

ALGORITMA GENETIKA DENGAN *ROULETTE WHEEL SELECTION* DAN *ARITHMETIC CROSSOVER* UNTUK PENGELOMPOKAN

Jessen Yaputra Setiawan ¹⁾ Dyah Erny Herwindiati ²⁾ Tri Sutrisno ³⁾

¹⁾²⁾³⁾ Teknik Informatika Universitas Tarumanagara
Jl. Letjen S.Parman No.1, Jakarta Barat 11440 Indonesia

¹⁾ jessen.yaputra@gmail.com

²⁾ dyahh@fti.untar.ac.id

³⁾ tris@fti.untar.ac.id

ABSTRACT

Genetic algorithms are techniques that can be used to clustering data that has global search characters. This application is made using roulette wheel selection techniques and arithmetic crossover techniques. The purpose of this research is to implement a genetic algorithm that produces good results in clustering image data. The result is clustering flower images with different colors has good results, while clustering flower images with similar colors do not have good results. Several experiments were carried out on each scenario to determine the effect of the parameters used on the fitness value obtained, the result was a clustering with parameter color characteristics, the parameter with the largest fitness value are the number of population = 100, iterations = 200, and mutations = 0.02. while clustering with color plus texture characteristic, the parameter with the largest fitness value are the number population=200, iterations=300, and mutations=0.02.

Key words

Arithmetic Crossover, Fitness, Flower, Genetic algorithms, Roulette wheel selection.,

1. Pendahuluan

Data Mining atau penggalian data merupakan proses kecerdasan metode yang diterapkan untuk mengestrak pola dari suatu data secara otomatis [1]. Di dalam penggalian data tersebut terdapat beberapa fungsi utama salah satunya adalah pengelompokan suatu data atau yang sering disebut dengan *clustering*.

Clustering adalah sebuah proses untuk mengelompokkan data ke dalam beberapa *cluster* atau kelompok sehingga data dalam satu *cluster* memiliki tingkat kemiripan yang maksimum dan data antar *cluster* memiliki kemiripan yang minimum [2]. *Clustering* merupakan salah satu metode data mining

yang bersifat tanpa arahan (*unsupervised*). Yang dimaksud metode *unsupervised* yaitu metode ini diterapkan tanpa adanya latihan (*training*) dan guru (*teacher*) serta tidak memerlukan target output. Didalam *Clustering* terdapat dua macam pencarian, yang pertama adalah pencarian lokal dan yang kedua adalah pencarian global. Sebagian metode *Clustering* menerapkan pencarian lokal sehingga mudah terjebak dalam optimasi lokal. Yang dimaksud dengan pencarian lokal adalah pencarian tersebut memiliki solusi baru yang diwarisi dari iterasi sebelumnya sehingga pusat *cluster* yang ditemukan masih menyerupai iterasi sebelumnya [3]. Hal tersebut akan berdampak buruk apabila pusat *cluster* yang ditunjuk pertama merupakan sebuah *outlier* karena akan mempengaruhi pemilihan pusat *cluster* selanjutnya sehingga pusat *cluster* yang dihasilkan nanti tidak optimal. Berbeda dengan pencarian global, dengan pencarian global pemilihan pusat *cluster* tidak mudah terjebak dalam optimasi lokal sehingga pusat yang dihasilkan dapat lebih optimal. Salah satu metode untuk *Clustering* dengan sifat pencarian global adalah Algoritma Genetika.

Algoritma Genetika adalah suatu algoritma yang terinspirasi oleh suatu teori evolusi yang telah diciptakan oleh Charles Darwin, algoritma ini sering digunakan sebagai alat untuk memecahkan suatu masalah. Algoritma Genetika mencerminkan proses seleksi alam di mana individu yang paling cocok dipilih untuk reproduksi untuk menghasilkan keturunan generasi berikutnya. Berikut adalah tahapan tahapan dari Algoritma Genetika, yang pertama adalah inialisasi populasi awal yang terdiri dari beberapa kromosom, berikutnya adalah menentukan fungsi *fitness* semakin besar *fitness* semakin besar juga kemungkinan kromosom terpilih untuk iterasi berikutnya, selanjutnya adalah seleksi individu berdasarkan nilai *fitness*, selanjutnya adalah *Crossover* atau kawan silang, dilanjutkan dengan *Mutation*. Dengan *Mutation* maka terbentuklah individu-individu

baru yang nantinya akan diproses kembali ketahap seleksi [4].

Tujuan dari penelitian ini adalah menerapkan algoritma genetika sederhana yang menghasilkan pengelompokan yang baik, dengan menggunakan data citra bunga sebagai percobaannya. Bunga dipilih sebagai skenario percobaan karena bunga memiliki variasi warna dan tekstur yang beragam sehingga cocok sebagai skenario percobaan ini.

2. Landasan Teori

2.1 Algoritma Genetika



Gambar 1 Flowchart Algoritma Genetika

Algoritma Genetika adalah algoritma yang berusaha menerapkan pemahaman mengenai evolusi alamiah pada tugas-tugas pemecahan-masalah (problem solving). Pendekatan yang diambil oleh algoritma ini adalah dengan menggabungkan secara acak berbagai pilihan solusi terbaik di dalam suatu kumpulan untuk mendapatkan generasi solusi terbaik [5]. Algoritma genetika dipilih karena algoritma ini memiliki sifat pencarian solusi global sehingga tidak mudah terjebak dalam optimasi lokal. Optimasi lokal sangat dihindari karena optimasi lokal memiliki hasil pengelompokan yang kurang baik dikarenakan pusat *cluster* yang didapat memiliki sifat seperti iterasi sebelumnya yang dimana apabila data observasi terdapat *Outlier* maka hasil pengelompokan akan semakin tidak baik. Berikut adalah tahapan-tahapan dalam Algoritma Genetika :

1. Inisialisasi populasi awal.
2. Menentukan fungsi *fitness*.
3. Seleksi
4. Kawin silang
5. Mutasi

2.2 Populasi Awal

Populasi awal merupakan tahapan awal dalam algoritma genetika, Populasi awal dibangkitkan secara random sehingga diperoleh solusi awal. Populasi ini sendiri terdiri atas individu, didalam individu terdapat kromosom dan didalam kromosom terdapat kumpulan gen. Dalam Algoritma genetika terdapat jenis-jenis golongan kromosom, diantaranya adalah kromosom biner, kromosom *float*, kromosom string, dan kromosom kombinatorial. Dalam hal ini jenis kromosom yang digunakan adalah kromosom *float*. Kromosom jenis *float* ini merupakan kromosom yang disusun dari gen-gen dengan nilai pecahan. Gen integer dapat digolongkan dalam bentuk kromosom ini [6].

2.3 Fungsi *Fitness*

Nilai *fitness* merupakan proses untuk mengevaluasi setiap populasi dengan menghitung nilai *fitness* setiap kromosom dan mengevaluasinya sampai terpenuhi kriteria berhenti. Suatu individu dievaluasi berdasarkan suatu fungsi tertentu sebagai ukuran performansinya [7]. Fungsi objektifitas yang digunakan adalah *Minimum Distance*, fungsi objektifitas *Minimum Distance* digunakan untuk mengecek validitas tiap pusat *cluster* yang dihasilkan tiap kromosom. Semakin kecil nilai objektifitas maka semakin besar nilai *fitness* yang dihasilkan, dengan besarnya nilai *fitness* maka kemungkinan kromosom terpilih pada tahap seleksi semakin besar. Untuk menentukan nilai *fitness* terbaik ditentukan berdasarkan nilai dari *fitness* terbesar pada suatu kromosom. Berikut adalah alur perhitungan nilai *fitness* [8] :

1. Menghitung jarak setiap data asli terhadap centroid yang direpresentasikan dalam kromosom
2. Mengambil nilai minimum dari jarak setiap data ke centroid ke I
3. Yang terakhir adalah dengan Menjumlahkan jarak terpendek data terhadap centroid yang dimiliki masing-masing kromosom sebagai nilai *fitness*.

Berikut adalah rumus-rumus yang digunakan untuk mendapatkan nilai *fitness* :

$$J = \sum_{i=1}^n \min r D(X_i, W_r) \quad (1)$$

$$Fitness = \frac{1}{J} \quad (2)$$

Keterangan :

J = minimum distance

min = jarak terkecil

X_i = data ke-i

W_r = centroid ke-r

n = jumlah data

2.4 Seleksi

Seleksi merupakan proses untuk menentukan individu-individu mana saja yang akan dipilih untuk dilakukan crossover. Proses seleksi yang digunakan adalah metode Roulette Wheel Selection Roulette Wheel Selection dipilih karena proses seleksi menggunakan probabilitas nilai fitness sehingga nilai fitness yang besar memiliki kesempatan untuk terpilih lebih besar dibandingkan dengan nilai fitness yang memiliki nilai lebih kecil. Caranya adalah membangkitkan bilangan acak R dari 0 sampai 1 sebanyak jumlah populasi kromosom. Nilai acak tersebut akan dihitung berdasarkan nilai kumulatif fitness, jika $R[k] < C[1]$ maka kromosom 1 terpilih sebagai induk, selain itu pilih kromosom ke-k sebagai induk dengan syarat $C[k-1] < R < C[k]$ [9].

2.5 Kawin Silang

Proses kawin silang ini merupakan proses untuk menambah keanekaragaman string dalam satu populasi. Operator pindah silang mempunyai peran yang paling penting dalam algoritma genetika karena didalamnya terdapat proses perkawinan (persilangan) gen antara dua individu (*parent*) yang menghasilkan dua individu baru (*offspring*) pada generasi berikutnya. Metode *Crossover* yang digunakan adalah metode *Arithmetic Crossover*, metode ini dipilih karena data berupa bilangan *real*. Berikut adalah rumus perhitungan dengan menggunakan metode *Arithmetic Crossover* [10] :

$$\text{offspring 1} = a * \text{parent1} + (1 - a) * \text{parent2} \quad (3)$$

$$\text{offspring 2} = (1 - a) * \text{parent1} + a * \text{parent2} \quad (4)$$

Keterangan:

Offspring 1 = keturunan pertama

Offspring 2 = keturunan kedua

a = Crossover rate dengan nilai 0 sampai 1

Parent1 = orangtua pertama

Parent2 = orangtua kedua

2.6 Mutasi

Mutasi merupakan proses mengubah nilai dari satu atau beberapa gen dalam suatu kromosom. Mutasi menciptakan individu baru dengan melakukan modifikasi

satu atau lebih gen dalam individu yang sama. Mutasi berfungsi untuk menggantikan gen yang hilang dari populasi selama proses seleksi serta menyediakan gen yang tidak ada dalam populasi awal. Didalam proses mutasi dibutuhkan *Mutation rate* yang berfungsi untuk melakukan mutasi dengan *rate* yang telah ditentukan. *Mutation rate* memiliki nilai range antara 0 sampai 1. Jenis *Mutation* yang digunakan adalah *Random Mutation*, proses mutasi jenis ini adalah dengan mengganti gen yang termutasi dengan nilai acak.

2.7 Manhattan Distance

Manhattan Distance atau taxicab norm merupakan penjumlahan besaran vector suatu ruang, Manhattan Distance adalah cara yang paling alami untuk mengukur jarak antar vektor, yaitu dengan cara menjumlahkan selisih mutlak dari komponen vektor [11]. Manhattan distance ini digunakan untuk mengukur jarak antara data observasi terhadap titik pusat Cluster atau Centroid. Manhattan Distance ini digunakan karena sangat akurat dalam data dengan dimensi yang cukup tinggi. Berikut adalah rumus perhitungan Manhattan Distance [12] :

$$d = \sum_{i=1}^n |X_i - Y_i| \quad (5)$$

Keterangan :

d = jarak

n = jumlah data

i = iterasi ke-i

X_i = vektor b iterasi ke-i

Y_i = vektor a iterasi ke-i

2.8 RGB

RGB adalah suatu model warna yang terdiri atas 3 buah warna merah, hijau, dan biru yang ditambahkan dengan berbagai cara untuk menghasilkan bermacam warna. Oleh karena itu RGB disebut juga *Additive Color*. Jadi data citra RGB adalah data yang berbentuk gambar atau citra yang pada setiap piksel citra mewakili warna yang merupakan kombinasi antara 3 warna dasar yaitu merah, hijau, dan biru. Setiap warna dasar memiliki intensitas tersendiri dengan nilai minimum 0 dan nilai maksimum 255. Dalam penelitian ini ciri data RGB yang digunakan merupakan nilai rata-rata *Red*, *Green*, dan *Blue* data setiap citra bunga.

2.9 GLCM

Gray Level Co-occurrence Matrix (GLCM) adalah teknik untuk memperoleh tekstur citra dengan menggunakan perhitungan pada orde kedua. Pengukuran tekstur pada orde pertama menggunakan perhitungan statistika didasarkan pada nilai pixel citra asli semata, seperti varians, dan tidak memperhatikan hubungan ketetanggaan pixel. Pada orde kedua, hubungan antarpasangan dua pixel citra asli diperhitungkan [13].

Pada hal ini jarak piksel yang digunakan adalah 1 dan orientasi derajat yang digunakan adalah 0, 45, 90, 135 derajat. Berikut adalah *feature* GLCM yang digunakan [14].

1. Contrast
kontras adalah ukuran intensitas atau variasi tingkat keabuan antara piksel dan tetangganya. Dengan rumus :

$$f = \sum_i \sum_j (i - j)^2 \tag{6}$$

Keterangan :
f = *feature contrast*
i = baris ke-i
j = kolom ke-j

2. Correlation
Fitur korelasi menunjukkan ketergantungan linear dari nilai-nilai tingkat keabuan dalam matriks. Dengan rumus :

$$f = \sum_i \sum_j \frac{(i-\mu_x)(j-\mu_y)}{\sigma_x \sigma_y} \tag{7}$$

Keterangan :
f = *feature correlation*
i = baris ke-i
j = kolom ke-j
 μ_x, μ_y = rata-rata px, py
 σ_x, σ_y = standar deviasi px, py

3. Energy
Energi berasal dari Angular Second Moment (ASM). ASM mengukur keseragaman lokal dari tingkat abu-abu. Ketika piksel sangat mirip, nilai ASM akan menjadi besar. Dengan rumus :

$$ASM = \sum_i \sum_j (i, j)^2 \tag{8}$$

$$f = \sqrt{ASM} \tag{9}$$

Keterangan :
ASM = *Angular Second Moment*
i = baris ke-i
j = kolom ke-j
f = *feature energy*

4. Homogeneity
Homogenitas mengukur seberapa dekat distribusi elemen dalam GLCM terhadap diagonal GLCM. Dengan rumus:

$$f = \sum_i \sum_j \frac{1}{1+(i-j)^2} \tag{10}$$

Keterangan :
f = *feature homogeneity*
i = baris ke-i
j = kolom ke-j

3. Hasil Percobaan

Percobaan dilakukan sebanyak tiga kali untuk mengetahui pengaruh parameter algoritma genetika

terhadap hasil pengelompokan, skenario yang dilakukan adalah sebanyak tujuh yaitu sebagai berikut :

1. skenario 1 : 30 mawar merah muda, dan 20 mawar putih
2. skenario 2 : 10 mawar merah tua, dan 20 mawar merah muda
3. skenario 3 : 30 kencana ungu, 10 kamboja jepang, dan 10 mawar putih
4. skenario 4 : 30 kencana ungu, 25 mawar merah muda, 15 mawar putih, dan 10 tabebuia
5. Skenario 5 : 20 jarum tujuh bintang, 10 kencana, 10 mawar putih, 15 tabebuia, dan 30 telang
6. Skenario 6 : 30 aster putih, dan 20 mawar putih
7. Skenario 7 : 15 mandevilla, dan 30 tabebuia

Dari ketujuh skenario diatas akan dilakukan percobaan terhadap masing-masing skenario, pada skenario 1 sampai 5 akan dikelompokkan berdasarkan ciri warnanya, sedangkan skenario 6 dan 7 dikelompokkan berdasarkan ciri warna ditambah tekstur, karena pada skenario 6 dan skenario 7 citra yang digunakan memiliki kesamaan warna. Hasilnya dari ketujuh skenario hanya skenario satu sampai lima yang menghasilkan pengelompokan yang baik dengan menggunakan ciri warna saja. Sedangkan skenario enam dan tujuh belum memperoleh hasil yang baik dengan pengelompokan ciri warna ditambah tekstur.

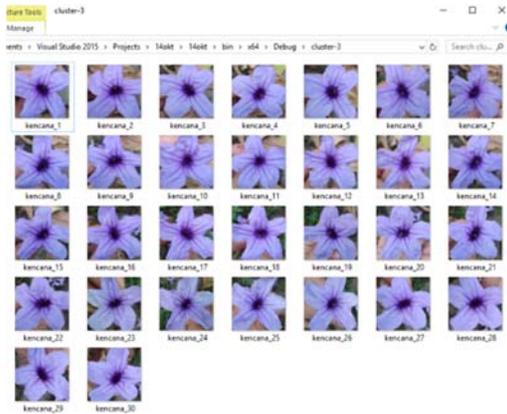
Berikut adalah contoh pengelompokan yang berhasil yaitu hasil pengelompokan data citra bunga skenario tiga, percobaan dua:



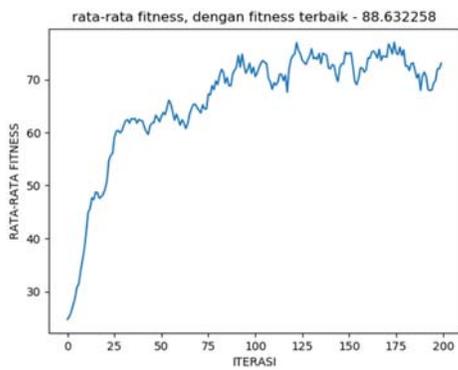
Gambar 2 anggota cluster 1 skenario 3 percobaan 2



Gambar 3 anggota cluster 2 skenario 3 percobaan 2

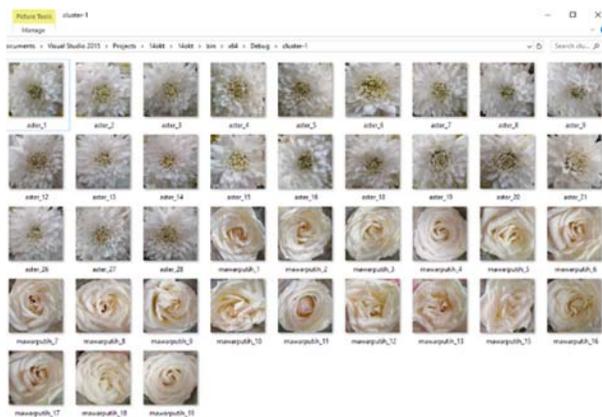


Gambar 4 anggota cluster 3 skenario 3 percobaan 2



Gambar 5 grafik fitness skenario 3 percobaan 2

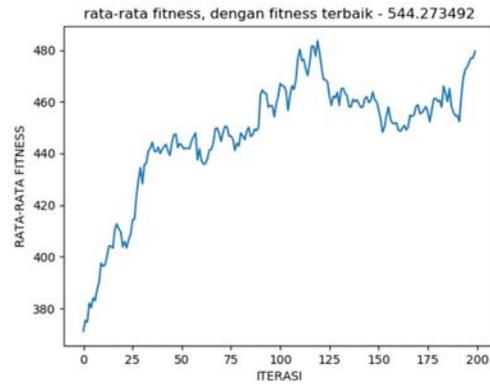
Berikut adalah contoh pengelompokan yang belum memperoleh hasil baik, yaitu hasil pengelompokan data citra bunga skenario enam, percobaan satu dengan ciri warna ditambah tekstur:



Gambar 6 anggota cluster 1 skenario 6 percobaan 1



Gambar 7 anggota cluster 2 skenario 6 percobaan 1



Gambar 8 grafik fitness skenario 6 percobaan 1

Berikut adalah hasil nilai *fitness* terbaik dari setiap percobaan pada tiap skenario, dari tabel ini dapat dilihat pengaruh parameter algoritma genetika terhadap *fitness* terbaik yang didapat:

Tabel 1 Tabel skenario 1

Percobaan	Jumlah populasi	Jumlah iterasi	mutasi	<i>Fitness</i> terbaik
1	50	50	0.02	72,629733
2	100	200	0.02	101,940727
3	100	200	0.2	70,71827

Tabel 2 Tabel skenario 2

Percobaan	Jumlah populasi	Jumlah iterasi	mutasi	<i>Fitness</i> terbaik
1	50	50	0.02	55,299487
2	100	200	0.02	168,052964
3	100	200	0.2	120,676495

Tabel 3 Tabel skenario 3

Percobaan	Jumlah populasi	Jumlah iterasi	mutasi	<i>Fitness</i> terbaik
1	50	50	0.02	63,789062
2	100	200	0.02	88,632258
3	100	200	0.2	57,715196

Tabel 4 Tabel skenario 4

Percobaan	Jumlah populasi	Jumlah iterasi	mutasi	<i>Fitness</i> terbaik
1	50	50	0.02	25,966524
2	100	200	0.02	58,843111
3	100	200	0.2	23,542764

Tabel 5 Tabel skenario 5

Percobaan	Jumlah populasi	Jumlah iterasi	mutasi	<i>Fitness</i> terbaik
1	50	50	0.02	22,167017
2	100	200	0.02	56,820466
3	100	200	0.2	20,856613

Tabel 6 Tabel skenario 6

Percobaan	Jumlah populasi	Jumlah iterasi	mutasi	<i>Fitness</i> terbaik
1	100	200	0.02	544,273492
2	200	300	0.02	570,333980
3	300	400	0.02	565,830259
4	100	200	0.4	469,112270
5	200	300	0.4	476,127146
6	300	400	0.4	463,153989

Tabel 7 Tabel skenario 6

Percobaan	Jumlah populasi	Jumlah iterasi	mutasi	<i>Fitness</i> terbaik
1	100	200	0.02	590,462570
2	200	300	0.02	705,574933
3	300	400	0.02	675,666776
4	100	200	0.4	495,561090
5	200	300	0.4	486,286317
6	300	400	0.4	497,658247

Pada tabel diatas dapat disimpulkan bahwa parameter sangat berpengaruh terhadap nilai *fitness* terbaik yang didapat, terlihat bahwa dengan menggunakan jumlah populasi dan iterasi yang cenderung besar menghasilkan nilai *fitness* yang besar, sebaliknya mutasi yang relatif kecil menghasilkan nilai *fitness* yang besar.

4. Kesimpulan

Berikut adalah kesimpulan yang dapat diambil dalam penelitian ini:

1. Algoritma genetika menghasilkan kinerja yang baik dalam mengelompokkan suatu data, dengan memperhatikan parameter yang digunakan untuk memperoleh hasil yang maksimal. Pada pengelompokan dengan menggunakan ciri warna parameter dengan nilai *fitness* terbesar adalah jumlah populasi=100, iterasi=200, dan mutasi=0,2. Sedangkan untuk pengelompokan dengan menggunakan ciri warna ditambah tekstur parameter dengan nilai *fitness* terbesar adalah jumlah populasi=200, iterasi=300, dan mutasi=0,02.
2. Pengelompokan citra bunga dengan warna yang berbeda memperoleh hasil yang baik sedangkan pengelompokan citra bunga dengan warna yang serupa belum memperoleh hasil yang baik.

REFERENSI

- [1] Jiawei Han, Micheline Kamber, and Jian Pei, DATA MINING Concepts and Techniques, 3rd Edition, (Waltham: Morgan Kaufmann, 2012), h. 8
- [2] Edy Irwansyah, Clustering, <https://socs.binus.ac.id/2017/03/09/clustering>, 21 Agustus 2018
- [3] Rachmat Saidi, ANALISIS DAN IMPLEMENTASI ALGORITMA GENETIKA UNTUK PENINGKATAN KINERJA FUZZY C-MEANS CLUSTERING, <http://repository.telkomuniversity.ac.id/pustaka/94420/analisis-dan-implementasi-algoritma-genetika-untuk-peningkatan-kinerja-fuzzy-c-means-clustering.html>, 22 September 2018
- [4] Dwips, Algoritma Genetika Step by Step, <http://www.inicatan.com/2017/04/11/algoritma-genetika-lengkap-step-by-step/>, 22 September 2018
- [5] Ardy Al-Maqassary, Pengertian Algoritma Genetika, <https://www.e-jurnal.com/2013/09/pengertian-algoritma-genetika.html>, 25 Agustus 2018
- [6] Achmad Basuki, Strategi Menggunakan Algoritma Genetika, <http://basuki.lecturer.pens.ac.id/lecture/StrategiAlgoritmaGenetika.pdf>, 27 September 2018
- [7] Ni Luh Gede Pivin Suwirmayanti, I Made Sudarsana, dan Suta Darmayasa, Penerapan Algoritma Genetika Untuk Penjadwalan Mata Pelajaran, <https://publikasi.dinus.ac.id/index.php/jais/article/download/1255/942>, 25 Agustus 2015
- [8] Ali Ridho Barakbah, Optimasi Titik Pusat K-means dengan Algoritma Genetika, http://rengga.lecturer.pens.ac.id/training_it/soft_computing/day-02/Ali_Ridho/Optimasi%20Titik%20Pusat%20K-means%20dengan%20GA.pdf, 22 September 2018

- [9] Eka Risky Firmansyah, Syukri Sayyid Ahmad, dan Nurul Hikmah Agustin, Algoritma Genetika, <https://ekarisky.com/content/uploads/Algoritma-Genetika.pdf>, 22 September 2018
- [10] Neurodimension, Arithmetic Crossover, http://www.neurodimension.com/genetic/documentation/OptiGenLibraryCPlusPlus/GeneticLibrary/Arithmetic_Crossover.htm, 22 September 2018
- [11] Sarah Iris Garcia, L0 Norm, L1 Norm, L2 Norm, & L-infinity Norm, <https://medium.com/@montjoile/l0-norm-l1-norm-l2-norm-l-infinity-norm-7a7d18a4f40c>, 25 Agustus 2018
- [12] Improved Outcomes Software, Manhattan, http://www.improvedoutcomes.com/docs/WebSiteDocs/Clustering/Clustering_Parameters/Manhattan_Distance_Metric.htm, 22 September 2018
- [13] Abdul Kadir, Teori dan Aplikasi Pengolahan Citra, (Yogyakarta: Andi Offset, 2013), h. 672
- [14] Nourhan Zayed, dan Heba A.Elnemr, Statistical Analysis of Haralick Texture Features to Discriminate Lung Abnormalities, <https://www.hindawi.com/journals/ijbi/2015/267807/>, 17 desember 2018

Jessen Yaputra Setiawan, merupakan mahasiswa program Sarjana S1, program studi Teknik Informatika, Fakultas Teknologi Informasi, Universitas Tarumanagara

Dyah Erny Herwindiati, memperoleh gelar Doktor MIPA, Matematika & Ilmu Pengetahuan Alam dari Institut Teknologi Bandung, Indonesia tahun 2006. Saat ini sebagai Dekan Fakultas Teknologi Informasi Universitas Tarumanagara.

Tri Sutrisno, memperoleh gelar S.Si dari Universitas Diponegoro tahun 2011. Kemudian memperoleh gelar M.Sc dari Universitas Gadjah Mada tahun 2015. Saat ini sebagai Staf pengajar program studi Teknik Informatika, Fakultas Teknologi Informasi Tarumanagara.