

See discussions, stats, and author profiles for this publication at: <https://www.researchgate.net/publication/324038397>

Deteksi Dini Penyakit Gagal Ginjal menggunakan Gabungan Genetic Algorithm dan Fuzzy K-Nearest Neighbor (GAFKNN)

Article · January 2018

CITATIONS

0

READS

375

3 authors, including:



Dian Eka Ratnawati
Brawijaya University

47 PUBLICATIONS 16 CITATIONS

[SEE PROFILE](#)



Candra Dewi
Brawijaya University

40 PUBLICATIONS 31 CITATIONS

[SEE PROFILE](#)

Some of the authors of this publication are also working on these related projects:



Prediksi Jumlah Pengangguran Terbuka di Indonesia menggunakan Metode Genetic-Based Backpropagation [View project](#)



Optimasi Asupan Makanan Harian Ibu Hamil Penderita Hipertensi Menggunakan Algoritme Genetika [View project](#)

Deteksi Dini Penyakit Gagal Ginjal menggunakan *Gabungan Genetic Algorithm dan Fuzzy K-Nearest Neighbor (GAFKNN)*

Muhyidin Ubaidillah¹, Dian Eka Ratnawati², Candra Dewi³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹ubaid.muhyidin@gmail.com, ²dian_ilkom@ub.ac.id, ³dewi_candra@ub.ac.id

Abstrak

Penyakit gagal ginjal kronik adalah salah satu jenis penyakit tidak menular tapi mematikan. Menurut Riset Kesehatan Dasar (2013), penyakit ini di Indonesia memiliki nilai prevensi nilai kurang lebih 0.2 persen. Namun beberapa orang masih belum sadar kalau mereka telah mengalami penyakit ini, dan penyakit gagal ginjal mereka telah pada tahap gagal ginjal kronik sehingga salah satu pengobatannya adalah dengan melakukan cuci darah. Padahal bila orang tersebut masih pada gagal ginjal stadium awal atau stadium dua, masih dapat melakukan terapi tanpa cuci darah. Selain itu beberapa orang masih malas untuk melakukan konsultasi, sehingga dibutuhkan sebuah program agar orang tersebut dapat mengetahui kondisi mereka dan termotivasi untuk melakukan *checkup* ke dokter. Dari permasalahan tersebut dibutuhkan sebuah deteksi dini yang dapat dilakukan oleh klasifikasi. Salah satu metode yang dapat melakukan klasifikasi adalah Fuzzy KNN, namun Fuzzy KNN memiliki kelemahan yaitu menentukan nilai k dan m yang menghasilkan nilai optimal. Sehingga dilakukan penggabungan dengan GAs. Dari hasil penggabungan tersebut program dapat menghasilkan akurasi yang cukup optimal yang mencapai 98%, dengan parameter pada GAs yaitu populasi 40, generasi 15, CR 0.5 dan MR 0.8.

Kata Kunci: *fuzzy k – nearest neighbor, fuzzy knn, algoritme genetika, GAs, gagal ginjal*

Abstract

Chronic kidney disease is one type of non-infectious but deadly disease. According to Basic Health Research (2013), this disease in Indonesia has a value of prevention value of approximately 0.2 percent. But some people are still not aware that they have experienced this disease, and their kidney failure disease has been at the stage of chronic renal failure so that one of the treatment is to do dialysis. Whereas if the person is still in early or stage 2 kidney failure, can still do therapy without dialysis. In addition, some people are still lazy to consult, so it takes a program so that people can know their condition and motivated to do a checkup to the doctor. From these problems requires an early detection that can be done by classification. One method that can classify is Fuzzy KNN, but Fuzzy KNN has a weakness that is determining the value of k and m that yield optimal value. So do the merger with GAs. From the results of the merger the program can produce a fairly optimal accuracy of 98%, with parameters on the GAs of population 40, generation 15, CR 0.5 and MR 0.8.

Keywords: *fuzzy k – nearest neighbor, fuzzy knn, genetic algorithm, GAs, kidney disease*

1. PENDAHULUAN

Kesehatan merupakan hal yang didambakan oleh seluruh manusia. Karena dengan kesehatan tersebut tubuh dapat melakukan segala aktivitas dengan maksimal. Namun setiap orang tetap memiliki kemungkinan terserang penyakit. Salah satu penyakit yang sangat mematikan antara lain adalah gagal ginjal (Riskesdas, 2008). Gagal ginjal termasuk dalam penyakit yang tidak menular (PTM).

Menurut Riset Kesehatan Dasar (Riskesdas) (2013), nilai prevalensi di seluruh Indonesia untuk penyakit gagal ginjal memiliki nilai rata rata berkisar kurang lebih 0.2 persen. Gagal ginjal menyerang organ ginjal dari makhluk hidup sehingga ginjal mereka tidak berfungsi sebagai mana fungsinya. Akibat dari penyakit ini ginjal tidak dapat melakukan penyarangan elektrolit, tidak dapat membuat *urine* atau air seni dan masih banyak lagi (Smeltzer & Bare, 2008). Penyakit ini juga dapat menyerang siapa saja tidak memandang umur. Kategori umur

sendiri dibagi menjadi 3 yakni Muda, Parubaya dan Tua (Kusumadewi & Hartati, 2013). Salah satu perawatan bagi orang tersebut adalah melakukan cuci darah yang dilakukan 3 kali selama seminggu (Smeltzer & Bare, 2008).

Namun, dari beberapa orang masih malas untuk melakukan *checkup*. Dikarenakan terdapat dua proses yaitu cek lab dan konsultasi dokter, dan kebanyakan lokasi dari dokter spesialis dan lab lumayan jauh sehingga beberapa orang malas untuk kembali ke dokter spesialis. Untuk permasalahan tersebut dapat menggunakan program yang dapat melakukan deteksi dini terhadap penyakit gagal ginjal. Deteksi dini tersebut dapat diselesaikan dengan sebuah klasifikasi, yang dimana klasifikasi akan melakukan identifikasi terhadap seseorang, apakah termasuk dalam penderita gagal ginjal kronik atau tidak. Klasifikasi sendiri terdapat beberapa metode antara lain *K-Nearest Neighbor* (KNN), *Fuzzy K-Nearest Neighbor* (FKNN) dan *Genetic Algorithm Fuzzy K-Nearest Neighbor* (GAFKNN).

Dalam beberapa penelitian banyak yang telah membandingkan metode KNN, Fuzzy KNN, dan GAFKNN. Penelitian pertama yaitu yang dilakukan oleh Juan dan Wenbin (2008) untuk membandingkan hasil akurasi dari metode Fuzzy KNN dan GAFKNN, dan hasil dari penelitian tersebut adalah Fuzzy KNN nilai akurasi 94.67% dan GAFKNN nilai akurasi 98.69%. Penelitian lain yang dilakukan adalah penelitian oleh Aggarwal dan Rupal (2015), penelitian tersebut melakukan perbandingan antara KNN, Fuzzy KNN dan GAFKNN yang memiliki nilai akurasi masing masing 87%, 93% dan 98%. Kemudian juga terdapat penelitian yang dilakukan Meristika, Ridok dan Muflikhah (2013), yang menyimpulkan bahwa KNN sangat lemah untuk data yang banyak sedangkan Fuzzy KNN dapat dengan mudah untuk mengatasi data yang banyak. Namun Fuzzy KNN memiliki kelemahan lain yaitu dalam penentuan nilai k dan m , sehingga diperlukan sebuah algoritma yang dapat melakukan pencari k dan m .

Dari penjelasan diatas maka penelitian ini akan menggunakan algoritma gabungan yaitu GA dengan Fuzzy KNN atau sering disebut GAFKNN untuk melakukan deteksi dini penyakit gagal ginjal.

2. LANDASAN TEORI

2.1 Fuzzy K-Nearest Neighbor (Fuzzy KNN)

Fuzzy KNN sendiri merupakan sebuah pengembangan dari K-NN. Algoritma ini lebih memfokuskan class membership ke dalam suatu *pattern* (pola) daripada menempatkan *pattern* (pola) ke dalam kelas yang tersedia (Alfiyanto, Ratnawati, & Ridok, 2014). Karena Fuzzy KNN berfokus terhadap *class membership* menggunakan konsep dari logika *fuzzy*, untuk menentukan derajat keanggotaan dari setiap kelas. Inti pokok dari algoritma Fuzzy KNN ini adalah untuk memberikan derajat keanggotaan sebagai satu fungsi untuk mencari jarak terdekat dari tetangga terdekatnya dan menentukan kelas yang memungkinkan (Wijayanti, 2013).

Menurut Keller, Gray dan Givens (1985) Dalam Fuzzy KNN terdapat beberapa tahapan untuk melakukan klasifikasi, berikut ini adalah tahapan tersebut:

- **Tentukan jarak terdekat**

Di Tahap ini akan mencari jarak terdekat dengan data *input*. Untuk mencari jarak terdekat dapat menggunakan salah satu rumus pencarian jarak yaitu *Euclidean Distance*, yang dapat dilihat pada Persamaan 1.

$$d(x, y) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (1)$$

Keterangan:

X_i : Data latih ke $- i$
 Y_i : Data Uji ke $- i$

- **Mencari tetangga terdekat.**

Setelah mendapatkan jarak maka yang dilakukan adalah mencari tetangga dengan nilai jarak yang terpendek sejumlah k yang telah ditentukan.

- **Hitung jarak antara tetangga yang terpilih dengan data latih lain.**

Setelah mendapatkan tetangga terdekat sejumlah k , langkah selanjutnya adalah menghitung jarak antara tetangga yang terpilih dengan semua data latih. Untuk mencari jarak tersebut dapat menggunakan Persamaan 1.

Sama seperti tahap ke dua, yaitu mengambil tetangga terkecil sejumlah k per data. Jika $k = 5$, maka tetangga pertama memiliki 5 tetangga, tetangga kedua memiliki 5 tetangga dan seterusnya.

- **Proses klasifikasi Fuzzy K-NN**

Dalam tahap klasifikasi ini terdapat beberapa tahapan yaitu:

Menentukan nilai keanggotaan kelas ke- i pada tetangga ke- j dihitung menggunakan Persamaan 2.

$$u_j(x) = \begin{cases} 0.51 + \left(\frac{n_j}{K}\right) \times 0.49 \rightarrow j = i \\ \left(\frac{n_j}{K}\right) \times 0.49 \rightarrow j \neq i \end{cases} \quad (2)$$

Keterangan:

U_{ij} : nilai keanggotaan tetangga

n_j : jumlah tetangga kelas- i pada himpunan k-nn

Menentukan nilai membership tiap kelas pada data uji, untuk proses tersebut dapat menggunakan rumus pada Persamaan 3. setelah nilai membership tiap kelas telah didapatkan, kemudian dibandingkan dengan kelas yang lain.

$$u_i(x) = \frac{\sum_{j=1}^k u_{ij} \left(\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}} \right)}{\sum_{j=1}^k \left(\frac{1}{\|x - x_j\|^{\frac{2}{m-1}}} \right)} \quad (3)$$

Keterangan:

$U_i(x)$: nilai membership kelas pada data uji x

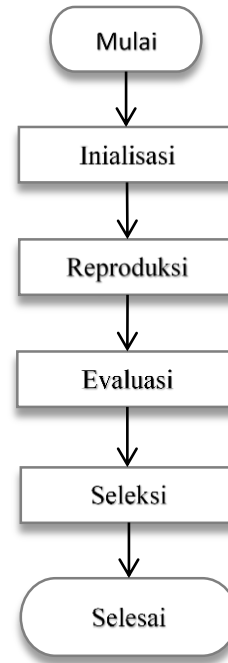
m : bobot pangkat

• Memilih membership kelas tertinggi.

Tahap ini akan memilih kelas dengan membership yang tertinggi, yang nantinya akan dijadikan kelas dari data yang diuji.

2.2 Genetic Algorithm (GAs)

Genetic Algorithms (GAs) atau sering disebut algoritma genetika merupakan tipe Evolution Algorithm yang lebih populer daripada yang lain. GAs dapat menyelesaikan atau memecahkan berbagai masalah yang kompleks, algoritma ini banyak digunakan dalam berbagai bidang yang memiliki masalah optimasi yang model matematikanya sangat sulit dibangun (Mahmudi, 2013). Siklus dari *Genetic Algorithm* Menurut Mahmudi (2013), terdapat beberapa siklus dalam GAs. Berikut ini adalah siklus dari GAs yang dapat dilihat pada Gambar 1.



Gambar 1 Siklus *Genetic Algorithm*
Sumber: (Mahmudi, 2013)

Berikut ini adalah penjelasan pada Gambar 1:

• Inisialisasi

Tahap ini adalah awal, dimana untuk ditahap ini membangkitkan himpunan individu secara acak yang terdiri dari string chromosome dan ditempatkan pada penampungan yang disebut populasi.

• Reproduksi

Di tahap ini berfungsi untuk menghasilkan keturunan dari induk yang terpilih dari tahap inisialisasi, dan ditempatkan di offspring. Di tahap ini ada 2 jenis reproduksi yaitu *Mutation* dan *Crossover*.

• Evaluasi

Di tahap ini berfungsi untuk mencari nilai *fitness* dari tiap individu yang telah terbuat dan beberapa keturunan yang telah terbuat.

• Seleksi

Di tahap ini berfungsi untuk memilih individu yang dapat bertahan pada proses atau generasi selanjutnya.

2.3 Gabungan GAs dan Fuzzy KNN

Didalam Fuzzy KNN terdapat dua parameter yang menentukan tingkat akurasi yang tinggi. Dua parameter tersebut adalah nilai jumlah tetangga terdekat (k) dan nilai bobot (m). Namun di Fuzzy KNN terdapat masalah yaitu kesulitan dalam menentukan nilai k dan nilai m ,

jika menggunakan nilai acak (*random*) yang memiliki range yang besar membutuhkan waktu yang *relative* cukup lama (Zhang Juan & He Wenbin, 2008).

GAs merupakan sebuah algoritma yang dipercaya dapat menyelesaikan masalah pencarian terhadap nilai yang memiliki range besar atau masalah yang lebih kompleks. Dan GAs merupakan sebuah algoritma yang sangat sesuai untuk melakukan masalah pencarian nilai parameter yang optimal (Goldberg, 1989). Dari penjelasan Fuzzy KNN dan GAs tersebut, maka untuk mengatasi masalah pencarian nilai parameter yaitu nilai k dan m tersebut dapat digunakan algoritma GAs. Sehingga *Genetic Algorithm* dan *Fuzzy K-Nearest Neighbor* dilakukan penggabungan.

Dalam pencarian dua parameter tersebut dapat menggunakan *binary code* untuk pembentukan kromosom awal. Kromosom tersebut terdiri dari 2 genes. Genes pertama (g_1) digunakan untuk nilai k dan genes kedua (g_2) digunakan untuk nilai m . Kemudian secara teori nilai m berkisar 1 sampai tak hingga, namun semakin tak hingga nilai maka berakibat kegagalan sistem melakukan pendeteksian. Sehingga untuk batasan nilai m adalah kisaran 1 sampai bilangan hingga, sedangkan nilai k berkisar dari 1 sampai jumlah sampel (Juan & Wenbin, 2008). Maka dapat dirumuskan sebagai berikut $k \in [1, a]$ dan $m \in [1, b]$, dimana a bilangan integer antara 1 sampai jumlah data dan b bilangan asli 1 sampai angka terhitung.

Karena kromosom nantinya menggunakan *binary code*, sesuai dengan penelitian Zhang Juan dan He Wenbin (2008) maka untuk mendapatkan nilai k dan m dapat melihat Persamaan 4 dan Persamaan 5.

$$k = 1 + decimal(s_1) \times \frac{a-1}{2^{n_1}-1} \quad (4)$$

Untuk lebih jelas dapat melihat keterangan Persamaan 4 sebagai berikut:

S_1 : hasil decode dari genes kedua
 a : jumlah data
 n_1 : jumlah bits dari jumlah tetangga.

$$m = 2 + decimal(s_2) \times \frac{b-1}{2^{n_2}-1} \quad (5)$$

Untuk lebih jelas dapat melihat keterangan Persamaan 5 sebagai berikut:

S_2 : hasil decode dari genes kedua

b : nilai bilangan terhitung yang ditentukan

n_2 : jumlah bits dari jumlah tetangga.

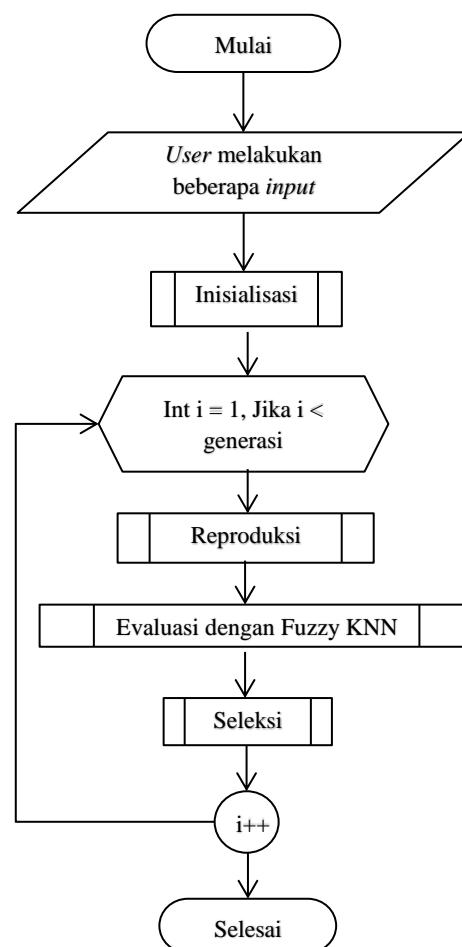
Menurut Juan dan Wenbin (2008) untuk mempertimbangkan nilai tak hingga yang digunakan sebagai nilai b , dapat menggunakan Persamaan 6 dan Persamaan 7. Kemudian nilai fitness didapatkan dari nilai akurasi tiap individu yang telah tercipta.

$$\frac{b-1}{2^{n_2}-1} \leq 0.01 \quad (6)$$

$$n_2 \geq \lg_2(100(b-1)+1) \quad (7)$$

3. PERANCANGAN

Penelitian ini termasuk kedalam penelitian implementatif. Dan hasil dari penelitian ini adalah program untuk melakukan deteksi dini terhadap penyakit gagal ginjal. Untuk alur dari program ini dapat melihat Gambar 2.



Gambar 2 Alur program

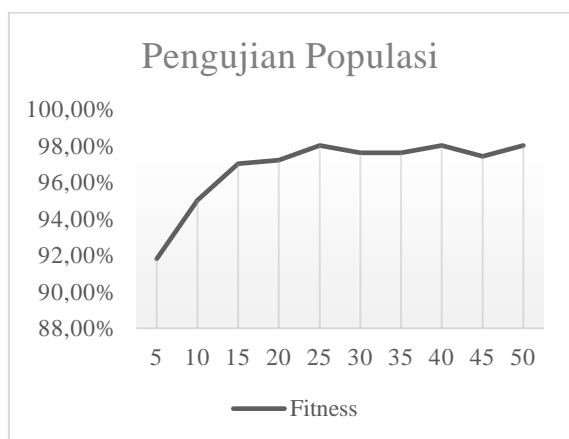
Pada Gambar 2, pertama *user* akan menginputkan data uji, jumlah popsize, jumlah

generasi, jumlah *crossover rate* dan jumlah *mutation rate*. Kemudian program akan melakukan proses inisialisasi, di tahap ini akan melakukan pembuatan kromosom awal, pencarian jarak data uji dan data latih kemudian data latih dan data latih. Setelah itu program akan melakukan proses reproduksi untuk menghasilkan kromosom baru. Tahap selanjutnya yaitu melakukan pencarian nilai akurasi menggunakan Fuzzy KNN dan nilai tersebut nantinya akan digunakan sebagai nilai *fitness*. Dan yang terakhir adalah mencari kromosom yang akan lolos pada generasi selanjutnya.

4. PENGUJIAN

4.1 Pengujian populasi

Pengujian populasi dilakukan untuk mencari populasi yang akan menghasilkan nilai *fitness* yang optimal. Untuk pengujian ini digunakan varian kelipatan 5, dan setiap varian akan dilakukan pengujian sebanyak 10 kali. Kemudian rata rata *fitness* didapatkan dari rata rata nilai *fitness* setiap kali percobaan. Untuk nilai generasi = 5, *Crossover Rate* (CR) = 0.5 dan *Mutation Rate* (MR) = 0.5. Untuk melihat hasil pengujian populasi dapat dilihat pada Gambar 3.



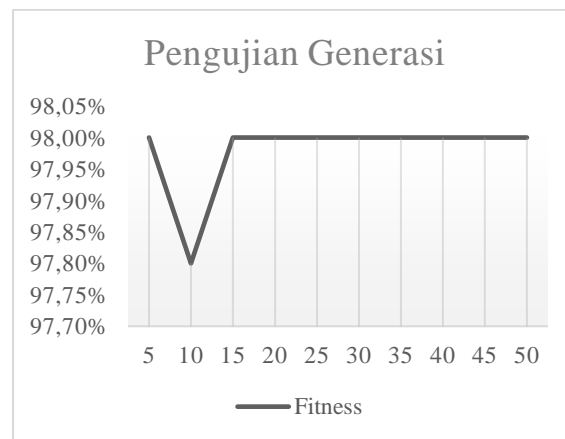
Gambar 3 hasil pengujian populasi

Dari gambar 3, didapatkan nilai *fitness* akan semakin meningkat jika jumlah populasi bertambah. Dan nilai tertinggi *fitness* adalah 98%.

4.2 Pengujian generasi

Pengujian generasi dilakukan untuk mencari generasi yang akan menghasilkan nilai *fitness* yang optimal. Untuk pengujian ini digunakan varian kelipatan 5, dan setiap varian

akan dilakukan pengujian sebanyak 10 kali. Kemudian rata rata *fitness* didapatkan dari rata rata nilai *fitness* setiap kali percobaan. Untuk nilai populasi = 5, CR = 0.5 dan MR = 0.5. Untuk melihat hasil pengujian populasi dapat dilihat pada Gambar 4.

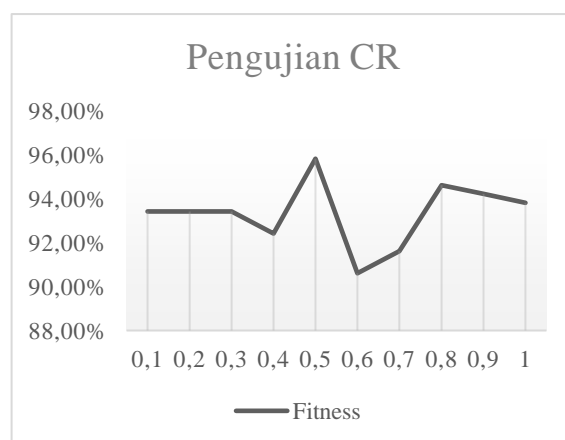


Gambar 4 hasil pengujian generasi

Dari gambar 3, nilai *fitness* akan mengalami konvergen setelah generasi ke 15 dengan nilai rata rata *fitness* 98.00%

4.3 Pengujian CR

Pengujian CR dilakukan untuk mencari CR yang akan menghasilkan nilai *fitness* yang optimal. Untuk pengujian CR digunakan varian generasi kelipatan 0.1, dan setiap varian akan dilakukan pengujian sebanyak 10 kali. Kemudian rata rata *fitness* didapatkan dari rata rata nilai *fitness* setiap kali percobaan. Untuk nilai populasi = 5, generasi = 5 dan mr = 0.5. Untuk melihat hasil pengujian populasi dapat dilihat pada Gambar 5.



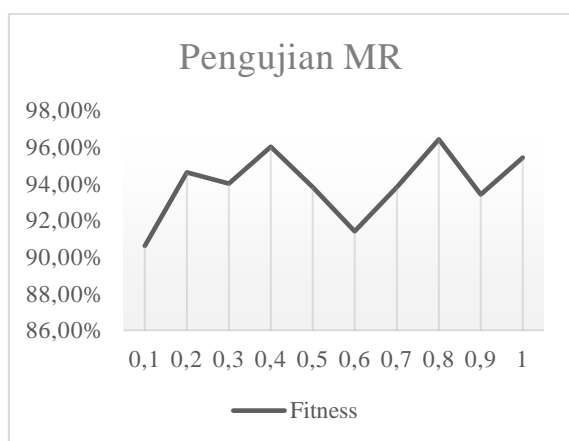
Gambar 5 hasil pengujian CR

Dari gambar 5, didapatkan nilai *fitness* akan mengalami kenaikan pada nilai 0.5 dengan nilai

fitness 95.80%. Dan mengalami penurunan nilai *fitness* setelah melebihi nilai 0.5.

4.4 Pengujian *Mutation Rate* (MR)

Pengujian ini dilakukan untuk mencari MR yang akan menghasilkan nilai *fitness* yang optimal. Untuk pengujian MR, sama seperti pengujian CR, namun berbeda di inputan saja, yaitu nilai populasi = 5, generasi = 5 dan CR = 0.5. Untuk melihat hasil pengujian populasi dapat dilihat pada Gambar 6.



Gambar 6 hasil pengujian MR

Dari gambar 6, nilai *fitness* terendah pada MR 0.1 dengan nilai 90.60%, dan setelah itu nilai *fitness* mengalami kenaikan dan puncaknya pada MR 0.8 dengan nilai 96.40%.

4.5 Perbandingan Fuzzy KNN dengan GAFKNN

Untuk melihat hasil perbandingan antara Fuzzy KNN dengan Fuzzy KNN dengan gabungan GAs, dapat melihat pada Tabel 1.

Tabel 1 Perbandingan Fuzzy KNN dan GAFKNN

Fuzzy KNN		GAFKNN	
Nilai Akurasi	Waktu	Nilai Akurasi	Waktu
96.0 %	0 ms	98.0 %	422 ms

5. KESIMPULAN

Dari penelitian yang telah dilakukan, penulis dapat memberikan kesimpulan sebagai berikut.

Untuk tingkat akurasi yang diciptakan oleh GAFKNN lebih optimal dibandingkan dengan Fuzzy KNN dengan perbandingan akurasi Mahmudi, W. F. 2013. *Algoritma Evolusi*. PTIIK Universitas Brawijaya, Malang.

Meristika, Y. S., Ridok, A., & Muflikhah, L.

98.00% dengan 96.00% namun dari segi estimasi waktu yang dihasilkan Fuzzy KNN lebih cepat daripada GAFKNN.

Dari pengujian parameter yang telah disusun dilakukan nilai akan optimal ketika mendapatkan nilai parameter yang tepat. Dan nilai dari parameter tersebut adalah 40 untuk populasi, 15 untuk generasi, 0.5 untuk CR dan 0.8 untuk MR.

Dan untuk kedepannya dibutuhkan penambahan fitur untuk mengatasi *missing value*. Karena masih error saat data mengalami *missing value*.

DAFTAR PUSTAKA

- Aggarwal, P., & Rupal, N. 2015. Data Mining of CRM dataset using GAF Algorithm. *International Journal of Advanced Trends in*, 2, 46-50.
- Alfiyanto, E., Ratnawati, D. E., & Ridok, A. 2014. Pengklasifikasian Kritik dan Saran dengan Menggunakan Fuzzy K-Nearest Neighbor (FK-NN). *Skripsi*. PTIIK Universitas Brawijaya, Malang.
- Goldberg, D. E. 1989. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Reading.
- Kaur, J., & Bhatia, R. 2016. Customer Relationship Management Relationship Management Classification by Hybridizing Genetic Algorithm and Fuzzy K-Nearest Neighbor. *International Journal of Computer Applications*, 147, 13-17.
- Kaur, S., & Arora, J. (2016). Implementing Hybrid GA & Fuzzy KNN to Optimize the Classification of Fuzzy KNN. *International Journal of Science and Research*, 5(6), 1988-1991.
- Keller, J., Gray, M., & Givens, J. 1985. A fuzzy k-nearest neighbor algorithm. *IEEE Trans System Man Cybernet.*, 580-585.
- Maghfirah, N., Dewi, C., & Hidayat, N. (2014). Implementasi Algoritma Fuzzy K-Nearest Neighbor (FK-NN) Untuk Mengetahui Tingkat Resiko Penyakit Gagal Ginjal. *DORO Repository Jurnal Mahasiswa*, 4.
2013. Perbandingan K-Nearest Neighbor dan Fuzzy K-Nearest Neighbor pada Diagnosis Penyakit

- Diabetes Melitus. *Skripsi*. PTIIK Universitas Brawijaya, Malang.
- Riskesdas. 2008. *Laporan Hasil Riset Kesehatan Dasar (Riskesdas)*. Badan Penelitian dan Pengembangan Kesehatan Republik Indonesia, Jakarta.
- Riskesdas. 2013. *Laporan Hasil Riset Kesehatan Dasar (Riskesdas)*. Badan Penelitian dan Pengembangan Kesehatan Kementerian RI, Jakarta.
- Smeltzer, & Bare. 2008. *Buku Ajar Keperawatan Medikal Bedah*. EGC, Jakarta.
- Zhang Juan, N. Y., & He Wenbin. 2008. Using Genetic Algorithm to Improve Fuzzy k-NN. *International Conference on Computational Intelligence and Security*, 475-479.