

# Optimalisasi Regresi Logistik Menggunakan Algoritma Genetika Pada Data Klasifikasi

*by* Abdurrahman Salim

---

**Submission date:** 16-Oct-2020 09:17AM (UTC+0700)

**Submission ID:** 1416660480

**File name:** 109-Article\_Text-538-1-10-20191225.pdf (276.67K)

**Word count:** 2967

**Character count:** 17316

# Optimalisasi Regresi Logistik Pada Proses Klasifikasi Menggunakan Algoritma Genetika

Abdurrahman Salim  
Program Studi Budidaya Tanaman Perkebunan  
Jurusan Produksi Pertanian  
Politeknik Negeri Jember  
Jember, Jawa Timur Indonesia  
abdurrahman.salim@polije.ac.id

Muhammad Rijal Alfian  
Program Studi Teknik Komputer  
Univesitas Teknologi Mataram  
Mataram, Nusa Tenggara Barat Indonesia  
rijal.alfian@gmail.com

**Abstract**— Classification of data is a large amount of data and with a variety of attributes so that it can make accuracy results low. Low accuracy can result in a method that cannot be represented in applying a problem. One of method is Logistic Regression method. Logistic Regression is one of classification method, if response variable has binary characteristic and there are many predictor variable such as combination of category and continue. Method of Logistic Regression requires a stage selection independent variable in improving the model accuracy. So it takes a good method in fixing the deficiency is Genetic Algorithm (GA). This method is an iterative method to get global optimum. The results of the classification accuracy of Logistic Regression in the case of septictank data in East Surabaya with 11 independent variables and binary dependent variable is Logistic Regression accuracy of 54.55%. However when selected with GA, the classification accuracy of Binary Logistic Regression is 90.91%.

**Keywords**— *Genetic Algorithm (GA), Logistic Regression, Classification.*

**Abstrak**— Pengklasifikasian data merupakan sejumlah data dalam jumlah banyak dan dengan atribut yang beraneka ragam sehingga dapat membuat hasil akurasi menjadi rendah. Akurasi yang rendah dapat mengakibatkan metode tidak dapat direpresentasikan dalam menerapkan suatu masalah. Sehingga salah satu metode yang dapat menangani atribut yang beraneka ragam adalah metode Regresi Logistik. Regresi logistik merupakan salah satu metode dalam pengklasifikasian, dimana terdapat variabel respon yang bersifat biner dan terdapat banyak variabel prediktor berupa gabungan katagorik dan kontinu. Metode Regresi Logistik ini membutuhkan tahapan seleksi variabel Independen dalam meningkatkan keakurasian modelnya. Sehingga dibutuhkan metode yang bagus dalam memperbaiki kekurangan tersebut yaitu algoritma genetika.

Metode ini merupakan metode iteratif untuk mendapatkan global optimum. Dalam penerapannya menggunakan data septiktank di wilayah Surabaya Timur dimana hasil ketepatan klasifikasi yang dihasilkan dari Regresi Logistik dengan 11 variabel independen dan variabel dependennya berbentuk biner dengan akurasi sebesar 54.55 %. Namun ketika diseleksi dengan Algoritma Genetika, Regresi Logistik Biner memiliki ketepatan klasifikasi yaitu 90.91 %.

**Keywords**— *Algoritma Genetika, Regresi Logistik, Klasifikasi*

## PENDAHULUAN

Pengklasifikasian merupakan salah satu metode statistika untuk mengelompok atau mengklasifikasi suatu data yang disusun secara sistematis. Masalah klasifikasi sering dijumpai dalam kehidupan sehari-hari. Baik itu pengklasifikasian data pada bidang akademik, sosial, pemerintahan, maupun pada bidang lainnya. Masalah klasifikasi ini muncul ketika terdapat sejumlah ukuran yang terdiri dari satu atau beberapa kategori yang tidak dapat diidentifikasi secara langsung tetapi harus menggunakan suatu ukuran.

Dalam statistika ada beberapa metode klasifikasi yang digunakan untuk melakukan klasifikasi data seperti: Analisis Diskriminan, Regresi Logistik, Naïve Bayes, dll. Masing-masing metode tersebut memiliki kelebihan dan kelemahan. Dalam penelitian ini yang akan dibahas ialah metode klasifikasi regresi logistik.

Regresi logistik biner adalah salah satu metode statistika yang sering digunakan untuk mengklasifikasikan sejumlah pengamatan dengan respon biner ke dalam beberapa kelompok berdasarkan satu atau lebih variabel prediktor. Melalui metode ini akan dihasilkan peluang dari masing-masing kategori respon yang akan dijadikan sebagai pedoman pengklasifikasian dan suatu pengamatan akan masuk kedalam respon kategori tertentu berdasarkan nilai peluang yang terbesar [1]. Regresi logistik juga dapat diartikan salah satu pendekatan model matematis yang digunakan untuk menganalisis hubungan antara satu

atau beberapa variabel independen yang bersifat kontinu maupun biner dengan satu variabel dependen yang bersifat dikotomis (biner). Dalam penggunaan metode Regresi Logistik diperlukan tahapan penyeleksian variabel untuk meningkatkan keakuratan model dalam menjelaskan data. Sehingga diperlukan metode lain dalam seleksi variabel salah satunya adalah Algoritma Genetika. Algoritma genetik merupakan salah satu algoritma optimasi, yang diciptakan untuk meniru beberapa proses yang diamati dalam evolusi alam [2]. Metode ini termasuk salah satu metode optimasi yang mutakhir [3]. Algoritma Genetika adalah suatu teknik optimasi yang didasarkan pada prinsip genetik dan seleksi alam. Dalam Algoritma Genetika populasi terbentuk dari banyak individu yang berkembang sesuai aturan seleksi spesifik dengan memaksimalkan fitness. Menurut Sivanandam dan Deepa [4] kelebihan yang dimiliki Algoritma Genetika dibanding metode-metode yang lain diantaranya yaitu sangat cocok digunakan untuk menyelesaikan masalah global optimum, mudah diubah atau fleksibel untuk diimplementasikan pada berbagai masalah dan ruang solusi lebih luas. Menurut insani, 2019 [5] Algoritma Genetika dapat digunakan dalam optimasi biaya bahan dan jasa pembangunan rumah. Sedangkan menurut Buani, 2016 [6] bahwa setelah optimasi dengan menggunakan data yang sama untuk mengoptimalkan Naive Bayes dengan Algoritma Genetika hasilnya meningkat. Menurut Purba dkk, 2019 [7] bahwa penggunaan regresi logistik pada data microarray lebih kecil akurasi dibandingkan dengan akurasi regresi logistik dengan algoritma genetika yang lebih tinggi akurasi dan lebih baik dalam pengaplikasiannya.

Berdasarkan uraian di atas, tujuan dari penelitian ini adalah keakuratan klasifikasi dari metode Regresi Logistik Biner dengan pengoptimumannya menggunakan Algoritma Genetika.

## REGRESI LOGISTIK

Regresi logistik merupakan salah satu metode klasifikasi yang sering digunakan. Regresi logistik biner digunakan saat variabel dependen merupakan variabel dikotomis. Menurut Hosmer, D.W. dan Lemeshow, S. 2000 didalam Purwa, 2019 [7] bahwa secara umum model regresi logistik adalah :

$$\pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}} \quad (1)$$

Dimana  $\pi(x)$  merupakan nilai probabilitas dari  $0 \leq \pi(x) \leq 1$ , yang berarti bahwa regresi logistik menggambarkan suatu probabilitas. Dengan

mentransformasikan  $\pi(x)$  pada persamaan di atas dengan transformasi logit  $g(x)$ , dimana :

$$g(x) = \ln \left( \frac{\pi(x)}{1 - \pi(x)} \right) \quad (2)$$

maka diperoleh bentuk logit :

$$g(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k \quad (3)$$

## ALGORITMA GENETIKA

Algoritma Genetika menggunakan analogi secara langsung dari kebiasaan yang alami yaitu seleksi alam. Algoritma ini berkerja dengan sebuah populasi yang terdiri dari individu-individu, yang masing-masing individu merepresentasikan sebuah solusi yang mungkin bagi persoalan yang ada. Dalam kaitan ini, individu dilambangkan dengan dengan sebuah nilai fitness yang akan digunakan untuk mencari solusi terbaik dari persoalan yang ada [8]. Algoritma Genetika merupakan metode pencarian yang meniru perumpamaan evolusi biologis alami untuk menentukan kromosom atau individu berkualitas tinggi dalam suatu kawasan berhingga potensial yang disebut populasi. Proses pemilihan individu dari suatu populasi dievaluasi berdasarkan fungsi fitness. Kromosom berwujud string tersebut merupakan calon pada setiap siklus operasi yang disebut generasi [9]. Algoritma ini juga digunakan untuk mendapatkan nilai global optimum dengan cara melakukan perulangan atau iterasi pada konsep evolusi darwin.

Menurut Trevino dan Falcian [10] terdapat 7 tahapan untuk menjalankan algoritma genetika, yaitu :

1. Membentuk populasi awal terdiri dari beberapa kromosom yang didalamnya memuat gen kromosom pada algoritma genetika yang digunakan untuk menunjukkan kandidat sekelompok gen yang dapat digunakan sebagai solusi permasalahan. Gen pada algoritma genetika berisi variabel yang ingin dioptimumkan, pada penelitian ini gen berisi variabel independen.
2. Masing-masing kromosom dalam populasi dievaluasi kemampuannya dengan menggunakan fungsi fitness. Pada penelitian ini fungsi fitnessnya berupa kesalahan klasifikasi.
3. Ketika sebuah kromosom memiliki nilai fitness lebih optimum daripada nilai inisialnya, maka kromosom dihentikan, namun apabila tidak maka tahapan analisis dilanjutkan ke tahap 4. Nilai fitness terkecil yang dipilih sebagai solusi permasalahan dari penelitian ini, karena fungsi fitness yang digunakan adalah tingkat kesalahan klasifikasi.

4. Memilih kromosom dengan nilai fitness yang optimum yang dijadikan orang tua.
5. Mengkombinasikan informasi genetika yang ada dalam replikasi orang tua melalui pindah silang. Dua induk secara random dipilih dan digunakan untuk membentuk dua kromosom baru.
6. Melakukan mutasi untuk memperkenalkan unsur gen baru pada kromosom secara acak.
7. Tahapan diulangi dari tahapan 2 sampai kromosom yang memberikan nilai fitness paling optimum atau sudah mencapai konvergen.

### KLASIFIKASI DATA

Pada tahapan ini pengujian dilakukan dengan menghitung nilai Accuracy, Sensitivity, dan Precision [11]. Akurasi merupakan perhitungan jumlah proporsi prediksi yang benar [12].

Data aktual dan data hasil prediksi dari model klasifikasi disajikan dengan menggunakan Tabulasi silang (Confusion matrix), yang mengandung informasi tentang kelas data aktual direpresentasikan pada baris matriks dan kelas data hasil prediksi pada Tabel 1.

Tabel 1. Tabel Klasifikasi

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

dimana,

$$akurasi = \frac{TN + TP}{TN + TP + FN + FP} \quad (4)$$

Selain akurasi, performa klasifikasi dapat dinilai pula berdasarkan nilai sensitivity dan specificity. Sensitivity merupakan akurasi kelas positif sedangkan specificity merupakan akurasi pada kelas negatif. Rumus sensitivity dan specificity adalah sebagai berikut.

$$Sensitivity = \frac{TP}{(TP + FN)} \times 100\% \quad (5)$$

$$Specificity = \frac{TN}{(TN + FP)} \times 100\% \quad (6)$$

### METODE PENELITIAN

#### A. Sumber Data

Sumber data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari penelitian [13]. Data yang diambil adalah rumah tangga yang membuang limbah domestik (rumah tangga) atau blackwater (tinja) pada tangki septik dan rumah tangga yang membuang limbah langsung ke badan air (selokan, sungai, dan lain-lain) atau yang mencemari tanah. Jumlah data yang digunakan adalah 102 observasi.

#### B. Variabel Penelitian

Variabel respon yang digunakan dalam penelitian ini adalah terdapat pada Tabel 2., sebagai berikut :

Tabel 2. Variabel Respon

Variabel Respon (Y)	
Y = Pembuangan Air Limbah Domestik blackwater (tinja) yang dibuang oleh setiap rumah tangga	0 = mempunyai septictank 1 = tidak mempunyai septictank

Variabel prediktor yang digunakan untuk mengetahui faktor-faktor yang mempengaruhi rumah tangga membuang limbah domestik di Surabaya Timur pada Tabel 3., sebagai berikut.

Tabel 3 Variabel Prediktor

Variabel Prediktor(X)	
Variabel	Keterangan
X <sub>1</sub> = Jumlah keluarga	0 = ≤4 1 = >4
X <sub>2</sub> = Pendidikan Kepala Keluarga	0 = ≤SD 1 = SMP/Sederajat 2 = SMA/Sederajat 3 = Diploma/Sarjana
X <sub>3</sub> = Pekerjaan Kepala Keluarga	0 = PNS/BUMN 1 = Karyawan Swasta 2 = Wirausaha 3 = Pertukangan
X <sub>4</sub> = Pendapatan Kepala Keluarga	0 = < Rp 1.750.000 1 = ≥ Rp 1.750.000
X <sub>5</sub> = Pengeluaran Rumah Tangga	0 = < Rp 1.750.000 1 = ≥ Rp 1.750.000
X <sub>6</sub> = Status Kepemilikan Rumah	0 = Status Hak Milik (SHM) 1 = Kontrak/Kos 2 = Milik Orangtua 3 = Rumah Dinas
X <sub>7</sub> = Lama Waktu Pengurusan	0 = ≤4 1 = >4
X <sub>8</sub> = Jenis Kloset	0 = Kloset Duduk 1 = Kloset Jongkok
X <sub>9</sub> = Penyakit	0 = Diare (Sakit Perut)

Variabel Prediktor(X)	
Variabel	Keterangan
	1 = Tipus 2 = Demam Berdarah 3 = Lainnya
X <sub>10</sub> = Air Mandi dan Cuci	0 = Air Tanah/Sumur 1 = PDAM
X <sub>11</sub> = Perilaku Sanitasi	0 = Baik 1 = Kurang Baik

C. Tahapan Penelitian

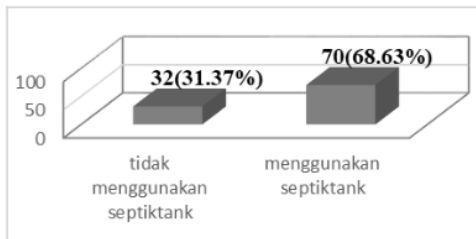
Langkah analisis yang digunakan secara garis besar adalah sebagai berikut :

1. Melakukan analisis karakteristik pada kasus data.
2. Mengolah data menggunakan Regresi Logistik dan Naïve Bayes, kemudian menentukan ketepatan klasifikasi.
3. Menggabungkan metode Regresi Logistik dan Naïve Bayes dengan Algoritma Genetika untuk seleksi variabel dan menentukan ketepatan klasifikasinya
4. Membandingkan ketepatan klasifikasi regresi logistik dan Naïve Bayes serta penggunaan algoritma genetika dalam penseleksian.
5. Perbandingan pengklasifikasian terbaik menggunakan Akurasi, APER, Sensitivity, Specificity, dan G-Mean

HASIL DAN PEMBAHASAN

A. Analisis Karakteristik Data

Karakteristik yang terdapat pada pembuangan limbah domestik blackwater (feces). Adapun jumlah rumah tangga yang sudah menggunakan septictank yang sesuai dengan standar SNI dan belum menggunakan septictank adalah sebagai berikut.



Gambar 1. Jumlah Pembuangan Limbah Domestik Rumah Tangga

B. Analisis Regresi Logistik Biner

Regresi Logistik merupakan salah satu metode dalam mengatasi data klasifikasi. Dalam metode ini

penentuan estimasi model yang digunakan adalah untuk mengetahui faktor mana yang mempengaruhi secara signifikan untuk model terbaik sehingga dapat dilihat dari akurasi. Maka dilakukan pembagian data Training 90% dan data Testing sebesar 10 % dari 102 observasi. Pada Tabel 4 akan dilihat estimasi secara keseluruhan dengan menggunakan data Training untuk melihat model yang signifikan sebelum penentuan klasifikasinya, sebagai berikut:

Tabel 4. Estimasi Parameter Regresi Logistik

Koefisien	Estimasi	Std. Error	z-value	Pr(> z )
(Intercept)	11,5660	2095,1010	0,0060	0,9960
<b>X<sub>1</sub> (1)</b>	<b>4,5360</b>	<b>1,5080</b>	<b>3,0070</b>	<b>0,0030</b>
X <sub>2</sub> (1)	0,7490	0,9860	0,7600	0,4480
X <sub>2</sub> (2)	2,5110	1,3220	1,9000	0,0570
X <sub>2</sub> (3)	3,1650	1,6270	1,9440	0,0520
X <sub>3</sub> (1)	3,3540	2,0650	1,6240	0,1040
X <sub>3</sub> (2)	3,4930	1,8870	1,8520	0,0640
<b>X<sub>3</sub> (3)</b>	<b>4,9380</b>	<b>2,1300</b>	<b>2,3190</b>	<b>0,0200</b>
X <sub>4</sub> (1)	-0,9020	0,9220	-0,9770	0,3280
<b>X<sub>5</sub> (1)</b>	<b>-3,8350</b>	<b>1,4560</b>	<b>-2,6350</b>	<b>0,0080</b>
<b>X<sub>6</sub> (1)</b>	<b>2,0290</b>	<b>1,0270</b>	<b>1,9770</b>	<b>0,0480</b>
X <sub>6</sub> (2)	2,2180	1,3590	1,6320	0,1030
X <sub>6</sub> (3)	-	4612,2030	-0,0030	0,9970
X <sub>7</sub> (1)	-	2095,0990	-0,0100	0,9920
X <sub>7</sub> (2)	-	2095,0990	-0,0090	0,9930
X <sub>8</sub> (1)	1,4240	1,6640	0,8560	0,3920
X <sub>9</sub> (1)	1,9470	1,6810	1,1580	0,2470
X <sub>9</sub> (2)	1,3750	2,1220	0,6480	0,5170
X <sub>9</sub> (3)	-1,8240	1,0430	-1,7490	0,0800
X <sub>10</sub> (1)	0,7970	0,9070	0,8790	0,3790
<b>X<sub>11</sub> (1)</b>	<b>2,1060</b>	<b>0,9550</b>	<b>2,2050</b>	<b>0,0270</b>

Dengan pengujian alpha 5 % atau 0.05, maka kriteria yang diperoleh untuk p-value<0.1 adalah variabel yang signifikan yang akan dibentuk menjadi model logitnya. Dari Tabel 4.16 terlihat bahwa variabel yang signifikan adalah X<sub>1</sub>, X<sub>3</sub>, X<sub>5</sub>, X<sub>6</sub>, dan X<sub>11</sub>. Jika dibuat model dari variabel yang signifikan, maka model regresi Logistiknya menjadi :

$$\pi(x) = \frac{\exp(4.536X_1 + 4.938X_3 - 3.835X_5 + 2.029X_6 + 2.106X_{11})}{1 + \exp(4.536X_1 + 4.938X_3 - 3.835X_5 + 2.029X_6 + 2.106X_{11})} \quad (8)$$

Kemudian menghitung ketepatan klasifikasi dari model Regresi Logistik Biner dengan data Testing 10

% pada Tabel 5.

Tabel 5. Tabel Klarifikasi Model Regresi Logistik

aktual	Biner	
	Prediksi	
	0	1
0	2	3
1	2	4

Dari Tabel 5 didapatkan ketepatan klasifikasi pada Tabel 6 sebagai berikut :

Tabel 6 Ketepatan Klasifikasi Regresi Logistik Biner

Akurasi	APER	Specificity	Sensitivity
54.55%	44.45%	0.57	0.50

### C. Analisis seleksi GA - Regresi Logistik Biner

Tahapan algoritma genetika akan berhenti jika nilai fitness minimum sudah konvergen dari generasi sebelumnya dan selanjutnya. Hasil dari algoritma genetika dengan model dari regresi logistik biner dapat dilihat pada Tabel 7 sebagai berikut:

Tabel 7. Hasil 100 kromosom dan nilai *fitness*

Kromosom	Nilai							Fitness
	X1	X2	X3	...	X9	X10	X11	
1	0	1	0		1	1	1	0,134
2	0	0	1		1	1	1	0,134
.								.
.								.
.								.
99	0	0	1	...	1	1	1	0,134
100	0	0	1	...	1	1	1	0,134

Dari Tabel 7 diberikan 100 kromosom dengan nilai fitness-nya dan hasil seleksi Algoritma Genetika di atas didapatkan 5 variabel yang terpilih yaitu X<sub>2</sub>, X<sub>7</sub>, X<sub>8</sub>, X<sub>9</sub>, X<sub>10</sub> dan X<sub>11</sub> yang terdapat pada Kromosom pertama dengan nilai *fitness* sebesar 0,134. Berikut Tabel 8. akan merangkum hasil ketepatan klasifikasinya.

Tabel 8. Ketepatan Klasifikasi GA-Reglog Biner

Akurasi	APER	Specificity	Sensitivity
90,91%	9,09%	1	0,75

### D. Perbandingan Hasil Klasifikasi

Dari hasil yang didapatkan, maka perbandingan klasifikasi antara Regresi Logistik dan Algoritma Genetika – Regresi Logistik yang dirangkum pada Tabel 9, sebagai berikut :

Tabel 9. Perbandingan Ketepatan Klasifikasi

Metode	Variabel Terpilih	Ukuran Ketepatan Klasifikasi				
		1	2	3	4	5

Regresi Logistik	X <sub>1</sub> , X <sub>2</sub> , ..., X <sub>11</sub>	54,55%	44,45%	0,57	0,5	0,54
GA - Regresi Logistik	X <sub>2</sub> , X <sub>7</sub> , X <sub>8</sub> , X <sub>9</sub> , X <sub>10</sub> , X <sub>11</sub>	90,91%	9,09%	1	0,75	0,875

Ket :

- 1 : Akurasi
- 2 : APER
- 3 : *Specificity*
- 4 : *Sensitivity*
- 5 : *G-Mean*

## KESIMPULAN

Berdasarkan analisis dari hasil dan pembahasan, maka Keakuratan klasifikasi pada Regresi Logistik untuk semua variabel independen yang digunakan adalah akurasi sebesar 54,55 %, Specificity sebesar 0,57, Sensitivity sebesar 0,50, dan G-Mean sebesar 0,68.

Pada analisis seleksi menggunakan Algoritma Genetika ini bertujuan untuk melihat seleksi variabel yang digunakan pada setiap Regresi Logistik. Dan hasil seleksi Algoritma Genetika untuk Regresi Logistik adalah X<sub>2</sub>, X<sub>7</sub>, X<sub>8</sub>, X<sub>9</sub>, X<sub>10</sub>, dan X<sub>11</sub> dengan nilai kesalahannya adalah 0,134. Dan penggunaan menggunakan algoritma genetika ini meningkatkan hasil klasifikasi pada regresi logistik yaitu akurasi sebesar 90,91%, APER sebesar 9,09%, specificity sebesar 1, dan sensitifity sebesar 0,75.

## DAFTAR PUSTAKA

- [1] A. D. Wulansari, "Perbandingan Ketepatan Klasifikasi antara Metode Regresi Logistik dan Klasifikasi Pohon: Suatu Aplikasi pada Kasus Program Wajardikdas 9 Tahun," Aug. 2019.
- [2] R. W.-A. B. Tasikmalaya, "Penerapan Algoritma Genetika Untuk Seleksi Fitur Pada Analisis Sentimen Review Jasa Maskapai Penerbangan Menggunakan Naive Bayes," *EVOLUSI J. Sains dan Manaj.*, vol. 4, no. 1, Jan. 2016.
- [3] B. S. Koentjoro, I. S. Sitanggang, A. K. Makarim, and A. K. Makarim, "Model Simulasi dan Visualisasi Prediksi Potensi Hasil dan Produksi Kedelai di Jawa Timur," *J. Penelit. Pertan. Tanam. Pangan*, vol. 34, no. 3, p. 195, Feb. 2016.
- [4] S. N. Sivanandam and S. N. Deepa, "Genetic Algorithms," in *Introduction to Genetic Algorithms*, Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, pp. 15–37.
- [5] F. Insani and A. R. Ramadhan, "Optimasi

- Biaya Bahan Dan Jasa Pembangunan Rumah Menggunakan Algoritma Genetika,” *Semin. Nas. Teknol. Inf. Komun. dan Ind.*, vol. 0, no. 0, pp. 222–229, Nov. 2019.
- [6] D. C. P. Buani, “Optimasi Algoritma Naïve Bayes dengan Menggunakan Algoritma Genetika untuk Prediksi Kesuburan (Fertility),” *EVOLUSI J. Sains dan Manaj.*, vol. 4, no. 1, Mar. 2016.
- [7] T. Purwa, “Perbandingan Metode Regresi Logistik dan Random Forest untuk Klasifikasi Data Imbalanced (Studi Kasus: Klasifikasi Rumah Tangga Miskin di Kabupaten Karangasem, Bali Tahun 2017),” *J. Mat. Stat. dan Komputasi*, vol. 16, no. 1, p. 58, Jun. 2019.
- [8] I. Ispandi and R. S. Wahono, *Journal of Intelligent Systems.*, vol. 1, no. 2. 2015.
- [9] E. Elyan and M. M. Gaber, “A genetic algorithm approach to optimising random forests applied to class engineered data,” *Inf. Sci. (Ny.)*, vol. 384, pp. 220–234, Apr. 2017.
- [10] V. Trevino and F. Falciani, “GALGO: an R package for multivariate variable selection using genetic algorithms,” *Bioinformatics*, vol. 22, no. 9, pp. 1154–1156, May 2006.
- [11] Y. I. Kurniawan, “Perbandingan Algoritma Naive Bayes Dan C.45 Dalam Klasifikasi Data Mining,” vol. 5, no. 4, pp. 455–464, 2018.
- [12] T. Vafeiadis, K. I. Diamantaras, G. Sarigiannidis, and K. C. Chatzisavvas, “A comparison of machine learning techniques for customer churn prediction,” *Simul. Model. Pract. Theory*, vol. 55, pp. 1–9, Jun. 2015.
- [13] A. Kusumawati, “Efektivitas Pemberian Paclotrazol Terhadap Keseimbangan Pertumbuhan Dua Varietas Kacang Tanah;” 2010.

# Optimalisasi Regresi Logistik Menggunakan Algoritma Genetika Pada Data Klasifikasi

---

## ORIGINALITY REPORT

---

**21** %

SIMILARITY INDEX

**21** %

INTERNET SOURCES

**2** %

PUBLICATIONS

**4** %

STUDENT PAPERS

---

## MATCH ALL SOURCES (ONLY SELECTED SOURCE PRINTED)

---

12%

★ id.123dok.com

Internet Source

---

Exclude quotes  On

Exclude bibliography  On

Exclude matches  < 2%