hari pertama laktasi dan tingkat maksimum produksi susu diketahui pada hari ke 20 setelah laktasi.

Pada kasus lain terdapat penelitian yang menggunakan pemodelan regresi linear dan algoritma genetika dalam konsumsi KWh listrik di Kota Batu yang dilakukan oleh Permatasari dan Mahmudy (2014). Pada koefisien yang dihasilkan oleh sistem menghasilkan nilai error yang lebih rendah dibandingkan dengan persamaan regresi.

Pada penelitian yang pernah dilakukan oleh Stonjavic, et al., (2013) yang bertujuan untuk menentukan model regresi linear (MLR) menggunakan algoritma genetika dalam sistem adaptif untuk bendungan Bocac di sungai Vrbas, Republic of Srpska. Hasil analisis menunjukkan bahwa, dalam periode waktu dimana satu atau lebih sensor yang aktif menunjukkan prediksi lebih baik daripada MLR asli yang mengasumsikan bahwa semua sensor selalu aktif.

3. METODE

3.1. Data dan Variabel Penelitian

Data yang digunakan pada penelitian ini berupa data time series bulanan pemakaian air PDAM mulai dari Januari 2008 sampai dengan Desember 2013. data ini diambil dari penelitian sebelumnya yang dilakukan oleh Irawan, et all (2015).

Model yang dibangun pada metode regresi non linear ini akan digunakan untuk menyelesaikan peramalan pemakaian air PDAM. Pada model non linear dalam parameternya bersifat kuadratik dan kubik dengan kurva yang dihasilkan membentuk garis lengkung (Yusnandar, 2004) selain itu, regresi non linear juga mempunyai banyak kurva yang bisa di bentuk(Sugiarto, 1992). Model regresi non linear terdapat variable terikat atau tidak bebas (Y) dan variable bebas (X). Nilai pemakaian air PDAM yang akan diprediksi dijadikan sebagai variable (Y) dan nilai pemakaian air PDAM pada t-1,t-2,t-3 dan t-4 dijadikan sebagai variable (X), dimana t menunjukkan bulan.

Tabel 1. Variabel pada regresi non linear

Variable	Deskripsi
t	Nilai yang akan diprediksi
t-1	Nilai pemakaian air PDAM pada bulan sebelumnya
t-2	Nilai pemakaian air PDAM pada 2 bulan sebelumnya
t-3	Nilai pemakaian air PDAM pada 3 bulan sebelumnya
t-4	Nilai pemakaian air PDAM pada 4 bulan sebelumnya
(t-1) ²	Nilai kuadrat pemakaian air PDAM pada bulan sebelumnya
(t-2) ²	Nilai kuadrat pemakaian air PDAM pada 2 bulan sebelumnya
(t-3) ²	Nilai kuadrat pemakaian air PDAM pada 3 bulan sebelumnya
(t-4) ²	Nilai kuadrat pemakaian air PDAM pada 4 bulan sebelumnya

Pada persamaan (1) akan ditunjukkan persamaan regresi non linear yang digunakan untuk penelitian ini.

$$t = a + b_1(t-1) + b_2(t-2) + b_3(t-3) + b_4(t-4) + b_5(t-1)^2 + b_6(t-2)^2 + b_7(t-3)^2 + b_8(t-4)^2 \quad (1)$$

Hasil yang didapatkan dari hasil peramalan akan menggunakan metode regresi non linear akan dibandingkan dengan nilai sebenarnya untuk mengetahui seberapa besar perbedaannya. Metode yang digunakan untuk mengukurnya adalah Mean Square Error (MSE). Semakin besar nilai MSE maka semakin besar perbedaan hasil peramalan dengan data sebenarnya. Untuk menghitung MSE dapat menggunakan persamaan 2 berikut ini :

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

Keterangan :

MSE = merupakan Nilai error MSE

n = merupakan jumlah data yang diprediksi

Y_i = merupakan nilai sebenarnya.

Y'I = merupakan nilai hasil prediksi
 I = indeks

3.2. Struktur Algoritma Genetika

Algoritma Genetika merupakan 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, 2015).

Algoritma genetika yang terdapat pada penelitian ini digunakan untuk mengoptimasi persamaan regresi non linear untuk mendapatkan solusi yang terbaik. Algoritma genetika akan bekerja untuk menentukan jumlah parameter yang tepat pada persamaan regresi non linear.

3.2.1. Representasi Kromosom

Kromosom tersusun dari sejumlah gen yang menjadi calon solusi untuk masalah yang akan dihadapi Representasi kromosom terdiri dari berbagai bentuk yaitu representasi integer, biner, permutasi, dan real code(Mahmudy., 2015)

Representasi yang digunakan pada permasalahan dengan menggunakan pemodelan regresi non linear ini adalah representasi bilangan biner. Setiap variabel diwakilkan oleh angka (0) atau (1), jika (0) maka variabel tersebut tidak digunakan dan jika (1) maka variabel tersebut digunakan (Stonjavic et al., 2013)

Pada penelitian ini 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 x_1, x_2, x_3, x_4 adalah data aktual pemakaian air mulai dari (t-1) sampai dengan (t-4) dan $x_1^2, x_2^2, x_3^2, x_4^2$ adalah hasil perhitungan pangkat dari data aktual x_1, x_2, x_3, x_4 .

Dalam menentukan model regresi non linear pada Tabel 2 indeks ke-1 hingga ke-8 menyatakan periode data yang akan digunakan yakni periode pemakaian air pada saat (t-1) sampai dengan (t-4) dan hasil kuadratnya. Apabila nilai biner pada setiap gen bernilai 1 berarti data pada periode tersebut dipakai. dan apabila biner bernilai 0 maka data pemakaian air pada periode tersebut tidak digunakan(Stonjavic et al., 2013).

Tabel 2 Pembentukan Kromosom

Kromosom							
1	2	3	4	5	6	7	8
x_2	x_2	x_2	x_4	x_1^2	x_2^2	x_3^2	x_4^2

0	0	1	1	0	0	0	0
---	---	---	---	---	---	---	---

3.2.2. Crossover

Pada kasus ini metode crossover yang dipilih menggunakan metode one-point-crossover yaitu proses pemilihan dua induk secara acak dari populasi dan menghasilkan offspring (anak) dari kedua induk tersebut (Mahmudy et al., 2013b)., 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(Stonjavic et al., 2013). Pada penelitian ini populasi awal sebanyak 5 individu dengan crossover rate (cr) sebanyak 0,6 maka akan dihasilkan offspring $0,6 \times 5 = 3$. Berikut akan ditunjukkan proses one cut point cross pada Gambar 1.

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

Gambar 1. Crossover

3.2.3 Mutasi

Proses mutasi yang digunakan adalah random mutation yaitu sebuah proses acak dimana memilih salah satu gen didalam kromosom yang ditentukan untuk kemudian dirubah nilainya dari 0 menjadi 1 dan sebaliknya(Mahmudy et al., 2013b) (Stonjavic et al., 2013). Dengan mutation rate sebesar 0.4 maka banyaknya offspring yang dihasilkan adalah $0.4 \times 5 = 2$. Proses mutasi akan ditunjukkan pada Gambar 2.

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

Gambar 2. Mutasi

3.2.4. Fungsi Fitness

Fungsi fitness digunakan untuk mengukur kebaikan solusi yang dibawa oleh suatu individu. Individu terbaik dengan nilai fitness yang tinggi cenderung menghasilkan solusi yang terbaik. Fungsi fitness yang digunakan untuk masalah pemodelan regresi non linear dalam kebutuhan air ditunjukkan pada Persamaan (2.25).

$$fitness = \frac{c}{error} + \frac{1}{1+(\alpha x n)} \quad (2-3)$$

Dimana :

- c = constanta
- n = Jumlah kromosom bernilai 1
- error = Nilai MSE yang dihasilkan
- α = 0,1

3.2.5. Seleksi

Proses selanjutnya adalah seleksi yaitu mengeliminasi kromosom yang memiliki fitness rendah. Tujuannya adalah untuk mempertahankan individu yang mempunyai nilai fitness tinggi agar dapat bertahan hidup pada generasi selanjutnya. Pada proses seleksi ini dilakukan dengan mengumpulkan semua kromosom parent dan offspring dalam satu penampungan (Stonjavic et al., 2013).

Pada penelitian ini metode yang digunakan adalah *elitism selection* yaitu memilih individu berdasarkan *popSize* terbaik dengan mengurutkan individu-individu dari nilai fitness tertinggi sampai terendah (Mahmudy et al., 2013a) 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, 2013b).

Hasil akhir berupa hasil seleksi individu-individu terbaik sejumlah populasi awal untuk dilanjutkan ke generasi selanjutnya. Hasil seleksi individu terbaik ditampilkan pada Gambar 3

P	P(t-1)		
P1	P3	[x2],[x3],[x4],[x1 ²]	1.023961717
P2	C3	[x1],[x2],[x4 ²]	0.970993768
P3	P1	[x3],[x4]	0.930121905
P4	C5	[x1],[x4],[x4 ²]	0.919619409
P5	P2	[x1],[x2],[x1 ²],[x4 ²]	0.910642509

Gambar 3. Seleksi

Setelah dilakukan proses seleksi, maka dilanjutkan dengan memilih kromosom terbaik yang dapat ditentukan berdasarkan nilai fitness terbesar. Pada Gambar 3, kromosom yang memiliki fitness terbesar adalah parent ke-1

4. HASIL DAN PEMBAHASAN

Pada penelitian ini, akan dilakukan beberapa uji coba untuk mendapatkan solusi terbaik. Uji coba dilakukan dengan mencari parameter algoritma genetika yang terbaik dan menguji jumlah periode yang tepat.

4.1. Pengujian Jumlah Periode.

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. Hasil percobaan banyaknya periode ditunjukkan pada Tabel 2.

Tabel.2 Pengujian Banyaknya Periode

No	Banyak Periode	Rata-rata nilai fitness
1	4	89.346
2	6	89.545
3	8	97.082
4	10	97.739
5	12	90.082
6	14	58.161
7	16	49.053
8	18	34.299
9	20	27.800

Pada Tabel 2 didapatkan jumlah periode yang terbaik adalah terdapat pada jumlah periode 10 dan dengan nilai rata-rata *fitness* 97.729.

Asal	Asal	Model	Fitness
------	------	-------	---------

4.2. Pengujian 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. Untuk hasil percobaan ukuran populasi ditunjukkan pada Tabel 3.

Tabel 3 Skenario Pengujian Ukuran Populasi (popSize)

Banyak PopSize	Rata-rata nilai fitness	rata-rata waktu (detik)
25	90.250	40.2
50	93.041	75.8
75	98.705	86.8
100	101.991	93
125	98.939	101.4
150	102.665	119.2
175	107.265	143.1
200	107.934	162.9
225	111.808	177.6
250	111.959	198.7

Dari 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, 2013b).

4.3. Pengujian Banyaknya 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 2000 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 percobaan. Hasil percobaan banyaknya generasi ditunjukkan pada Tabel 4.

Tabel 4 Pengujian Banyaknya Generasi

No	Jumlah Generasi	Rata-rata nilai fitness
1	250	88.228
2	500	94.943
3	750	97.222
4	1000	98.745
5	1250	100.339
6	1500	105.693
7	1750	109.482
8	2000	109.840

Dari hasil uji coba generasi didapatkan penggunaan ukuran generasi terbaik adalah 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 and Mahmudy, 2014).

4.4. Pengujian Kombinasi Crossover Rate dan Mutation Rate

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. Hasil percobaan kombinasi cr dan mr ditunjukkan pada Tabel 4.

Tabel 5 Hasil Uji Coba Kombinasi Crossover Rate (cr) dan Mutation Rate (mr)

Kombinasi	Rata-Rata Nilai Fitness
-----------	-------------------------

cr	Mr		Data Aktual	Prediksi dengan Regresi	Prediksi dengan Algen	Error Regresi	Error GA
						$(Y_i - Y'_i)^2$	
1	0	105.867	1606972	1613939.056	1628993.928	48539870.46	484965316
0.9	0.1	107.013	1558064	1559682.551	1543730.829	2619708.318	205439792
0.8	0.2	103.302	1773087	1765555.022	1804400.275	56730696.51	980521185.3
0.7	0.3	107.290	1658690	1652439.504	1699838.989	39068694.84	1693239282
0.6	0.4	110.252	1680750	1683713.16	1730606.593	8780319.056	2485679905
0.5	0.5	102.613	1765834	1771281.536	1728341.631	29675644.1	1405677740
0.4	0.6	109.939	1799518	1798813.23	1825805.532	496700.1629	691034331.8
0.3	0.7	103.365	1763284	1757549.303	1762365.357	32886748.63	843905.5054
0.2	0.8	106.688	1837567	1852024.338	1773967.301	209014625	4044921768
0.1	0.9	107.748	1698905	1696006.899	1797882.307	8398987.44	9796507213
0	1	106.505	1772154	1735972.362	1717332.663	1309110953	3005378952
			1732172	1759987.911	1773294.318	773724894.5	1691045051
			1694870	1717093.846	1717175.906	493899332.5	497553459.4
			1751939	1725442.68	1720383.558	702054995.6	995745931.4
			1807659	1793975.072	1717269.51	187249884.2	8170259929
			1752567	1817925.311	1763774.637	4271708789	125611116.3
			1773785	1806373.278	1732648.582	1061995843	1692204889
			1901589	1840326.386	1765789.678	3753107883	18441455811
			1858807	1834457.563	1822500.007	592895084.4	1318197719
			1825106	1807663.543	1779285.395	304239312.9	2099527821
			1869328	1879971.244	1854179.068	113278643.3	229490148.3
			1820043	1854904.558	1859285.51	1215328239	1539974627
			1780592	1786086.47	1789475.243	30189201.26	78912002.28
			1861409	1864996.847	1809906.88	12872647.42	2652468368
			1913115	1881624.33	1858277.714	991662309.1	3007127947
					Sumerror	16249530007	67333784210
					MSE	5098.945774	10379.50166
					Fitness	196.6745269	96.89929465

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.

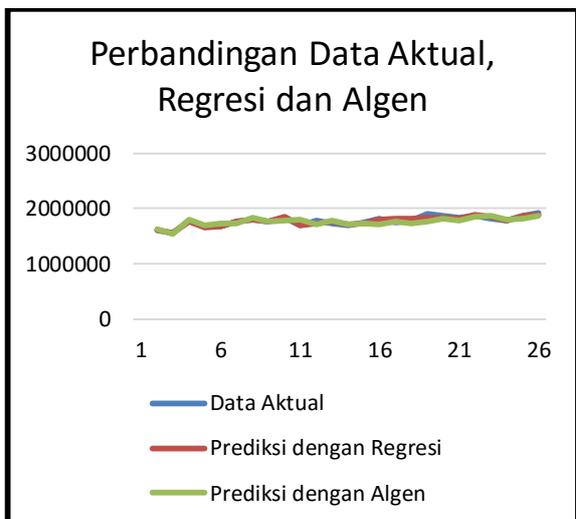
Persamaan dari model regresi non linear yang terbentuk dengan algoritma genetika ini akan dibandingkan dengan persamaan regresi non linear yang juga dapat melakukan prediksi di bulan selanjutnya.

Sampel data yang diambil untuk testing perbandingan menggunakan pemakaian selama 25 bulan. Perbandingan antara data aktual (yang sebenarnya) dengan hasil prediksi model regresi non linear dengan melibatkan semua variabel serta model regresi non linear yang dibentuk dengan algoritma genetika dapat dilihat pada Tabel 6.

Pada Tabel 6 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

Tabel 6. Detail Error Regresi dan Error Algoritma Genetika

genetika lebih baik dibandingkan dengan regresi non linear.



Gambar 4. Perbandingan Data Aktual, Regresi dan Hasil Prediksi

5. PENUTUP

Dari hasil uji coba yang dilakukan pada kasus prediksi pemakaian air PDAM menggunakan algoritma genetika untuk pemodelan regresi non linear mampu memberikan hasil yang baik. Parameter yang terdapat pada algoritma genetika memiliki pengaruh yang besar terhadap hasil yang didapatkan. Uji coba untuk mendapatkan jumlah populasi yang terbaik adalah pada saat jumlah populasinya sebesar 225 dengan nilai rata-rata fitnessnya adalah 111,808. Sedangkan jumlah generasi terbaik adalah sebanyak 1750 generasi dengan nilai fitness 109,842. Dan kombinasi terbaik untuk nilai crossover rate dan mutation rate adalah 0,6 dan 0,4 dimana nilai fitness yang didapatkan sebesar 110,252. Akan tetapi pada penelitian ini nilai error yang didapatkan algoritma genetika dengan menggunakan parameter terbaik lebih besar dibandingkan nilai error yang didapatkan regresi non linear. Hal ini dikarenakan model regresi non linear yang dibentuk dengan algoritma genetika menggunakan variable yang lebih sedikit dibandingkan dengan regresi non linear.

Pada penelitian selanjutnya hibridisasi algoritma genetika bisa dilakukan untuk menghasilkan model regresi yang lebih baik.

6. DAFTAR PUSTAKA

- INDONESIA, P., 2015. 7 Alasan Ini Akan Memperkuat Kamu Untuk Kuliah di Kota Malang. [Online] Available at: www.pendidikanindonesia.com [Accessed 21 Januari 2016].
- IRAWAN, DEWI, C. & CHOLISSODIN, I., 2015. Implementasi Metode Jaringan Syaraf Tiruan Backpropagation Pada Peramalan

Kebutuhan Air PDAM Kota Malang. DORO : Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya, Volume 6, p. 11.

- KESEHATAN, D. 1990. Peraturan Menteri Kesehatan RI No 416/Menkes/Per/IX/1990 Tentang Syarat-syarat dan Pengawasan Kualitas Air. Jakarta: Departemen Kesehatan RI.
- MAHMUDY, W. F. 2015. Dasar-Dasar Algoritma Evolusi, Program Teknologi Informasi dan Ilmu Komputer, Universitas Brawijaya, Malang.
- MAHMUDY, W. F., MARIAN, R. M. & LUONG, L. H. S. 2013a. Modeling and optimization of part type selection and loading problems in flexible manufacturing system using real coded genetic algorithms. *International Journal of Electrical, Computer, Electronics and Communication Engineering*, 7, 251-260.
- MAHMUDY, W. F., MARIAN, R. M. & LUONG, L. H. S. 2013b. Real coded genetic algorithms for solving flexible job-shop scheduling problem – Part I: modeling. *Advanced Materials Research*, 701, 359-363.
- PEMERINTAHAN, B. 2015. LKPJ Tahun Anggaran 2014.
- 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*, 5.
- STONJAVIC, B., MILIVOJEVIC, M. & IVANOVIC, M. 2013. Adaptive System For Dam Behavior Modeling Based In Linear Regression And Genetic Algorithms. *IEEE*, 65, 182-190.
- SYAIFUL & ACHMAD 2012. Segudang Prestasi dan Layanan PDAM.
- YUSNANDAR 2004. Aplikasi Analisis Regresi Non Linear Model Kuadratik Terhadap Produksi Susu Kambing Peranakan Etawah (PE) Selama 90 Hari Pertama Laktasi. *Informatika Pertanian*, 13, 736-743.