

Pengembangan Metode Klasterisasi Data Berbasis *Hybrid Improved Artificial Bee Colony (IABC)* dan *K – Harmonic Means*

**Tegar Palyus Fiqar¹, Saiful Bahri Musa², Fitriah Maharani Humaira³
I Made Widiartha⁴, Darlis Herumurti⁵, Agus Zainal Arifin⁶**

¹ Informatika, Institut Teknologi Kalimantan, Balikpapan. Email: tegar@itk.ac.id

² Teknik Informatika, Politeknik Gorontalo, Gorontalo. Email: saiful.bm@poligon.ac.id

³ Teknik Listrik Industri, Politeknik Negeri Madura, Sampang. Email: humaira@poltera.ac.id

⁴ Teknik Informatika, Universitas Udayana, Bali. Email: imadewidiartha@cs.unud.ac.id

⁵ Informatika, Institut Teknologi Sepuluh Nopember, Surabaya. Email: darlis@its-sby.edu

⁶ Informatika, Institut Teknologi Sepuluh Nopember, Surabaya. Email: agus.za@its-sby.edu

Abstract

One of data grouping process method is k-harmonic clustering method (KHM) which has a relatively short and simple process. However, it has a weakness at cluster center point. Randomly formed cluster center point causes difficulty to converge solutions. One way to solve the problem at the cluster center point requires a method which has a global solution for KHM. The method is Improved artificial bee colony (IABC), improvement of artificial bee colony (ABC) method based on behavior patterns of honey bee colony in food searching process. Advantage of the IABC method is able to have more optimum global solution. This research proposes a new method of clustering using improved artificial bee colony and K-Harmonic means (IABC-KHM) to optimize the center point in clusters that lead to global solution. In this study, the IABC is functioned for finding the most optimum cluster center point for the data clustering process using KHM. Furthermore, the performance test of the IABC-KHM clustering method is compared with ABC and ABC-KHM methods on three different datasets. The result of mean value of best function of IABC-KHM method of Iris dataset is 152,87, Contraceptive Method Choice dataset is 918,54, and Wine dataset is 31,01. Moreover, the result of the average value of the best F-Measure method IABC-KHM Iris dataset is 0.90, the Contraceptive Method Choice dataset is 0.41, the Wine dataset is 0.95. To conclude, IABC-KHM method has successfully optimized the position of cluster center point that directs the cluster result which has global solution.

Keywords: Cluster, K-Harmonic Means, Artificial Bee Colony, Improved Artificial Bee Colony.

Abstrak

Salah satu metode klasterisasi yang sering digunakan untuk proses pengelompokan data ialah metode k-harmonic clustering (KHM), hal ini dikarenakan KHM memiliki proses yang relatif singkat dan sederhana, namun KHM memiliki kelemahan pada titik pusat klaster yang terbentuk. Titik pusat klaster yang dibentuk secara random menyebabkan solusi yang diperoleh sulit untuk konvergen. Salah satu cara untuk mengatasi permasalahan pada titik pusat klaster membutuhkan suatu metode yang dapat memberikan solusi global ke dalam metode KHM. Metode improved artificial bee colony (IABC) merupakan peningkatan dari metode artificial bee colony (ABC) berdasar pada pola perilaku koloni lebah madu dalam proses pencarian makanan. Kelebihan dari metode IABC mampu memiliki solusi global yang lebih optimum. Penelitian ini mengusulkan metode baru klasterisasi menggunakan improved artificial bee colony dan K-Harmonic means (IABC-KHM) untuk optimalisasi titik pusat pada klaster yang mengarah pada solusi global. Pada penelitian ini, IABC berperan untuk mencari titik pusat klaster yang paling optimal selanjutnya proses klasterisasi data menggunakan KHM. Uji coba kinerja metode klasterisasi IABC-KHM ini dibandingkan dengan metode ABC dan ABC-KHM pada tiga dataset berbeda. Hasil nilai rerata fungsi tujuan terbaik metode IABC-KHM dataset Iris sebesar 152,87, dataset Contraceptive Method Choice sebesar 918,54, dataset Wine sebesar 31,01. Hasil nilai rerata fungsi F-Measure terbaik metode IABC-KHM dataset Iris sebesar 0,90, dataset Contraceptive Method Choice sebesar 0,41, dataset Wine sebesar 0,95. Kesimpulan dari hasil metode IABC-KHM menunjukkan perolehan posisi titik pusat klaster yang lebih optimal dengan mengarahkan hasil klaster yang memiliki solusi global.

Kata Kunci: Klasterisasi, K-Harmonic Means, Artificial Bee Colony, Improved Artificial Bee Colony.

1. Pendahuluan

Klasterisasi data (*clustering*) merupakan proses pengelompokan sekumpulan data atau sekumpulan objek ke dalam klaster yang sama dengan kemiripan *intra-class* yang tinggi sedangkan kemiripan *inter-class* rendah. Metode klaster yang populer dapat dikelompokkan menjadi dua cara yaitu dengan *partitional clustering* dan *hierarchical clustering* (Tan, 2006).

Metode *hierarchical clustering* merupakan metode pengelompokan data yang membangun hirarki dari klaster, metode ini dibangun dengan sebuah dendrogram (*binary tree*) yang pada awalnya terdapat elemen-elemen independen selanjutnya dilakukan proses penggabungan menjadi kelompok-kelompok yang memiliki nilai kedekatan tertentu. Pada metode *partitional clustering* setiap klaster memiliki titik pusat (*centroid*) yang selanjutnya dihitung fungsi tujuan dengan meminimumkan jarak dari seluruh data terhadap *centroid* dari masing-masing klaster.

Salah satu metode *partitional clustering* yang sering digunakan penelitian ialah *K-harmonic Means* (KHM). *K-harmonic Means* pertama kali diusulkan oleh Zhang, dkk (Zhang, 1999), metode ini merupakan perbaikan dari metode *K-Means clustering* (Pen, 1999) yang mana *K-means clustering* memiliki kelemahan pada sensitifitas keakuratan hasil yang sangat berpengaruh pada inisialisasi titik awal pusat klaster. Selanjutnya, Hammerly dan Elkan (2002) melakukan penelitian serta memperbaiki kinerja dari KHM yaitu pada nilai rata-rata harmonik yang dihasilkan dari elemen data terhadap pusat klaster. Rata-rata harmonik yang diterapkan pada KHM telah mampu mengurangi permasalahan inisialisasi, namun hal ini masih terdapat kemungkinan terjadi pada permasalahan lokal optimal (Yang, 2009).

Metode *Improved artificial bee colony* (IABC) (Gao, 2011) merupakan perbaikan dari *artificial bee colony* (ABC) dalam proses pencarian solusi global (Karaboga, 2009). Kedua metode ini diinspirasi dari pola perilaku koloni lebah madu dalam proses pencarian makanan. Pada kedua metode ini sama-sama menggunakan tiga jenis lebah yaitu lebah pengintai (*scout*), lebah penunggu (*onlooker bee*), dan lebah pekerja (*employed bee*). Perbaikan yang dilakukan oleh peneliti Gao pada tahun 2011 memiliki kelebihan yang terdapat pada dua bagian yaitu pada proses inisialisasi dan proses skema pencarian. Proses inisialisasi dan proses skema pencarian yang dibangun IABC bertujuan untuk meningkatkan hasil yang konvergen dan meningkatkan kualitas dari hasil akhir. Pada kesimpulan akhir penelitian Gao, IABC terbukti telah memiliki kualitas hasil yang lebih baik dibandingkan metode ABC untuk konvergensi untuk permasalahan lokal optimal. Salah satu cara untuk menyelesaikan permasalahan lokal optimal yang terjadi pada KHM yaitu dengan menggunakan metode yang memiliki karakteristik pencarian solusi global yang diterapkan pada metode KHM dalam hal ini metode IABC. Penelitian ini mengusulkan metode klasterisasi baru yang menggabungkan metode *improved artificial bee colony* dan *K-Harmonic Means* (IABC-KHM). Metode IABC-KHM ini diharapkan memberikan solusi global yang optimum serta mengatasi permasalahan posisi titik pusat data klaster.

2. Metode Penelitian

Pada bagian metode penelitian dibahas tentang konsep dasar dari teori-teori yang dipakai dalam makalah ini. Pemaparan tersebut meliputi konsep dasar dari analisis klaster, *Artificial Bee Colony* (ABC), *K-Harmonic means clustering*, dan *Improved Artificial Bee Colony* (IABC).

2.1. Analisis Klaster

Analisis klaster merupakan suatu analisa untuk mengelompokkan objek berdasarkan informasi deskriptif tentang objek tersebut dan relasi antar objek yang tersedia (Tan, 2006). Tujuan analisis klaster adalah membuat membentuk keseragaman yang tinggi antara masing-masing anggota dari sekumpulan data serta membentuk kesegeraman yang rendah terhadap antar kelompok. Secara umum, tipe klasterisasi dapat dibedakan menjadi dua yaitu klasterisasi secara partisi (*unnnested*) dan hirarki (*nested*) (Tan, 2006).

2.2. *K-Harmonic Means Clustering*

Metode *K-Harmonic Means* (KHM) merupakan salah satu metode pengelompokan data dengan berdasarkan *centeroid-based* yang mana berpusat pada sebuah vektor sebagai pusat klaster. Klaster-

kluster yang terbentuk menggunakan metode KHM mengalami penyempurnaan kalster secara iteratif berdasarkan posisi pusat kalster terhadap posisi dari masing-masing data serta terhadap posisi antar masing-masing kluster. Pada penentuan keanggotan suatu elemen data terhadap suatu kluster memiliki perbedaan untuk proses pembentukannya antara metode kluster *K-harmonic Means* dan metode kluster *K-Means*, jika pada metode kluster *K-Means* penentuan keanggotaan data pada suatu kluster ditentukan dengan perhitungan jarak minimum terhadap kluster, pada kluster *K-Harmonic Means* menggunakan perhitungan rata-rata harmoniknya. Misal $C = \{c_l \mid l = 1, 2, \dots, K\}$ dimana K adalah jumlah titik pusat kluster, dan $X = \{x_i \mid i = 1, 2, \dots, N\}$ dimana N adalah jumlah data.

$$\text{MIN}\{\|x - c\|^2 \mid c \in C\} \rightarrow \text{HA}\{\|x - c\|^2 \mid c \in C\} = \frac{|C|}{\sum_{c \in C} \frac{1}{\|x - c\|^2}} \quad (1)$$

Fungsi tujuan $\text{Perf}_{KM}(X, C)$ untuk *K-Means* dapat dipaparkan sebagai berikut:

$$\text{Perf}_{KM}(X, C) = \sum_{i=1}^N \text{MIN}\{\|x_i - c_l\|^2 \mid l = 1, \dots, K\}. \quad (2)$$

Jika fungsi $\text{MIN}()$ diganti dengan fungsi rata-rata harmonik $\text{HA}()$ maka didapatkan:

$$\text{Perf}_{KM}(X, C) = \sum_{i=1}^N \text{HA}\{\|x_i - c_l\|^2 \mid l = 1, \dots, K\} = \sum_{i=1}^N \frac{K}{\sum_{l=1}^K \frac{1}{\|x_i - c_l\|^2}}. \quad (3)$$

Hammerly menggunakan sebuah parameter p untuk mengganti *squared* pada persamaan 3, sehingga fungsi tujuan KHM menjadi (Hammerly, 2002):

$$\text{KHM}(X, C) = \sum_{i=1}^N \frac{K}{\sum_{l=1}^K \frac{1}{\|x_i - c_l\|^p}}. \quad (4)$$

Perbedaan antara proses klasterisasi dengan menggunakan metode *K-Means* dan metode KHM terletak pada nilai fungsi tujuan yang diperoleh. Jika pada metode *K-Means* fungsi tujuan diasumsikan setiap elemen data mempunyai bobot yang sama (statis), sedangkan metode KHM setiap elemen data diberi nilai bobot yang tidak statis berdasarkan dengan rata-rata harmoniknya. Penentuan rata-rata nilai harmonik ditentukan dengan cara pemberian bobot yang besar pada elemen data jika posisinya terletak jauh dari pusat kluster, berlaku sebaliknya jika elemen data terletak dengan posisi pusat kluster maka akan diberikan nilai bobot yang kecil. Hal ini sangat penting pada metode kluster KHM untuk menghindari terbentuknya suatu area yang berisikan banyak titik pusat kluster. Penerapan prinsip yang ada pada metode KHM ini memiliki peran untuk mengurangi sensitifitas terhadap inisialisasi awal titik pusat kluster yang terjadi pada metode kluster *K-Means* (Gungor, 2007). Beberapa tahapan proses kluster dengan menggunakan metode *K-Harmonic Means clustering* (Yang, 2009) yaitu:

1. Pemilihan posisi titik pusat kluster awal dari sekumpulan elemen data ditentukan dengan cara random.
2. Cari nilai fungsi tujuan dengan menggunakan persamaan (4), dimana variabel p merupakan input parameter, variabel p memiliki nilai sebesar $p \geq 2$.
3. Hitung nilai keanggotaan $m(c_l/x_i)$ untuk masing-masing posisi titik pusat kluster c_l menggunakan persamaan (5).

$$m(c_l \mid x_i) = \frac{\|x_i - c_l\|^{-p-2}}{\sum_{l=1}^K \|x_i - c_l\|^{-p-2}} \quad (5)$$

4. Cari nilai bobot $w(x_i)$ berdasarkan persamaan (6) untuk masing-masing x_i

$$w(x_i) = \frac{\sum_{l=1}^K \|x_i - c_l\|^{-p-2}}{\left(\sum_{l=1}^K \|x_i - c_l\|^{-p}\right)^2} \quad (6)$$

5. Hitung posisi titik pusat kluster dari semua elemen data berdasarkan nilai keanggotaan dan bobot yang diperoleh dari tahap sebelumnya untuk setiap titik pusat kluster c_j . Penentuan posisi titik pusat kluster menggunakan persamaan (7).

$$c_l = \frac{\sum_{i=1}^N m(c_l | x_i) \cdot w(x_i) \cdot x_i}{\sum_{i=1}^N m(c_l | x_i) \cdot w(x_i)} \quad (7)$$

6. Lakukan tahapan ke dua sampai kelima sehingga memperoleh nilai fungsi tujuan, tahapan ini dilakukan hingga tidak terdapat perubahan nilai fungsi tujuan yang signifikan pada setiap iterasi.
7. Tentukan label masing-masing elemen data x_i pada kluster terhadap titik pusat kluster c_j sesuai dengan nilai keanggotaan yang diperoleh x_i terhadap c_j .

Pada setiap elemen data x_i dinyatakan anggota dari kluster dari titik pusat kluster c_j apabila memiliki syarat nilai keanggotaan $m(c_j/x_i)$ lebih besar dibandingkan dengan nilai keanggotaan elemen data tersebut terhadap titik pusat kluster lainnya. Pada metode KHM nilai keanggotaan $m(c_j/x_i)$ sangat berguna ketika elemen data-data yang ada antar kluster tidak *well sparated* atau terpisah secara baik (Jiang, 2010).

2.3. *Artificial Bee Colony*

Artificial bee colony (ABC) merupakan salah satu metode optimasi yang sering digunakan untuk proses komputasi, metode ini terinspirasi dari perilaku lebah madu dalam proses pencarian makanan (nektar). Metode ABC juga termasuk dalam kecerdasan kolektif (*swarm intelligent*) yang mana kecerdasan yang dibangun dari sekumpulan object secara Bersama-sama. Salah satu kelebihan dari metode ABC ialah memiliki kemampuan optimasi penyelesaian masalah tanpa batasan sehingga diperlukan parameter tertentu yang menandakan ketercapaian dari sebuah target optimum. Metode ABC memiliki empat komponen penting antara lain lebah pekerja, lebah penunggu dan lebah *scout* serta sumber makanan (*food source*). Ada dua model perilaku pada lebah madu yang diadopsi metode ABC, Pertama yaitu proses perekrutan lebah menuju sumber makanan, kedua yaitu proses ditinggalkannya sumber makanan oleh lebah madu pekerja. Pada setiap sumber nektar hanya terdapat satu ekor lebah pekerja, sehingga jumlah lebah madu pekerja yang dimiliki oleh kawanan tersebut sama dengan jumlah sumber makanan/nektar yang berada terdekat dan berada disekitar sarang lebah madu. Adapun langkah-langkah secara umum dari tahapan metode ABC (Karaboga, 2009) seperti berikut.

1. Inisialisasi populasi sumber makanan dan populasi lebah.
2. Ulangi proses (a-d).
 - a. Tempatkan lebah-lebah madu pekerja menuju sumber nektar/makanan.
 - b. Letakkan lebah penunggu pada sumber makanan berdasarkan jumlah nektarnya.
 - c. Kirim lebah *scout* ke daerah-daerah pencarian baru untuk memperoleh sumber makanan baru.
 - d. Mengingat sumber makanan terbaik yang telah ditemukan oleh lebah *scout*.
3. Berhenti hingga persyaratan terpenuhi.

2.4. *Improved Artificial Bee Colony (IABC)*

Pada algoritma ABC memiliki kecepatan konvergensi untuk mencapai tujuan lebih lambat ketika menangani masalah unimodal. Selain itu, algoritma ABC juga mudah terjebak dalam lokal optima ketika menyelesaikan masalah multimodal yang kompleks (Karaboga, 2009). Oleh karena itu, keterbaharuan penelitian yang berkembang terkait dengan metode ABC ialah berfokus untuk mempercepat konvergensi dan menghindari lokal optima. Penelitian yang dikembangkan (Gao, 2011) berupaya untuk mencapai kedua tujuan tersebut. Keluaran dari penelitian tersebut menghasilkan metode optimasi baru yaitu IABC, Metode IABC sendiri terinspirasi oleh metode *differential evolution* (DE). Bagian pertama yang diadopsi metode DE dan diterapkan oleh metode IABC terletak pada strategi mutasi yaitu "*DE/best/1*" dan "*DE/rand/1*", pada IABC solusi tersebut digunakan pada penentuan kandidat sumber makanan, maka diperoleh solusi perbaikan *search equation* yaitu "*ABC/best/1*" dan "*ABC/rand/1*". Bagian kedua yang diadopsi dari DE pada IABC ialah parameter baru m , inisialisasi parameter m pada metode DE digunakan sebagai parameter kendali pertumbuhan populasi sedangkan pada IABC parameter m digunakan untuk parameter kendali jumlah dari selisih elemen antara kandidat sumber makan dan

sumber makanan yang telah diperoleh. Tabel 1 menunjukkan perbedaan fungsi antara metode ABC yang telah tersedia dan metode IABC yang mengadopsi metode DE.

Tabel 1. Fungsi pencarian sumber makanan metode ABC dan IABC

<i>ABC</i>	<i>IABC</i>
$v_{\alpha,\beta} = y_{\alpha,\beta} + z_{\alpha,\beta}(y_{\alpha,\beta} - y_{\gamma,\beta})$	" <i>ABC/best/1</i> " $\Rightarrow v_{\alpha,m} = y_{best,m} + z_{\alpha,\beta}(y_{\alpha,m} - y_{r1,m})$ " <i>ABC/rand/1</i> " $\Rightarrow v_{\alpha,m} = y_{r1,m} + z_{\alpha,\beta}(y_{\alpha,m} - y_{r2,m})$

(8)

Ket
 $v_{\alpha,\beta} / v_{\alpha,m}$: posisi kandidat sumber makanan
 $y_{\alpha,\beta} / y_{best,m} / y_{r1,m} / y_{\alpha,m} / y_{r2,m}$: Sumber makanan yang sudah terjelajahi
 $z_{\alpha,\beta}$: nilai random bernilai -1 hingga 1

Langkah-langkah IABC sebagai berikut (Gao, 2011):

1. Inisialisasi: *Preset* parameter m , probabilitas selektif p , ukuran populasi dan limit.
2. Buat inisialisasi populasi dan hitung nilai-nilai fungsi dari populasi, proses pembentukan inisialisasi populasi menggunakan persamaan *chaotic system* persamaan (9)

$$ch_{\phi+1} = \eta * ch_{\phi}(1 - ch_{\phi}), \quad ch_{\phi} \in (0,1), \phi = 0,1,2, \dots, K. \quad (9)$$

ϕ merupakan iterasi dari *chaotic system* dan K merupakan jumlah iterasi maksimum dari *chaotic system*. η merupakan parameter kontrol dari *choatic system*.

3. Dihasilkan solusi baru $v_{\alpha,m}$ untuk lebah pekerja kemudian dilakukan evaluasi. Terapkan proses *greedy selection* untuk semua lebah pekerja. Pada tahap ini menggunakan persamaan yang tersaji pada Tabel 1 untuk IABC.
4. Hitung nilai peluang p_i untuk $y_{best,m}$ dan $y_{r1,m}$ menggunakan persamaan (10) untuk mendapatkan solusi baru $v_{\alpha,m}$ (persamaan 8).

$$P_i = \frac{fit_i}{\sum_{i=1}^{SN} fit_i} \quad (10)$$

5. Tentukan solusi yang ditinggalkan oleh lebah *scout*, jika ada maka ganti dengan solusi $y_{\alpha,\beta}$ baru menggunakan persamaan (11).

$$y_{\alpha,\beta} = y_{min,\beta} + rand[0,1] * (y_{max,\beta} - x_{min,\beta}) \quad (11)$$

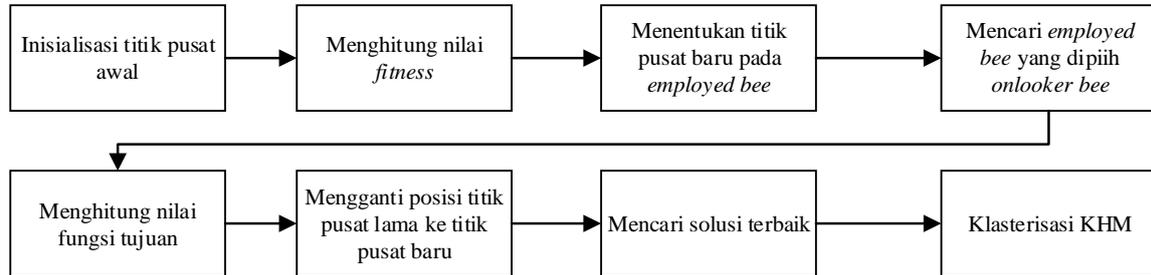
6. Jika kriteria pemberhentian terpenuhi, stop. Jika tidak maka kembali ke langkah ke-2.

3. Metode Usulan

Pada penelitian ini mengusulkan metode baru klasterisasi data yaitu metode IABC-KHM. Metode usulan ini dikembangkan dengan dua metode dari penelitian terpisah yang sudah ada, metode penelitian pertama terkait dengan metode klasterisasi yaitu KHM, metode penelitian kedua yaitu metode optimasi yang mana metode itu bertujuan untuk pencarian titik optimum dari suatu fungsi atau persamaan, metode yang digunakan yaitu IABC, metode IABC merupakan pengembangan dan perbaikan dari metode optimasi ABC. Secara garis besar skema dari usulan metode IABC-KHM ditunjukkan seperti pada Gambar 1. Metode IABC-KHM bertujuan melakukan klasterisasi data yang mana metode IABC memiliki peran untuk melakukan optimalisasi terhadap titik pusat klaster data yang digunakan pada metode klasterisasi KHM. Proses optimasi yang dilakukan pada pusat klaster terletak saat fase lebah yang sebagai titik pusat klaster akan menggunakan tetangga lebah pekerja sebagai acuan untuk bereksplorasi pada seluruh ruang pencarian.

Pada tahap pertama *dataset* yang dibaca antara lain jumlah fitur, jumlah data dan jumlah kelas yang digunakan. Inisialisasi awal dilakukan penentuan nilai maksimal dan nilai minimal setiap fitur pada masing-masing kelas dari data yang digunakan. Setelah memperoleh nilai maksimal dan minimal selanjutnya nilai ini digunakan untuk menentukan inisialisasi titik pusat awal. Proses inisialisasi ini menggunakan algoritma 1. Tahapan selanjutnya menghitung nilai *fitness* dari titik pusat klaster yang dihasilkan oleh masing-masing lebah. Nilai *fitness* dalam fase lebah ini diperoleh dari perhitungan *cost* terhadap posisi pusat klaster yang telah dihasilkan. Nilai *cost* ini didapat dengan mencari nilai total rata-

rata harmonik yang dihasilkan terhadap jarak antara titik pusat kluster yang terbentuk dan titik elemen data. Setelah tahapan inialisasi titik pusat dan perhitungan *fitness*, selanjutnya masuk pada fase *employed bee*. Pada fase ini menentukan titik pusat baru dari setiap *employed bee*. Proses penentuan titik pusat ini sesuai dengan algoritma 2.



Gambar 1. Blok diagram sistem

Algoritma 2. Penentuan titik pusat baru

```

1. For (i=1:i<=JumKls)
2.   rnd = rand;
3.   For (j=1:j<=JumKls)
4.     r1 = randi(JumKls);
5.     r2 = randi(JumKls);
6.     R = 1-2*rand;
7.     Tentukan Posisi Terbaik Secara Minimum : best = find(CenEB(:,i,a) == min(min(CenEB(:,i,a))));
8.   For (c=1:c<=JumFit)
9.     If rnd < prob (prob = 0.25)
10.      Tentukan Titik Pusat Berdasarkan Rand/1
11.      Baru(j,c,i) = CenEB(j,c,r1) + R*(CenEB(j,c,i) - CenEB(j,c,r2));
12.    Else
13.      Tentukan Titik Pusat Berdasarkan Best/1
14.      Baru(j,c,i) = CenEB(j,c,best) + R*(CenEB(j,c,i) - CenEB(j,c,r1));
15.    End If
16.  End For
17. End For
18. End For
  
```

Algoritma 1. Inialisasi titik pusat awal

```

1. Tentukan nilai parameter p
2. Menghitung Nilai Maksimum Dan Minimum kelas dan fitur dari Sumber Makanan(Data)
3. For (i=1:i<=jumKls)
4.   For(j=1:j<=jumKls)
5.     For(k=1:k<=JumFtr)
6.       R = 0+(1-0)*rand(1,1);
7.       For(i=1:i<=10) R = p*R*(1-R);
8.       Tentukan Titik Pusat1 CenT1(j,k,i) = R*(Max(j,k) - Min(j,k)) + Min(j,k)
9.     End For
10.   End For
11. End For
12. For (i=1:i<=jumKls)
13.   For(j=1:j<=jumKls)
14.     For(k=1:k<=JumFtr)
15.       Tentukan Titik Pusat2 CenT2(j,k,i) = (Max(j,k) - CenT1(j,k,i)) + Min(j,k)
16.     End For
17.   End For
18. End For
12. For (i=1:i<=jumKls)
13.   For(j=1:j<=jumKls)
14.     For(k=1:k<=JumFtr)
15.       Tentukan Titik Pusat Initalisasi CenT(j,k,i) = Max(CenT1,CenT2)
16.     End For
17.   End For
18. End For
  
```

Selanjutnya fase *onlooker bee*. langkah pertama pada fase ini diperlukan penentuan probabilitas dari solusi yang dihasilkan dari masing-masing *employed bee*. Penentuan probabilitas ini dilakukan untuk

memilih *employed bee* yang diproses pada tahapan selanjutnya. Penentuan probabilitas yang dilakukan menggunakan metode *roulette wheel* berdasarkan solusi (titik pusat kluster) yang dihasilkan oleh *employed bee*. Setelah proses penentuan maka proses pencarian titik pusat oleh *onlooker bee*. Proses pencarian titik pusat dilakukan dengan menggunakan algoritma 2. Setelah titik pusat diperoleh, selanjutnya dilakukan perhitungan fungsi tujuan, fungsi tujuan ini bertujuan untuk menentukan jarak kedekatan antar kluster. Semakin dekat jarak datanya terhadap pusat kluster, maka semakin baik solusi yang ditemukan oleh lebah. Metode *Euclidean distance* digunakan pada penelitian untuk melakukan perhitungan jarak antara data kluster maupun terhadap titik pusat kluster. Perhitungan jarak yang paling optimal diperoleh, selanjutnya tahapan pembaharuan titik pusat yang diperoleh dari titik pusat *onlooker bee*, sehingga titik pusat yang diperoleh oleh lebah *employed bee* akan diganti dengan hasil perolehan titik pusat *onlooker bee*.

Tahapan terakhir IABC adalah pencarian solusi terbaik dari sekumpulan lebah yang memiliki nilai optimal. Lebah inilah yang diproses pada tahapan klasterisasi selanjutnya dengan menggunakan KHM. Tahapan klasterisasi KHM diawali dengan menghitung fungsi tujuan, yang mana pada fungsi tujuan dilakukan perhitungan jarak titik pusat dengan menggunakan metode *Euclidean*. Selanjutnya dilakukan proses perhitungan keanggotaan dari masing-masing data dalam *dataset* terhadap kelas. Setelah proses perhitungan keanggotaan maka dilakukan perhitungan nilai bobot dari setiap data, nilai bobot ini digunakan dalam perhitungan posisi dari titik pusat selanjutnya, proses ini dilakukan hingga batas iterasi yang digunakan terpenuhi. Langkah berikutnya menentukan keberadaan anggota dalam kluster dan langkah terakhir melakukan transformasi dari kluster menjadi kelas label. Keluaran dari proses kluster ini yaitu data x_i menjadi anggota dari sebuah kluster c_i .

4. Akuisisi *Dataset* dan Skenario Uji Coba

Pada penelitian ini menggunakan *dataset* acuan yang sering digunakan untuk data kluster. *Dataset* yang digunakan antara lain: *dataset* Wine, Iris dan *Contraceptive Method Choice (CMC)*. Penelitian ini menggunakan *dataset* yang diperoleh dari *UC Irvine (UCI) Machine Learning Repository* dengan alamat *website ftp://ftp.ics.uci.edu/pub/machine-learning-databases/*. *Dataset* yang digunakan memiliki tipe numerik. Setiap *dataset* memiliki jumlah data dan jumlah fitur yang berbeda-beda. Informasi mengenai *dataset* yang digunakan penelitian seperti nama *dataset*, jumlah kelas, jumlah fitur dan jumlah data disajikan pada Tabel 2.

Pada masing-masing *dataset* yang digunakan penelitian ini dibagi menjadi data *training* dan data *testing*. Porsi yang digunakan dalam pembagian data *training* sebesar 80% dan data *testing* sebesar 20%. Porsi data *training* dan data *testing* masing-masing kelas pada setiap *dataset* disajikan pada Tabel 3. Data *training* digunakan untuk melihat kinerja dari metode yang diusulkan dan metode pembanding yang digunakan pada proses klasterisasi data. Data *testing* atau data uji digunakan untuk menguji kelas label yang diperoleh dari proses klasifikasi setelah terjadinya proses pembelajaran dengan menggunakan data *training* terhadap titik pusat kluster.

Tabel 2. Informasi *Dataset*

<i>Dataset</i>	Jumlah Fitur	Jumlah Kelas	Jumlah Data
<i>Iris</i>	4	3	150
<i>Contraceptive Method Choice (CMC)</i>	9	3	1473
<i>Wine</i>	13	3	178

Tabel 3. Informasi *Dataset Iris*

<i>Dataset</i>	Kelas	Jumlah Data	Jumlah Data	Jumlah Data
		<i>Training</i>	<i>Testing</i>	
<i>Iris</i>	1	40	10	50
	2	40	10	50
	3	40	10	50

<i>Dataset</i>	<i>Kelas</i>	<i>Jumlah Data Training</i>	<i>Jumlah Data Testing</i>	<i>Jumlah Data</i>
CMC	1	503	126	629
	2	267	67	334
	3	408	102	510
Wine	1	47	12	59
	2	57	14	71
	3	38	10	48

Tahapan uji coba dirancang dengan beberapa skenario untuk menguji kinerja metode. Skenario uji coba dibuat berdasarkan penggunaan parameter pada masing-masing metode ABC, ABC-KHM dan IABC-KHM. Pada metode ABC, parameter yang digunakan mengacu pada penelitian Zhang (2009), parameter yang digunakan yaitu nilai limit, nilai limit ini merepresentasikan penentuan sumber makanan yang ditinggalkan (*abandoned food source*) telah dilewati oleh lebah *scout*, lebah *scout* memiliki peran untuk menemukan solusi baru yang bersifat global. Nilai limit yang digunakan pada penelitian Zhang sebesar 10 sedangkan pada skenario ujicoba penelitian ini nilai limit dirancang bervariasi yaitu bernilai 5, 10, dan 20. Hal ini bertujuan untuk melihat pengaruh dari penentuan nilai limit terhadap hasil *clustering* yang diperoleh. Pada metode KHM, parameter yang digunakan yaitu nilai p , nilai p merupakan nilai parameter yang digunakan untuk menghasilkan nilai *objective function* (Yang, 2009), nilai p yang digunakan pada skenario uji coba yaitu bernilai 2, 3, dan 4. Pada skenario metode ABC-KHM dan IABC-KHM, parameter yang diubah-ubah yaitu nilai p , nilai limit, nilai jumlah *cycle* maksimum. Pada setiap skenario metode klasterisasi yang diuji dilakukan sebanyak 10 kali.

Proses evaluasi kinerja dilakukan dengan membandingkan hasil metode usulan IABC-KHM dengan metode pembandingan ABC-KHM, serta ABC klasik. Tolok ukur yang digunakan untuk pengukuran evaluasi kinerja metode klaster antaralain *F-measure*, *running time* serta fungsi tujuan. *F-measure* merupakan nilai pengukuran *precision* dan *recall* antara kelas label sebenarnya dengan kelas label klasterisasi yang diperoleh dari data masukan. Semakin besar nilai *F-measure* maka semakin mirip antara hasil klaster dengan kelas label dari tiap data dalam klaster. Nilai *F-measure* ini mengindikasikan bagaimana performa suatu metode untuk menghasilkan suatu anggota hasil klaster yang sesuai dengan anggota dari setiap kelas label yang ada. Persamaan yang digunakan untuk mengukur *F-measure* seperti pada persamaan 12 (Handl, 2003).

$$F(\alpha, \gamma) = \frac{(\beta^2 + 1) \cdot (p(\alpha, \gamma) \cdot r(\alpha, \gamma))}{\beta^2 \cdot p(\alpha, \gamma) + r(\alpha, \gamma)} \quad (12)$$

$r(\alpha, \gamma) = n_{\alpha, \gamma} / n_{\alpha}$ dan $p(\alpha, \gamma) = n_{\alpha, \gamma} / n_{\gamma}$ dimana $n_{\alpha, \gamma}$ adalah jumlah elemen data dari kelas α yang berada pada klaster γ , n_{α} adalah jumlah elemen data kelas α yang diharapkan sebagai *query*, n_{γ} adalah jumlah elemen data dari klaster γ yang dihasilkan oleh *query*, agar memperoleh nilai pembobotan ideal antara *recall* dan *precision* pada perhitungan nilai *F-measure* maka menggunakan nilai $\beta=1$. *Running time* merupakan salah satu tolok ukur parameter yang digunakan dalam evaluasi dengan melihat durasi waktu yang dibutuhkan untuk masing-masing metode dalam melakukan proses klasterisasi data. Perhitungan *running time* yang dilakukan pada penelitian ini menggunakan fungsi perhitungan waktu yang disediakan MATLAB. Fungsi Tujuan merupakan nilai rata-rata penjumlahan harmonik antara titik elemen data terhadap seluruh titik pusat klaster. Nilai fungsi tujuan ini digunakan untuk mencari posisi optimal dari titik pusat klaster. Semakin kecil nilai fungsi tujuan maka semakin baik klaster yang terbentuk.

5. Hasil dan Pembahasan

Hasil dari penelitian menggunakan evaluasi kinerja metode dengan menggunakan fungsi tujuan, *F-measure*, *running time*, dan kesalahan kelas (*misclassification*). Tabel 4 menunjukkan hasil evaluasi kinerja dengan fungsi tujuan, semakin kecil nilai dari fungsi tujuan maka semakin baik, karena semakin kecil nilai fungsi tujuan maka titik pusat klaster semakin optimal. Tabel 5 menunjukkan hasil evaluasi kinerja dengan fungsi *F-measure*, semakin besar nilai dari fungsi *F-measure* maka semakin baik, karena

semakin besar nilai fungsi *F-measure* maka semakin mirip antara hasil kluster dengan kelas label dari tiap data dalam kluster. Tabel 6 menunjukkan hasil evaluasi kinerja dengan proses *running time*, semakin kecil nilai dari proses *running time* maka semakin baik, karena semakin kecil nilai proses *running time* maka semakin cepat proses klasterisasi yang diterapkan pada suatu metode. Tabel 7 menunjukkan hasil evaluasi kinerja dengan kesalahan klasterisasi, semakin kecil nilai dari kesalahan klasterisasi maka semakin baik, karena semakin kecil nilai kesalahan klasterisasi maka semakin sedikit data yang salah pada suatu kelas.

Tabel 4. Rerata Fungsi Tujuan

<i>Dataset</i>	Parameter (<i>p</i>)	Nilai Fungsi Tujuan		
		ABC	ABC-KHM	IABC-KHM
<i>Iris</i>	2	162,16	152,87	152,87
	3	179,2	154,40	154,40
	4	173,15	172,48	209,88
<i>CMC</i>	2	1183,98	978,83	980,43
	3	1102,58	947,29	947,40
	4	1214,88	917,95	918,54
<i>Wine</i>	2	77,91	68,69	68,69
	3	56,21	45,67	45,67
	4	39,06	31,01	31,01

Hasil dari penelitian fungsi tujuan, fungsi *F-measure*, proses *running time* dan kesalahan kelas diperoleh dari sepuluh kali percobaan pada setiap masing-masing *dataset* dan setiap nilai parameter, selanjutnya setiap nilai yang dihasilkan dirata-rata seperti ditunjukkan Tabel 4, Tabel 5, Tabel 6 dan Tabel 7. Nilai pada masing-masing *dataset* dan setiap nilai parameter yang dibandingkan menunjukkan kinerja terbaik dari ketiga metode.

Kinerja fungsi tujuan pada *dataset CMC* pada Tabel 4, menunjukkan secara umum metode ABC-KHM lebih baik dibandingkan dengan metode IABC-KHM dan ABC, tetapi metode IABC-KHM lebih baik dibandingkan ABC. Turunnya kinerja metode IABC-KHM dibandingkan ABC-KHM ini disebabkan oleh data yang digunakan proses *training* memiliki nilai yang sangat mirip antar kelas. Oleh karena itu untuk proses klasterisasi yang memiliki karakteristik seperti data *CMC* disarankan menggunakan metode ABC-KHM.

Tabel 5. Rerata Fungsi *F-Measure*

<i>Dataset</i>	Parameter (<i>p</i>)	Nilai <i>F-Measure</i>		
		ABC	ABC-KHM	IABC-KHM
<i>Iris</i>	2	0,88	0,90	0,90
	3	0,89	0,90	0,90
	4	0,89	0,90	0,90
<i>CMC</i>	2	0,39	0,38	0,38
	3	0,39	0,40	0,41
	4	0,39	0,41	0,41
<i>Wine</i>	2	0,85	0,93	0,93
	3	0,85	0,95	0,95
	4	0,89	0,94	0,95

Tabel 6. Rerata *Running Time*

<i>Dataset</i>	Parameter (<i>p</i>)	<i>Running Time</i> (s)		
		ABC	ABC-KHM	IABC-KHM
<i>Iris</i>	2	36,34	4,35	5,22
	3	37,57	4,77	5,18
	4	36,36	6,07	5,65
<i>CMC</i>	2	129,99	28,45	40,63
	3	155,31	29,36	42,99
	4	156,02	31,39	42,88
<i>Wine</i>	2	58,65	8,77	7,12
	3	59,57	8,97	7,28
	4	59,06	8,96	7,15

Tabel 7. Rerata Kesalahan Kelas

<i>Dataset</i>	Jumlah Kesalahan Kelas		
	ABC	ABC-KHM	IABC-KHM
<i>Iris</i>	3	3	3
<i>CMC</i>	190	184	185
<i>Wine</i>	5	2	2

Kinerja fungsi tujuan pada *dataset Wine* metode ABC-KHM dan metode IABC-KHM sama-sama memiliki kinerja optimalisasi terbaik (Tabel 4). Sehingga kedua metode ini dapat dijadikan alternatif untuk digunakan klasterisasi untuk data yang memiliki karakteristik seperti data *Wine*. Secara umum baik untuk setiap *dataset* dan masing-masing parameter yang digunakan, evaluasi kinerja dengan menggunakan fungsi *F-measure* menunjukkan hasil metode IABC-KHM lebih unggul dibandingkan metode ABC-KHM dan ABC (Tabel 5). Pada data *Iris* dengan menggunakan nilai parameter dua, tiga dan empat diperoleh nilai fungsi *F-measure* sebesar 0,9. Pada data *CMC* nilai fungsi *F-measure* dengan nilai parameter dua sebesar 0,38. Nilai parameter tiga dan empat menghasilkan nilai sebesar 0,41. Sehingga penggunaan metode IABC-KHM untuk proses klasterisasi lainnya yang memiliki karakteristik data seperti *CMC* disarankan menggunakan nilai parameter tiga dan empat. Pada data *Wine* nilai fungsi *F-measure* dengan nilai parameter dua sebesar 0,93. Nilai parameter tiga dan empat menghasilkan nilai terbaik sebesar 0,95. Sehingga penggunaan metode IABC-KHM untuk proses klasterisasi lainnya yang memiliki karakteristik data seperti *Wine* disarankan menggunakan nilai parameter tiga dan empat.

Kinerja untuk proses *running time* dari hipotesis yang dilakukan, seharusnya metode IABC-KHM lebih lambat dibandingkan metode ABC-KHM, hipotesis ini didasarkan pada penambahan algoritma metode IABC-KHM untuk proses pengolahan data klasterisasi. Namun hasil dari evaluasi kinerja berdasarkan proses *running time* pada Tabel 6 yang diperoleh tidak semua *dataset* dan nilai parameter yang digunakan menjadi lambat, seperti *dataset Iris* dengan nilai parameter empat, metode IABC-KHM lebih unggul dibandingkan dengan metode ABC dan metode ABC-KHM. Hasil dari *dataset Wine* dengan nilai parameter dua, tiga dan empat, metode IABC-KHM lebih unggul dibandingkan dengan kedua metode pembanding. Secara umum untuk proses klasterisasi yang memiliki data seperti data *Iris* dan data *Wine* disarankan untuk menggunakan metode IABC-KHM, sedangkan proses klasterisasi data yang memiliki karakteristik seperti data *CMC* disarankan untuk menggunakan metode ABC-KHM.

Hasil kesalahan kelas dari data masing-masing *dataset* ditunjukkan pada Tabel 7. Pada *dataset Iris* ketiga metode memiliki kesalahan kelas sebanyak tiga buah data. Pada *dataset CMC* metode ABC-KHM memiliki kesalahan kelas yang paling sedikit dibandingkan kedua metode. Meskipun IABC-KHM

memiliki kesalahan yang lebih banyak dibandingkan dengan metode ABC-KHM, IABC-KHM memiliki kesalahan yang lebih sedikit dibandingkan dengan ABC yang memiliki kesalahan kelas sebanyak 190 buah data. Selisih kesalahan kelas antara metode IABC-KHM dan metode ABC-KHM sebesar 1 buah data yaitu 185 buah data dan 184 buah data. Pada *dataset* Wine metode ABC-KHM dan metode IABC-KHM memiliki kesalahan kelas yang sama besar yaitu sebanyak 2 buah data. Kedua metode ini lebih unggul dibandingkan dengan metode ABC.

6. Kesimpulan dan Saran

Kesimpulan yang diperoleh berdasarkan hasil dan analisis dari serangkaian ujicoba yaitu posisi titik pusat kluster telah mampu menuju solusi global optimum dengan menggunakan metode *hybrid* IABC-KHM, hal ini diperkuat oleh nilai hasil fungsi tujuan yang relatif lebih kecil pada *dataset* dengan jumlah keanggotaan data yang sedikit. Hasil pengukuran berdasarkan klas label dengan menggunakan *F-Measure*, Metode IABC-KHM lebih unggul daripada metode ABC-KHM pada data *CMC* dengan nilai parameter 3 dan *Wine* dengan nilai parameter 4, sedangkan pada data *Iris* memiliki nilai yang sama antara metode IABC-KHM dan ABC-KHM. Nilai *F-measure* antara metode IABC-KHM dan metode ABC menunjukkan metode IABC-KHM lebih unggul. Waktu pemrosesan yang dibutuhkan klasterisasi data metode IABC-KHM lebih cepat pada *dataset* *Wine* dari metode pembanding yang lain, sedangkan pada *dataset* *CMC* metode ABC-KHM lebih cepat dibandingkan dengan metode lainnya.

Saran untuk penelitian selanjutnya yaitu perlunya pengembangan metode IABC-KHM untuk mempercepat proses klasterisasi karena hasil *running time* menunjukkan hanya unggul pada *dataset* tertentu.

Daftar Pustaka

- Gao, Weifeng., & Liu., Sanyang. Improved artificial bee colony algorithm for global optimization. *Information Processing Letters*. Hal 871-882. 2011
- Gungor, Z., & Unler, A. K-Harmonic Means Data Clustering with Simulated Annealing Heuristic. *Applied Mathematics and Computation*. 184. 199–209. 2007.
- Hammerly, G., & Elkan, C. Alternatives to The K-Means Algorithm that Find Better Clusterings. *Proceedings of the 11th international conference on information and knowledge management*. 600–607. 2002
- Handl, J., Knowles, J., dan Drigo, M. On the Performance of Ant-based Clustering. *Design and Application of Hybrid Intelligent System*. 104. 204-213. 2003
- Jiang, H., Yi, S., Li, J., Yang, F., dan Hu, X. Ant Clustering Algorithm with K-Harmonic Means Clustering. *Expert Systems with Applications*. 37. 8679–8684. 2010
- Karaboga, Dervis., Ozturk, Celal. A novel clustering approach: Artificial Bee Colony (ABC) algorithm. *Applied Soft Computing*. 11. 652-657. 2011.
- Pen, J.M., Lozano, J.A., & Larranga, P. An Empirical Comparison of Four Initialization Methods for The K-Means Algorithm. *Pattern Recognition Letters*. Vol. 20. Hal. 1027-1040. 1999
- Tan, P.N., Steinbach, M., dan Kumar, V. Introduction to Data Mining 4th edition. Pearson Addison Wesley. New York. 2006.
- Yang, F., Sun. T., dan Zhang, C. An Efficient Hybrid Data Clustering Method Based on K-Harmonic Means and Particle Swarm Optimization. *Expert Systems with Application*. 36. 9847-9852. 2009
- Zhang, B., Hsu, M., & Dayal, U. K-Harmonic Means – A Data Clustering Algorithm. Technical Report HPL-1999-124. Hewlett-Packard Laboratories. 1999
- Zhang C., Ouyang, D., dan Ning, J. An Artificial Bee Colony Approach for Clustering. *Expert Systems with Applications*. 37. 4761-4767. 2009.

Halaman ini sengaja dikosongkan