Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)
Vol. x, No. x, April 2016, hlm. x-x

Optimasi Model Regresi Menggunakan Algoritma Genetika untuk Memprediksi Jumlah Mahasiswa yang Mengambil Mata Kuliah

Sutrisno[1], Wayan Firdaus Mahmudy[2]

Program Studi Informatika/ Ilmu Komputer
Fakultas Ilmu Komputer
Universitas Brawijaya, Malang 65145, Indonesia
Email : sutris.ub[at]gmail.com[1], wayanfm[at]ub.ac.id[2]

Abstrak

Pembuatan jadwal mata kuliah dilakukan oleh pihak akademik pada setiap awal semester. Saat awal semester itulah mahasiswa akan memprogram kartu rencana studi dengan menempuh perkuliahan sesuai dengan jumlah SKS maksimal sesuai indeks prestasi yang diperoleh pada semester sebelumnya. Pembuatan jadwal membutuhkan tenaga, pikiran dan waktu yang banyak karena harus mempertimbangkan ketersediaan ruang, waktu, dosen dan perkiraan jumlah mahasiswa yang mengambil mata kuliah. Seringkali pihak akademik mengalami kesulitan memprediksi jumlah mahasiswa yang akan menempuh mata kuliah tertentu, khususnya memprediksi mahasiswa yang sebelumnya telah menempuh mata kuliah tersebut namun masih memperoleh nilai yang kurang memuaskan apakah akan memperbaiki nilainya atau tidak. Sehingga sering terjadi kekurangan kelas, karena mata kuliah yang sedang dibuka banyak diminati. Dilain sisi ada pula kelas yang masih banyak kosong, dimana mahasiswa kurang minat terhadap kelas yang dibuka. Sehingga diperlukan sebuah sistem yang mampu membantu pihak akademik dalam menentukan perkiraan kelas yang akan dibuka. Dalam optimasi menggunakan algoritma genetika, representasi kromosom yang digunakan adalah binary, proses crossover yang digunakan adalah one cut point crossover, proses mutasi menggunakan random mutation, serta seleksi menggunakan metode seleksi elitism. Hasil penelitian memberikan nilai parameter terbaik yaitu ukuran populasi 600, jumlah generasi 400, dengan kombinasi cr : mr berturut-turut adalah 0,4 : 0,6. Dalam pengujian yang dilakukan peneliti, hasil terbaik dari optimasi ini dibuktikan dengan nilai MSE terkecil 5,64 dengan kromosom terpilih 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 . Sehingga dapat disimpulkan model regresi linear berganda yang dibangun menggunakan algoritma genetika mampu untuk memprediksi jumlah mahasiswa pengambil mata kuliah dengan akurasi yang baik dan dengan usaha yang lebih mudah dibandingkan dengan model prediksi lainnya.

Kata kunci: *optimasi, algoritma genetika, prediksi mahasiswa mengulang*

Abstract

Subject schedule making is conducted by academic employees in every beginning of the semester. In the beginning of the semester, the students will programmed the study program by taking the lecture based on the number of maximum SKS based on the achievement index which was gained in the previous semester. Schedule making needs power, thought, and much time since it needs to consider the availability of classroom, time, lecturer and the number of student's estimation who will take the subjects. Academic employees often find difficulties in predicting the number of those who will take the certain subjects, but they have not got the satisfying score yet, whether they will revise the score or not. Thus, it causes the lack of classroom, since the subject which is being opened has high proclivity of students in choosing the subject. On the other hand, there is opened classroom which still has high capacity because the number of the students who take the subject is lesser. Thus, it needs a system which can help the academic employees in deciding classrooms estimation which are going to be opened. In this genetic algorithm optimization, chromosomes representation using binary, crossover process using one cut point crossoer, mutation peocess using elitism method. The results of this study provide the best parameter value that is population size is 600, generation size is 400, combination cr:mr is 0,4 : 0,6. The best result of this optimization is evidence by the smallest MSE value of 5,64 with selected chromosome 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1. It can be concluded linear regression models were constructed using genetic algorithm were able to predict the number of students estioation have a good accuracy than another prediction models.

Keywords: *optimization, genetics algorithm, the repeating student's estimation.*

1. PENDAHULUAN

1.1 Latar Belakang

Pembuatan jadwal perkuliahan pada suatu jurusan atau fakultas dilakukan rutin setiap pergantian semester. Pembuatan jadwal mata kuliah ini merupakan hal yang penting dalam proses perkuliahan, proses penyusunan jadwal mata kuliah harus disusun dengan baik. Salah satu kriteria jadwal

mata kuliah yang baik memiliki distribusi mata kuliah yang merata dan jumlah mahasiswa yang sesuai dengan kapasitas kelas untuk menghasilkan suasana perkuliahan yang kondusif.

Proses penyusunan jadwal perkuliahan bukanlah sebuah pekerjaan yang mudah. Ada banyak aspek dan kendala yang mempengaruhi dalam penyusunan jadwal mata kuliah yaitu dosen, mahasiswa, mata kuliah, ruang perkuliahan, slot waktu dan prediksi jumlah mahasiswa yang memiliki minat untuk mengambil mata kuliah tertentu. Setiap aspek tersebut memiliki keadaan yang dapat menjadi masalah dan konflik dalam proses perkuliahan. Misalnya masalah yang dihadapi adalah jumlah kelas yang dibuka tidak sesuai dengan jumlah permintaan dari mahasiswa, sehingga ada mata kuliah tertentu mengalami kekosongan karena sedikitnya minat mahasiswa yang mengambil mata kuliah tersebut. Dilain sisi ada mata kuliah yang sangat diminati oleh mahasiswa sehingga dengan sangat cepat kelas terpenuhi. Hal ini mengakibatkan tidak sesuainya persediaan jumlah kelas yang dibuka dengan jumlah mahasiswa yang akan mengambil mata kuliah.

Menentukan jumlah perkiraan atau prediksi mahasiswa yang akan mengambil suatu mata kuliah pada jurusan Informatika harus dilakukan pada setiap pembuatan jadwal atau saat pergantian semester. Hal ini penting untuk menentukan berapa kelas yang akan dibuka dan nantinya akan mempengaruhi pada proses pembuatan jadwal perkuliahan sesuai dengan daya tampung, jumlah peminat, ketersediaan dosen.

Menentukan jumlah perkiraan atau prediksi mahasiswa yang akan mengambil suatu mata kuliah harus dilakukan pada setiap pembuatan jadwal atau saat pergantian semester. Hal ini penting untuk menentukan berapa kelas yang akan dibuka dan nantinya akan mempengaruhi pada proses pembuatan jadwal perkuliahan sesuai dengan daya tampung, jumlah peminat, ketersediaan dosen dan jam pelajaran.

Penentuan jumlah kelas yang dibuka diperlukan sebuah perhitungan dengan menggunakan sebuah metode yang mampu memprediksi perkiraan jumlah mahasiswa yang akan mengambil mata kuliah tersebut. Dari total mahasiswa yang mengambil mata kuliah tertentu terdiri dari mahasiswa yang baru yang pertama kali mengambil mata kuliah dan mahasiswa yang mengulang untuk memperbaiki nilai yang pernah diperoleh pada semester sebelumnya. Mahasiswa mengulang mata kuliah yang sebelumnya memiliki nilai C dan C+, yang berpotensi untuk mengulang pada setiap mata kuliah tersebut. Hal ini karena untuk mahasiswa yang sudah mendapatkan nilai A, B+ dan B sudah dipastikan tidak akan mengulang karena nilai mereka sudah baik dan dipastikan lulus, sedangkan mahasiswa yang memperoleh nilai D dan E hampir dipastikan akan mengambil mata kuliah tersebut. Dari mahasiswa yang memperoleh nilai C+, C, D+, D da E tidak diketahui kapan mahasiswa tersebut akan mengulang.

Untuk dapat memberikan solusi dari permasalahan perkiraan jumlah mahasiswa yang akan mengulang mata kuliah yang sebelumnya memperoleh nilah C+, C, D+, D dan E pada setiap mata kuliah, maka ditawarkan sebuah sistem yang mampu membantu memecahkan masalah tersebut dengan menggunakan algotima genetika.

Penggunaan algoritma genetika untuk menyelesaikan permasalahan perkiraan atau prediksi telah dibahas dalam beberapa penelitian sebelumnya. Misalkan (Permatasari & Mahmudy, 2015) yang menerapkan algoritma genetika untuk melakukan prediksi konsumsi kWh listrik di kota Batu. Penelitian sejenis dilakukan pula oleh (Wahyuni, Sutojo, & Luthfiarta, 2014) yang menerapkan Naïve Bayes dan Algoritma Genetika untuk memprediksi hasil pemilu legislatif DKI Jakarta. Penelitian selanjutnya yang dilakukan oleh (Rahmi, Mahmudy, & Setiawan, 2015) yang menggunakan model regresi yang dibangun dengan menggunakan algoritma genetika untuk memprediksi harga saham berdasarkan data historis. Dari ketiga hasil penelitian tersebut menyimpulkan bahwa algoritma genetika dapat digunakan untuk memprediksi dengan cara membentuk koefisien terbaik pada analisis regresi linear berganda.

Algoritma genetika yang telah berhasil diterapkan pada berbagai masalah kompleks, maka pada skripsi ini agoritma genetika dgunakan untuk optimasi persamaan regresi linear berganda dalam menentukan jumlah mahasiswa yang mengambil mata kuliah.

1.2. Rumusan Masalah

Dari paparan pendahuluan yang telah dikemukakan diatas, dapat dirumuskan permasalahan sebagai berikut:

1. Bagaimana menerapkan metode algoritma genetika untuk optimasi model regresi linear berganda pada prediksi jumlah mahasiswa yang mengambil mata kuliah?
2. Bagaimana menentukan representasi kromosom terbaik untuk optimasi model regresi linear berganda menggunakan algoritma genetika dalam memprediksi jumlah mahasiswa yang mengambil mata kuliah?
3. Bagaimana parameter terbaik dalam optimasi menggunakan agoritma genetika untuk menentukan jumlah mahasiswa yang mengambil mata kuliah?

1.3 Batasan Masalah

Adapun batasan masalah dalam penelitian ini yaitu:

1. Metode yang digunakan yaitu optimasi menggunakan algoritma genetika untuk

menentukan model persamaan analisis regresi linear berganda.

2. Penerapan prediksi mahasiswa yang mengambil mata kuliah hanya untuk memprediksi jumlah mahasiswa yang memungkinkan mengulang pada mata kuliah tertentu, yaitu mahasiswa yang memperoleh nilai C+, C, D+, D dan E pada perkuliahan sebelumnya.

1.4 Tujuan

Adapun tujuan dari penelitian ini adalah:

1. Untuk menerapkan optimasi menggunakan algoritma genetika dalam melakukan proses prediksi jumlah mahasiswa yang mengambil mata kuliah.
2. Untuk mengetahui representasi kromosom yang terbaik dalam proses optimasi algoritma genetika dalam proses prediksi jumlah mahasiswa yang mengambil mata kuliah.
3. Untuk mengetahui parameter terbaik dalam menentukan jumlah mahasiswa yang akan mengambil mata kuliah.

1.5 Manfaat

Manfaat yang diperoleh dari skripsi ini yaitu:

1. Mengetahui prediksi jumlah mahasiswa yang akan mengambil mata kuliah dari mahasiswa yang sebelumnya memperoleh nilai C+, C, D+, D dan E.
2. Membantu pihak akademik dalam menentukan jumlah kelas yang akan dibuka pada tiap semesternya.

2. TINJAUAN PUSTAKA

Algoritma genetika (Genetic Algorithm, GAs) mampu menyelesaikan berbagai permasalahan kompleks dan dapat mengalami perkembangan seiring perkembangan teknologi (Mahmudy, 2013). GAs telah banyak digunakan untuk menyelesaikan berbagai macam kasus, terutama masalah optimasi, termasuk untuk membuat prediksi dari suatu kondisi. Pada penelitian ini akan dijabarkan beberapa referensi yang berhubungan dan relevan dengan kasus yang diangkat sebagai rujukan dalam menyelesaikan penelitian ini.

Pada penelitian sebelumnya yang pernah dilakukan dapat dilihat perbandingan objek, metode dan perbandingan dari masing-masing studi literatur skripsi peneliti. Perbandingan tersebut dapat dilihat pada Tabel 1 berikut.

Tabel 1. Perbandingan Objek dan Metode Penelitian Sebelumnya

No	Judul	Objek	Metode
1.	Optimasi Jaringan Syarat Tiruan dengan Algoritma Genetika untuk Peramalan Curah Hujan.	Peramalan Curah Hujan	Jaringan Syaraf Tiruan dan Algoritma Genetika
2.	Evolving RBF Neural Networks for Rainfall Prediction Using Hybrid Particle Swarm Optimization And Genetic Algorithm Review	Peramalan Curah Hujan	Hybrid PSO (Particle Swarm Optimization) dan Algoritma Genetika
3.	Perbandingan Metode Fuzzy dengan Regresi Linier Berganda dalam Peramalan Jumlah Produksi (Studi Kasus: Produksi Kelapa Sawit di PT. Perkebunan III (PERSERO) Medan Tahun 2011-2012)	Peramalan Jumlah Produksi	Perbandingan Metode Fuzzy dengan Regresi

2.1 Pengambilan Mata Kuliah

Mahasiswa akan mengambil suatu mata harus dilakukan setiap awal semester. Pada saat seperti itulah, mahasiswa akan melakukan program pengisian KRS (Kartu Rencana Studi) di Sistem Informasi Akademik Mahasiswa atau SIAM.

Mahasiswa yang telah menyelesaikan proses registrasi dan pembayaran finansial untuk pendidikan tiap semester akan dilanjutkan untuk melakukan pengisian KRS pada SIAM. Mahasiswa akan memilih mata kuliah, kelas, jumlah SKS maksimal dan hari perkuliahan sesuai dengan keminatan dan IP (indeks prestasi) yang diperoleh pada semester sebelumnya. Penentuan nilai IP merupakan rata-rata nilai yang diperoleh mahasiswa dalam rentang nilai 0,0-4,0. Setiap universitas akan memiliki aturan penilaian sebagai standar baku seperti yang terlihat pada Tabel.2 berikut:

Tabel 2. Aturan Penilaian Indeks Prestasi

Nilai Angka (N)	Nilai Huruf	Bobot
$80 < N \leq 100$	A	4
$75 < N \leq 80$	B+	3,5
$69 < N \leq 75$	B	3
$60 < N \leq 69$	C+	2,5
$55 < N \leq 60$	C	2
$50 < N \leq 55$	D+	1,5
$44 < N \leq 50$	D	1
$0 \leq N \leq 44$	E	0

Sumber: buku pedoman mahasiswa PTIIK Universitas Brawijaya

Dalam pemilihan kelas, mahasiswa yang akan mengambil mata kuliah harus memilih kelas yang telah disediakan atau dibuka oleh akademik dari fakultas. Idealnya jumlah mahasiswa dalam setiap kelas yang dibuka tidak lebih dari 40. Pada umumnya kelas yang dibuka ditandai dengan sebuah variabel, yaitu kelas A, B, C, D dan seterusnya.

Mahasiswa yang mengambil mata kuliah berasal dari dua jenis, yaitu mahasiswa yang baru pertama kali mengambil mata kuliah dan mahasiswa yang mengulang karena pada semester sebelumnya telah menempuh mata kuliah tersebut namun masih perlu untuk memperbaiki nilai yang diperolehnya.

2.2 Algoritma Genetika

Pada algoritma genetika, teknik pencarian dilakukan sekaligus atas sejumlah solusi yang mungkin dikenal dengan istilah populasi. Individu yang terdapat dalam satu populasi disebut dengan istilah kromosom. Kromosom ini merupakan suatu solusi yang masih berbentuk simbol. Populasi awal dibangun secara acak, sedangkan populasi berikutnya merupakan hasil evolusi kromosom-kromosom melalui iterasi yang disebut dengan generasi. Pada setiap generasi, kromosom akan melalui proses evaluasi dengan menggunakan alat ukur yang disebut dengan fungsi *fitness*.

Nilai fitness dari suatu kromosom akan menunjukkan kualitas dari kromosom dalam populasi tersebut. Semakin besar nilai *fitness*, maka menunjukkan bahwa kualitas dari kromosom semakin baik pula. Generasi berikutnya dikenal dengan istilah anak *(offspring)* terbentuk dari gabungan dua kromosom generasi sekarang yang bertindak sebagai induk *(parent)* dengan menggunakan operator penyilangan *(crossover)*.

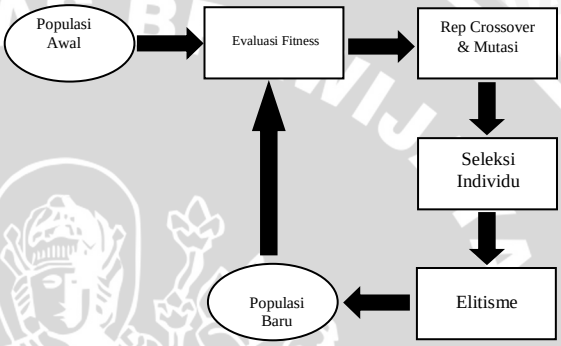
Selain operator penyilangan, suatu kromosom dapat juga dimodifikasi dengan menggunakan operator mutasi. Populasi generasi yang baru dibentuk dengan cara menyeleksi nilai *fitness* dari kromosom induk *(parent)* dan nilai *fitness* dari kromosom anak *(offspring)*, serta menolak kromosom-kromosom yang lainnya sehingga ukuran populasi (jumlah kromosom dalam suatu populasi) konstan. Setelah melalui beberapa generasi, maka algoritma ini akan konvergen ke kromosom terbaik.

2.3 Struktur Algoritma Genetika

Algoritma genetika merupakan teknik *search stchohastic* yang berdasarkan pada mekanisme seleksi alam dan genetika natural. Yang menjadi perbedaan antara algoritma genetika dengan algoritma konvensional lainnya adalah algoritma genetika menggunakan teknik pencarian secara sekaligus atas sejumlah solusi yang mungkin dikenal dengan istilah populasi.

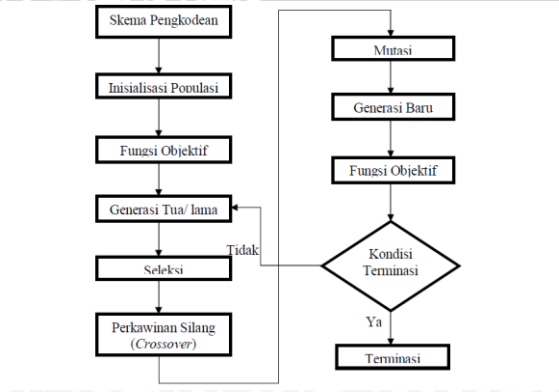
Setiap individu didalam populasi disebut kromosom, yang merepresentasikan suatu penyelesaian terhadap masalah yang ditangani. Sebuah kromosom terdiri dari sebuah string yang berisi berbagai simbol, dan biasanya, tetapi tidak mutlak, string tersebut berupa sederetan bit-bit biner "0" dan "1". Sebuah kromosom tumbuh atau berkembang biak melalui berbagai iterasi yang berulang-ulang, dan disebut sebagai generasi. Pada setiap generasi, berbagai kromosom yang dihasilkan akan dievaluasi menggunakan suatu pengukuran fitness.

Secara skematis, siklus algoritma genetika dapat digambarkan seperti pada Gambar 1 berikut:



Gambar 1. Siklus Algoritma Genetika

Algoritma genetika memberikan sebuah pilihan untuk penentuan nilai dari parameter yang digunakan dengan meniru proses reproduksi genetic, melalui pembentukan kromosom baru serta seleksi alami seperti yang terjadi pada makhluk hidup. Algoritma genetika secara umum dapat diilustrasikan dalam diagram alir pada Gambar 2 berikut:



Gambar 0. Diagram Alir Algoritma Genetika

Golberg (1989) mengemukakan bahwa algoritma genetika mempunyai karakteristik yang perlu diketahui sehingga dapat terbedakan dari prosedur pencarian atau optimasi yang lain, yaitu:

1. Algoritma genetika dengan pengkodean dari himpunan solusi permasalahan berdasarkan

parameter yang telah ditetapkan dan bukan parameter itu sendiri.
2. Algoritma genetika mencari sebuah solusi dari sejumlah individu-individu yang merupakan solusi permasalahan bukan hanya dari sebuah individu.
3. Algoritma genetika memiliki informasi fungsi objektif *(fitness)*, sebagai cara untuk mengevaluasi individu yang mempunyai solusi terbaik, bukan turunan dari sebuah fungsi.
4. Algoritma genetika menggunakan aturan-aturan transisi peluang, bukan aturan deterministik.

Parameter dan variabel yang digunakan dalam algoritma genetika yaitu:

1. Fungsi *fitness* yang dimiliki oleh setiap individu untuk menentukan tingkat kesesuaian individu tersebut dengan kriteria yang dicapai.
2. Populasi jumlah individu yang dilibatkan dalam setiap generasi.
3. Peluang terjadinya persinlangan *(crossover)* pada suatu generasi.
4. Peluang terjadinya mutasi pada setiap individu.
5. Jumlah generasi yang akan dibentuk menentukan lamanya penerapan algoritma genetika.

2.4 Operator Algoritma Genetika

Operator genetika yang digunakan setelah proses evaluasi tahap pertama membentuk poulasi baru dari genetika yang ada sekarang. Operator-operator tersebut adalah operator seleksi, *crossover* dan mutasi.

2.4.1 *Seleksi*

Seleksi bertujuan memberikan kesempatan reproduksi yang lebih besar bagi anggota populasi yang paling fit. Langkah pertama dalam seleksi ini adalah pencarian nilai fitness. Kemampuan algoritma genetika untuk melakukan produksi kromosom yang lebih baik secara progresif tergantung pada penekanan selektif *(selective pressure)* yang diterapkan ke populasi. Penekanan selektif dapat diterapkan dalam dua cara, yaitu:

1. Membuat lebih banyak kromosom anak yang dipelihara dalam populasi dan memilih hanya kromosom-kromosom terbaik bagi generasi berikut. Walaupun orang tua dipilih secara acak, metode ini akan terus menghasilkan kromosom yang lebih baik berhubungan dengan penekakan selektif yang diterapkan pada individu anak tersebut.
2. Memilih orang tua yang lebih baik ketika membuat keturunan baru. Dengan metode ini, hanya kromosom sebanyak yang dipelihara dalam populasi yang perlu dibuat bagi generasi berikutnya. Walaupun penekanan selektif tidak diterapkan ke level keturunan, metode ini akan terus menghasilkan kromosom yang lebih baik karena adanya penekanan selektif yang diterapkan ke orang tua.

2.4.2 *Perkawinan Silang (Crossover)*

Crossover (perkawinan silang) bertujuan menambah keanekaragaman string dalam populasi dengan penyilangan antar-string yang diperoleh dari sebelumnya. Beberapa jenis crossover tersebut adalah:

Crossover 1-titik

Pada crossover dilakukan dengan memisahkan suatu string menjadi dua bagian dan selanjutnya salah satu bagian dipertukarkan dengan salah satu bagian dari string yang lain yang telah dipisahkan dengan cara yang sama. Proses yang demikian dinamakan operator crossover satu titik seperti diperlihatkan pada tabel berikut:

Tabel 3. Contoh Crossover 1-titik

Kromosom orang tua 1	11001011
Kromosom orang tua 2	11011111
Keturunan	**11001111**

2.4.3 *Mutasi*

Mutasi merupakan proses mengubah nilai dari satu atau beberapa gen dalam suatu kromosom. Operasi *crossover* yang dilakukan pada kromosom dengan tujuan untuk memperoleh kromosom-kromosom baru sebagai kandidat solusi pada generasi mendatang denga *fitness* yang lebih baik, dan semakin lama menuju solusi optimum yang diinginkan. Akan tetapi, untuk mencapai hal ini, penekanan selektif juga memegang peranan yang penting. Jika dalam proses pemilihan kromosom-kromosom cenderung pada kromosom yang memiliki *fitness* yang tinggi saja, maka akan terjadi konvergensi *premature*. Konvergensi *premature* yaitu mencapai solusi yang optimal lokal sangat mudah terjadi.

2.5 Perhitungan *Fitness*

Mahmudy (2013), menjelaskan bahwa *Fungsi fitness* ditentukan untuk mengukur seberapa baik suatu individu. Individu dengan nilai *fitness* terbaik diakhir iterasi (generasi) akan tetap bertahan hidup dan dapat dikodekan sebagai solusi terbaik hasil dari persilangan dan mutasi. Maka semakin besar nilai fitnes semakin baik individu tersebut untuk mnejadi calon solusi.

Fungsi *fitness* ditunjukkan pada persamaan. Individu yang baik memiliki nilai *fitness* yang besar. Sedangkan fungsi hasil analisis regresi akan dikatakan baik apabila memiliki nilai yang error dalam jumlah sedikit. Sehingga nilai *fitness* yang baik diperoleh dari perhitungan satu MSE yaitu nilai error rata-rata kuadrat dari selisih Y' dan Y.

$$Fitness = \frac{1000}{\text{MSE}+1} + \frac{1}{\alpha n+1} \quad (1)$$

Keterangan :

MSE
= merupakan error rata
− rata kuadrat dari selisih Y dan Y′
α
= nilai alfa untuk melakukan normalisasi fitness.
n = banyaknya nilai 1 dalam sebuah kromosom.

2.6 Regresi Linier Berganda

Analisis regresi linier berganda merupakan hubungan secara linier antara dua atau lebih variabel *independen* (X1,X2,…,Xn) dengan variabel *dependen* (Y). Analisis ini untuk mengetahui arah hubungan antara variabel *independen* dan variabel *dependen* apakah memiliki berhubungan positif atau negatif dan untuk memprediksi nilai dari variabel *dependen* apabila variabel *independen* mengalami kenaikan atau penurunan. Data yang digunakan biasanya berskala interval atau rasio.

Analisa regresi berganda adalah sebuah teknik statistik yang dapat digunakan untuk menganalisis hubungan linier satu variabel respond antara beberapa *variabel* predictor. Model regresi linear berganda yang melibatkan *k* buah variabel *predictor* (X_1, X_2, X_3,…. X_k dapat dilihat pada persamaan berikut (Daniel, 1989):

$$y_i = \beta_0 \sum_{j=1}^{p} \beta_j \; x_{ij} + \; \varepsilon_i \quad (2)$$

Taksiran persamaannya adalah

$$\hat{y} = b_0 \sum_{j=1}^{p} b_j \; x_{ij} \quad (3)$$

Apabila dijabarkan akan menjadi persamaan berikut.

$$\hat{y} = b_0 + b_1 X_1 + b_2 X_2 + \cdots + b_n X_n \quad (4)$$

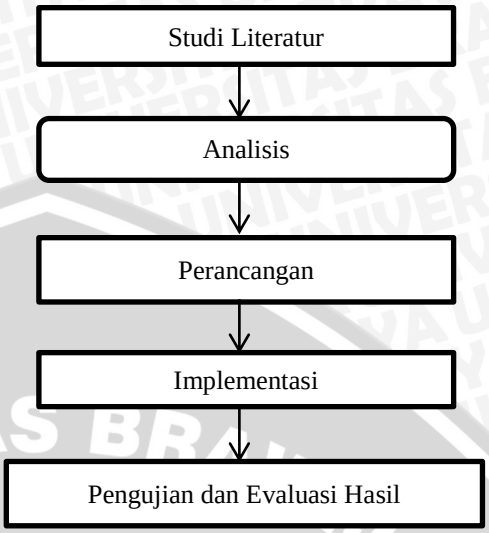
I = 1,2,…. N dan j = 1,2,…p, dengan n adalah banyaknya pengamatan dan (p+1) adalah banyaknya parameter.
Keterangan:
y = variabel respon.
ŷ = taksiran variabel respon.
x = variabel *predictor*.
β = variabel regresi.
b = taksiran variabel regresi.
ε = residual.

3. METODOLOGI PENELITIAN

Dalam tahapan ini menjelaskan susunan kegiatan penelitian yang dilakukan untuk membangun sebuah system algoritma genetika secara terstruktur serta perancangan system yang baik.

3.1 Tahapan Penelitian

Langkah-langkah yang dilakukan dalam proses perancangan penelitian implementatif menggunakan analisis regresi yang dioptimasi menggunakan algoritma genetika dapat dilihat pada Gambar 3.1 berikut:



Gambar 3. Alur dan Tahapan Penelitian

3.2 Data yang Digunakan

Untuk menyelesaikan permasalahan yang akan dibahas pada penelitian ini, peneliti menggunakan data historit jumlah mahasiswa yang memperoleh nilai C+,C,D+,D dan E dari jangka waktu tertentu. Dalam kasus ini, peneliti menggunakan data hasil generate yang merepresentasikan kondisi realistis mahasiswa yang mengulang dalam rentang waktu 24 tahun. Dari data generate ini, dapat digunakan untuk membangun model persamaan yang efektif untuk memprediksi jumlah mahasiswa yang akan mengulang.

3.3 Formulasi Permasalahan

Dimisalkan pihak akademik akan memprediksi jumlah mahasiswa yang akan mengulang pada mata kuiah ABC berdasarkan data 24 tahun terakhir. Maka untuk proses perhitungan regresi adalah sebagai berikut:

1. Penentuan jumlah mahasiswa yang telah mengambil mata kuliah ABC
   Sebelum data jumlah historis mahasiswa dihitung menggunakan fungsi regresi perlu diketahui bahwa data dengan periode tertetu (pertahun) dalam notasi X membutuhkan data 3 tahun periode historis. Sehingga, dengan data jumlah mahasiswa yang memperoleh nilai C+, C, D+, D dan E 24 tahun terakhir membutuhkan variabel dependen sebanyak 15 (X1 – X15).

Tabel 4. Data Perolehan Nilai Mata Kuliah ABC

Th MK	Jl	M (u)	Perolehan Nilai				
			C+	C	D+	D	E
1992	180	50	40	1	37	6	8
1993	145	43	35	16	26	4	6
1994	150	61	24	19	11	5	6

1995	163	72	38	27	16	6	8
1996	157	52	20	61	0	9	12
1997	176	66	38	4	41	8	11
1998	177	56	51	11	18	8	10
1999	175	51	31	5	52	7	10
2000	180	62	11	36	10	6	8
2001	172	53	24	26	11	6	6
2002	176	63	15	41	39	5	5
2003	168	64	37	32	46	6	6
2004	178	60	46	4	41	10	5
2005	174	76	36	13	31	8	4
2006	171	69	17	12	13	11	11
2007	183	74	43	30	0	12	9
2008	186	82	41	63	18	19	13
2009	175	75	50	26	9	22	11
2010	162	69	13	21	42	11	6
2011	202	88	48	43	39	11	6
2012	194	72	51	26	12	11	6
2013	209	78	24	35	38	10	6
2014	194	74	36	42	33	8	9
2015	190	80	43	34	45	12	9
1992	181	50	23	22	37	7	6

Tabel 4 menjelaskan bahwa jumlah mahasiswa yang menempuh perkuliahan ABC pada tahun mata kuliah (Th MK) tertentu terdiri dari dua jenis mahasiswa, yaitu mahasiswa baru yang ditandai dengan M(br) dan mahasiswa mengulang M(u). Dari jumlah M(br) dan M(u) ini akan diketahui jumlah sebenarnya mahasiswa yang menempuh perkuliahan ABC. Tabel 5 selanjutnya dimodelkan menjadi model analisis regresi linear berganda. Pada analisis regresi linear berganda akan dimodelkan dengan variabel regresi seperti yang ditunjukkan pada Tabel 5 berikut.

Tabel 5. Variabel dalam Regresi Linear Berganda

Variabel	Keterangan
Y(u)	Jumlah mengulang pada tahun ajaran (u).
X1	Jumlah mhs dengan nilai C+ pada Y(u)-1.
X2	Jumlah mhs dengan nilai C+ pada Y(u)-2.
X3	Jumlah mhs dengan nilai C+ pada Y(u)-3.
X4	Jumlah mhs dengan nilai C pada Y(u)-1.
X5	Jumlah mhs dengan nilai C pada Y(u)-2.
X6	Jumlah mhs dengan nilai C pada Y(u)-3.
X7	Jumlah mhs dengan nilai D+ pada Y(u)-1.
X8	Jumlah mhs dengan nilai D+ pada Y(u)-2.
X9	Jumlah mhs dengan nilai D+ pada Y(u)-3.
X10	Jumlah mhs dengan nilai D pada Y(u)-1.
X11	Jumlah mhs dengan nilai D pada Y(u)-2.
X12	Jumlah mhs dengan nilai D pada Y(u)-3.
X13	Jumlah mhs dengan nilai E pada Y(u)-1.
X14	Jumlah mhs dengan nilai E pada Y(u)-2.
X15	Jumlah mhs dengan nilai E pada Y(u)-3.

Tabel 5 menjelaskan variabel yang akan digunakan untuk analisis regresi dari data historis 3 tahun terakhir pada masing-masing nilai (C+, C, D+, D dan E) yang direpresentasikan dalam notasi X1, X2 hingga X15.

2. Persamaan Fungsi Regresi

Pada data yang ditunjukkan pada Tabel 5 akan dilakukan pembentukan fungsi regresi menggunakan persamaan (4) yang telah dijabarkan pada bab sebelumnya, maka pembentukan regresi yang terbentuk akan menjadi sebuah persamaan sebagai berikut:

Y' = 45.02 - 0.13X1 -0.01X2 -0.19X3 + 0.19X4 + 0.02X5 + 0.04X6 + 0.07X7 + 0.09X8 + 0.38X9 + 0.11X10 + 0.96X11 + 1.39X12 + 0.30X13 + 0.68X14 - 2.01X15

3.4 Alur Penyelesaian Masalah

3.4.1 Representasi Kromosom, Mencari *Error* dan Perhitungan *Fitness*

Representasi kromosom yang akan digunakan yaitu kromosom dengan representasi bilangan biner (nilai 0 dan 1). Dalam representasi kromosom, panjang kromosom yang digunakan yaitu sepanjang 15. Contoh representasi kromosom yang dibentuk sebagai berikut:

Kromosom individu : 00000 00001 10100

Dalam representasi kromosom dengan menggunakan bilangan biner diatas menjelaskan banyaknya jumlah variabel yang mempengaruhi jumlah mahasiswa yang akan mengulang mata kuliah tertentu. Nilai 1 menyatakan bahwa variabel tersebut mempengaruhi mahasiswa untuk mengulang mata kuliah dan nilai 0 berarti variabel tersebut tidak mempengaruhi jumlah mahasiswa yang mangambil mata kuliah.

Untuk menghitung nilai *Mean Square Error* menggunakan persamaan (1), sehingga diperoleh nilai *fitness*:

Fitness = 16.45863076923077

Dengan hanya melibatkan 3 variabel X sesuai dengan kromosom yag terpilih, yaitu 00000 00001 10100.

3.4.2 Perhitungan *Crossover*

Dalam penelitian ini menggunakan metode *one cut point crossover* yang ditentukan secara *random* oleh sistem untuk posisi *one cut point.* Dalam proses *crossover* bilangan biner dilakukan dengan menggantikan kromosom 2 induk dengan posisi yang *random.* Contoh *crossover* yang dilakukan yaitu misalkan individu 1 dan individu 3 terpilih sebagai induk yang akan dilakukan *crossover.* Misalkan posisi *one cut point* adalah 8, maka dapat dilihat pada Tabel 6 berikut:

Tabel 3.3.6. Proses *Crossover*

Individu Ke	Kromosom
P1	1 0 1 0 1 1 0 0 0 0 0 1 1 1 1
P3	1 0 1 0 0 0 0 1 1 0 1 1 1 1 0
Hasil *Crossover*	

C1	1 0 1 0 1 1 0 0 1 0 1 1 1 1 0
C2	1 0 1 0 0 0 0 1 0 0 0 1 1 1 1

3.4.3 Perhitungan Mutasi

Mutasi yang digunakan dalam memprediksi jumlah mahasiswa yang mengulang mata kuliah dilakukan dengan menggunakan *random mutation*. Yaitu melakukan pergeseran kromosom secara *random*. Dalam mutasi ini, nantinya akan menggantikan nilai kromosom pada posisi tertentu dengan nilai lain pada kromosom yang sama namun posisi yang berbeda. Contoh mutasi yang digunakan adalah sebagai berikut:

Posisi mutasi 1 = 5;
Posisi mutasi 2 = 11;
Individu ke 6 = 1 1 1 0 0 0 1 1 0 0 1 1 1 1 0
Offspring C3 = 1 1 1 0 0 1 1 1 0 0 1 1 1 1 0

3.4.4 Evaluasi dan Seleksi

Untuk mendapatkah hasil yang terbaik, maka dilakukan evaluasi dan seleksi pada hasil reproduksi yang telah ada sebelumnya. Dengan menghitung nilai *fitness* dari hasil reproduksi maka akan diperoleh individu-individu pilihan yang mampu bertahan hidup. Semakin tinggi nilai *fitness*, maka akan semakin baik nilai kromosom tersebut menjadi calon solusi untuk menyelesaikan permasalahan yang ada.

Proses seleksi yang dilakukan yaitu dengan menggunakan metode *replacement selection*. Proses yang dilakukan dalam *replacement selection* yaitu dengan mengganti induk dengan *offsring*-nya (anak) yang memiliki nilai *fitness* yang lebih besar dibandingkan induknya.

4. IMPLEMENTASI

Implementasi antarmuka sistem terdiri dari dua halaman yaitu halaman Proses Algoritma dan halaman About. Halaman Proses Algoritma ada 2 tabel yaitu tabel data mahasiswa dan tabel hasil perbandingan dari proses prediksi menggunakan analisis regresi linear berganda dan hasil prediksi optimasi menggunakan Algoritma Genetika.

Hasil implementasi aplikasi sebagai berikut:

Gambar 4. Halaman Utama Tab Proses Algoritma

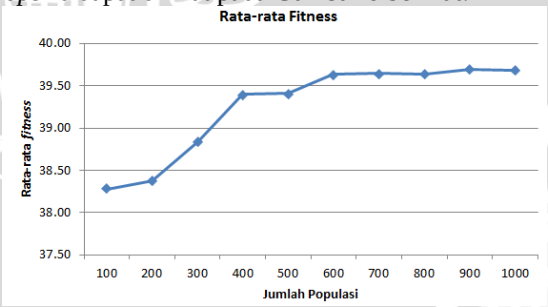
Gambar 5. Halaman Tab About

5. PENGUJIAN DAN ANALISIS

5.1 Pengujian dan Analisis Ukuran Populasi

Uji coba ini bertujuan untuk melihat pengaruh dari ukuran populasi atau *Popsize* terhadap perubahan pada nilai *fitness*. Pada pengujian ukuran populasi ini menggunakan parameter banyaknya generasi, *crossover rate*, dan *mutation rate* yang telah ditetapkan pada parameter sebelumnya. Sedangkan pada nilai *popsize* akan diganti secara bertahap dari 50 hingga 500 individu dengan kelipatan 50.

Data historis mahasiswa mengulang yang digunakan pada pengujian ini menggunakan formulasi data hasil *generate* yang merepresentasikan kondisi realistis selama 24 tahun. Jumlah generasi yang digunakan adalah 10 hingga 100 generasi. *Crossover rate* yang digunakan adalah 0.8 dan *mutation rate* adalah 0.2. Setiap *popsize* yang dimasukkan akan diuji sebanyak 10 kali agar dapat diketahui rata-rata *fitness* tersebut. Dari pengujian ini akan diketahui nilai *popsize* yang optimal untuk menyelesaikan malasah pemodelan analisis regresi linear berganda untuk jumlah mahasiswa yang mengulang. Hasil dari pengujian *popsize* dapat dilihat pada Gambar 6 berikut:



Gambar 6. Grafik Ukuran Populasi

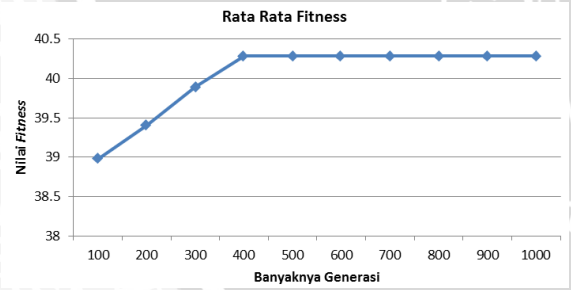
Dari gambar 6 dapat dijelaskan bahwa pada populasi 100 menjadi populasi yang memiliki nilai rata-rata *fitness* yang terendah, hal ini dikarenakan populasi dengan jumlah 100 belum mencakup secara optimal daerah pencarian pada algoritma genetika. Pada populasi 600 telah terjadi konvergensi hingga populasi 1000. Namun pada populasi 900 menjadi populasi terbaik karena memiliki nilai rata-rata *fitness* yang paling besar, selanjutnya pada populasi 1000 tidak terjadi perbedaan rata-rata nilai *fitness* yang jauh. Hal ini menandakan bahwa telah terjadi konvergensi pada populasi 600 hingga 1000.

Banyaknya populasi yang digunakan akan mempengaruhi lamanya komputasi yang diperlukan, namun dengan populasi yang besar akan menghasilkan nilai *fitness* yang lebih tinggi dan menghasilkan solusi yang semakin baik pula. Pada penelitian sebelumnya yang dilakukan oleh Fakhiroh, D, Mahmudy, WF & Indriati (2015), menghasilkan pengujian dengan pola yang sama dimana mulai terjadi kovergensi pada ukuran populasi tertentu.

5.2 Pengujian dan Analisis Ukuran Generasi

Pada pengujian ini dilakukan untuk mengetahui banyaknya generasi terhadap perubahan nilai *fitness*. Parameter algoritma genetika seperti nilai *popsize*, *crossover rate*, dan *mutation rate* menggunakan nilai tetap selama proses pengujian ini. Banyaknya generasi yang akan diuji akan diganti secara bertahap dengan kelipatan 100 mulai dari 100 generasi hingga 1000 generasi.

Pada pengujian ini menggunakan data historis jumlah mahasiswa yang memperoleh nilai C+, C, D+, D dan E sebagai masukan system. *Popsize* yang digunakan dalam pengujian ini adalah 100 individu, dengan *crossover rate* sebesar 0.8 dan *mutation rate* 0.2. Setiap generasi akan melalui 10 kali percobaan kemudian dihasilkan rata-rata *fitness* dari percobaan tersebut. Dari hasil pengujian ini akan diketahui banyaknya generasi yang optimal untuk mengetahui faktor yang mempengaruhi jumlah mahasiswa yang mengulang dari model persamaan analisis regresi linear berganda untuk memprediksi jumlah mahasiswa yang mengulang. Hasil pengujian dari uji coba banyaknya generasi dapat diihat pada Gambar 7.

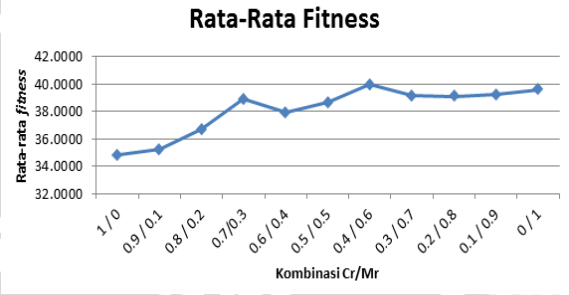


Gambar 7. Grafik Ukuran Generasi

Dari Gambar 7, grafik hasil pengujian banyaknya generasi 400 menunjukkan angka *fitness* yang konstan. Dengan melakukan pengujian sebanyak generasi 400, sudah tidak didapatkan nilai *fitness* yang lebih variatif dikarenakan telah terjadi konvergensi. Generasi 400-1000 telah memiliki rata-rata nilai *fitness* yang sama, yaitu 40,2861.

5.3 Pengujian dan Analisis Kombinasi Cr dan Mr

Pengujian untuk kombinasi nilai *crossover rate (cr)* dan *mutation rate (mr)* dilakukan untuk mendapatkan kombinasi yang paling optimal sehingga nilai *fitness* yang diperoleh juga optimal. Parameter algoritma seperti jumlah generasi dan *popsize* telah ditetapkan sesuai dengan data hasil pengujian terbaik pada proses sebelumnya. Untuk proses *crossover rate* dan *mutation rate* akan mengalami pergantian kombinasi secara bertahap dari 0 hingga 1 dengan kelipatan 0.1 setiap perubahan. Hasil pengujian kombinasi *crossover rate* dan *mutation rate* dapa dilihat pada Gambar 8 berikut.



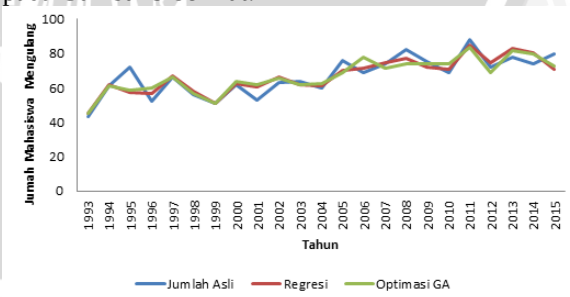
Gambar 8. Grafik Hasil Pengujian Kombinasi Crossover Rate (cr) dan Mutation Rate (mr)

Dari Gambar 6.3 diatas dapat diketahui bahwa kombinasi dari crossover rate dan mutation rate dapat diketahui bahwa kombinasi cr = 0.4 dan mr = 0.6 menjadi kombinasi terbaik dari hasil pengujian yang dilakukan.

5.4 Pengujian Perbandingan Hasil Prediksi dari Analisis Regresi dan Optimasi Algoritma Genetika

Pengujian dari hasil prediksi jumlah mahasiswa dengan menggunakan metode analisis regresi linear berganda dan hasil optimasi menggunakan algoritma genetika ini menggunakan beberapa parameter yang berbeda. Seperti jumlah individu, kombinasi *cr* dan *mr*, serta banyaknya iterasi yang digunakan.

Pengujian dengan jumlah populasi sebanyak 50 populasi, *crossover rate* 0.4, *mutation rate* 0.6 serta iterasi sebanyak 1 kali menghasilkan hasil seperti pada Gambar 9 berikut:



Gambar 9. Grafik Perbandingan Hasil Hasil Pengujian

Pada hasil pengujian Tabel 6.4, diperoleh selisih MSE sebesar 5,64 lebih baik pada analisis regresi linear berganda, individu yang terpilih adalah P89 dengan kromosm 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1. Pada kromosom P89 hanya menggunakan 13 variabel independen (nilai X) untuk menghasilkan hasil analisis yang sama akuratnya dengan regresi linear berganda. Apabila dari data pada Tabel 6.4 disajikan dalam bentuk grafik, maka akan terlihat seperti pada

Gambar 6.4 berikut. Pada Gambar 6.4 terdapat tiga parameter yaitu jumlah asli, jumlah yang diperoleh dari hasil analisis regresi dan jumlah yang diperoleh dari hasil optimasi algoritma genetika. Pada sumbu X (horizontal) merupakan tahun perkuliahan. Pada bidang vertical (sumbu Y) terdapat parameter jumlah mahasiswa yang mengulang.

6. PENUTUP

6.1 Kesimpulan

Kesimpulan yang diperoleh dari hasil implementasi dan pengujian yang telah diakukan dalam penelitian ini adalah:

1. Penerapan metode algoritma genetika dalam melakukan optimasi model regresi linear berganda dengan melakukan pengkodean kromosom secara binary, crossover menggunakan one cut point crossover, metode permutasi yang digunakan adalah random mutation dengan metode seleksi elitism.
2. Representasi kromosom terbaik yang dihasilkan dari optimasi model regresi linear berganda menggunakan algoritma genetika adalah 1 1 1 1 1 0 1 1 1 0 1 1 1 1 1 dengan selisih error sebesar 5,64 lebih baik pada analisis menggunakan regresi linear berganda dibandingkan dengan hasil otimasi menggunakan algoritma genetika. Namun, pada agoritma genetika tidak menggunakan semua variabel independen (variabel X) sebagai variabel yang mempengaruhi prediksi analisis regresi linear berganda. Hanya variabel X yang memiliki kromosom 1 yang digunakan dalam proses perhitungan prediksi jumlah mahasiswa yang mengulang.
3. Parameter yang dihasilkan untuk populasi terbaik yang terpilih adalah 600 dengan kombinasi crossover rate 0.4 dan mutation rate 0.6 dan banyaknya generasi sebesar 400 generasi.

6.2 Saran

Saran yang dapat diberikan untuk mengembangkan penelitian ini agar menjadi lebih baik adalah:

1. Perlu melakukan uji coba pada algoritma evolusi lain. Sehingga diharapkan memperoleh variasi hasil yang lebih banyak dan memungkinkan menghasilkan hasil yang lebih baik.
2. Perlu melakukan uji kombinasi *crossover rate*, *mutation rate* yang lebih kompleks dan lebih lengkap untuk menemukan kombinasi nilai *cr* dan *mr* yang paling baik.

7. DAFTAR PUSTAKA

Arianti, N. (2005). Optimasi Jaringan Syarat Tiruan dengan Algoritma Genetika untuk Peramalan Curah Hujan. IPB Repository.

Daniel, W. W. (1989). Statistika Non Parametrik Terapan. Jakarta: Gramedia.

Goldberg, D.E. (1989). Genetic Algorithm in Search, Optimization and Machine Learning. New York: Addison Weasley Publishing Company Inc.

Kurniawan, D. (2008). Regresi Linier.

Kustiyo, Aziz., Buono, Agus,. & Apriyanti, Novi. (2006). Optimasi Jaringan Syarat Tiruan dengan Algoritma Genetika untuk Peramalan Curah Hujan. IPB Repository Jurnal Ilmu Komputer (74).

Mahmudy, W. F. (2013). Algoritma Evolusi. Program Teknologi Informasi dan Ilmu Komputer Universitas Brawijaya Malang.

Mahmudy, W. F., Marian, R. M., & Luong, L. H. (2014). Hybrid Genetic Algorithms for Part Type Selection and Machine Loading Problems with Alternatif Production Plans in Flexible Manufacturing System. ECTI Transactions On Computer And Information Technology (ECTI-CIT), 8(1), 80-93.

Permatasari, A. I., & Mahmudy, W. F. (2015). Pemodelan Regresi Linear dalam Konsumsi Kwh Listrik di Kota Batu Menggunakan Algoritma Genetika. DORO: Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya, 5(14).

Rahmi, A., Mahmudy, W. F., & Setiawan, B. D. (2015). Prediksi Harga Saham Berdasarkan Data Historis Menggunakan Model Regresi yang Dibangun dengan Algoritma Genetika. DORO: Repository Jurnal Mahasiswa PTIIK Universitas Brawijaya, 5(12).

Wahyuni, D. T., Sutojo, T., & Luthfiarta, A. (2014). Prediksi Hasil Pemilu Legislatif DKI Jakarta Menggunakan Naïve Bayes dengan Algoritma Genetika Sebagai Fitur Seleksi.

Wati, S. E., Sebayang, D., & Sitepu, R. (2013). Perbandingan Metode Fuzzy dengan Regresi Linier Berganda dalam Peramalan Jumlah Produksi (Studi Kasus: Produksi Kelapa Sawit di PT. Perkebunan III (PERSERO) Medan Tahun 2011-2012). Saintia Matematika, 1(3), 273–284.

Wu, Jiansheng., Long, Jin., & Liu, Mingzhe. (2012). Evolving RBF Neural Networks for Rainfall Prediction using Hybrid Particle Swarm Optimization And Genetic Algorithm. Neurocomputing Vol 148. pp 136–142