

Implementasi Bee Colony Optimization Pada Pemilihan Centroid (Klaster Pusat) Dalam Algoritma K-Means

Ika Arfiani^{1,*}, Herman Yuliansyah¹, M Dzikrullah Suratin²

¹ Fakultas Teknologi Industri, Informatika, Universitas Ahmad Dahlan, Yogyakarta, Indonesia

² Fakultas Teknik, Informatika, Universitas Muhammadiyah Maluku Utara, Ternate, Indonesia

Email: ^{1,*}ika.arfiani@tif.uad.ac.id, ²herman.yuliansyah@tif.uad.ac.id, ³irul.dzi@gmail.com

Email Penulis Korespondensi: ika.arfiani@tif.uad.ac.id

Submitted: 23/03/2022; Accepted: 31/03/2022; Published: 31/03/2022

Abstrak—Klustering merupakan metode yang digunakan untuk membagi data menjadi beberapa kelompok bagian. K-means (KM) merupakan algoritma yang sering digunakan dalam klustering, hanya saja hasil dari KM sering kali terjebak di lokal optima yaitu solusi yang optimal (baik maksimal ataupun minimal) pada kandidat solusi di tetangga terdekatnya saja bukan keseluruhan dari semua solusi yang ada atau yang biasa disebut global optima. Pada penelitian ini bertujuan untuk melakukan memperbaiki proses penentuan klaster pada algoritma Kmeans dengan menggunakan algoritma Bee Colony Optimization (BCO). BCO merupakan algoritma yang bekerja berdasarkan cara lebah mencari makan, BCO terkenal mampu lolos dari jebakan lokal optima dengan mengenali mana hasil yang terbaik dari serangkaian hasil optimal. Menggabungkan BCO dengan KM dimulai dengan memilih sumber makanan awal secara acak dan menggunakan KM untuk menyelesaikan semua permasalahan klustering pada setiap langkah BCO berikutnya serta menyimpan sumber makanan terbaik disetiap iterasinya. Sumber-sumber terbaik tersebut akan dipilih sumber makanan terbaiknya berdasarkan probabilitas kecocokannya masing-masing. Hasil dari penelitian ini bahwa metode BCOKM telah terbukti dapat menyelesaikan permasalahan pembagian data, dimana metode BCOKM mampu membentuk cluster yang baik, ditunjukkan dari nilai fitness yang dihasilkan (nilai terendah 1221.53 dan nilai tertinggi 1233.28) semuanya lebih baik dari nilai fitness menggunakan K-means (1251.42). Begitu pula dalam hal akurasi, dimana penggunaan BCOKM seluruhnya menunjukkan hasil (83.16%-83.30%) lebih baik daripada penggunaan hanya K-means (83.09%).

Kata Kunci: Bee Colony Optimization; K-Means; Klustering

Abstract—Clustering is a method that is used to divide the data into several groups of parts. K-means (KM) is an algorithm that is often used in clustering, only just the result of KM often times get stuck in local optima i.e. the optimal solution (both maximum or minimal) on the candidate solution in the nearest neighbor only, not the whole of all existing solutions or what is commonly called the global optima. In this study aims to do improve the cluster determination process on the Kmeans algorithm using the Bee Colony Optimization (BCO) algorithm. BCO is an algorithm that works based on the way the bees search for food, BCO is famous for being able to escape from the local optima trap by recognizing which results are best from a series of optimal results. Combining BCO with KM begins with selecting a source of food early in random and using KM to resolve all the problems of clustering at every step BCO next and keep sources of food best in each iteration. The result of this research is that the BCOKM method has been proven to be able to solve the problem of data sharing, where the BCOKM method is able to form a good cluster, as shown by the resulting fitness value (the lowest value is 1221.53 and the highest value is 1233.28) all of which are better than the fitness value using K-means (1251.42). Likewise in terms of accuracy, where the use of BCOKM all showed better results (83.16%-83.30%) than the use of only K-means (83.09%).

Keywords: Bee Colony Optimization; K-Means; Clustering

1. PENDAHULUAN

Klustering digunakan untuk membagi data ke dalam kelompok-kelompok yang homogen. Dalam pekerjaan pengelompokan, label dari data belum diketahui dan dengan pengelompokan diharapkan dapat diketahui kelompok data untuk kemudian diberi label sesuai keinginan.[1] Klustering dalam penelitian Pham dkk digunakan untuk membagi data menjadi kelompok yang homogen, banyak aplikasi dari klustering yang telah dirasakan manfaatnya seperti segmentasi citra, mengenali segmentasi pasar dalam bisnis, temu kembali informasi dan juga dalam merangkum data. Pham menyajikan hasil pengujian metode pengelompokan berbasis Algoritma Bees terhadap algoritma k-means dan GA-clustering. Algoritma itu diterapkan pada lima set data nyata (Vokal, Iris, Minyak Mentah, Peta Kontrol dan Cacat Kayu). Metode ini menggunakan Algoritma Lebah untuk mencari set pusat klaster yang meminimalkan metrik pengelompokan tertentu[2].

K-means (KM) dalam penelitian yang dilakukan oleh Nainggolan merupakan algoritma clustering yang umum digunakan karena kemudahan dan waktu yang dibutuhkan untuk melakukan pembelajaran relatif cepat.[3] Permasalahannya, hasil cluster KM sering mengacu pada solusi local optima. Yaitu solusi optimal (baik maksimum maupun minimum) terhadap kandidat solusi pada tetangga terdekatnya, bukan seluruh solusi yang ada atau biasa disebut global optima.[4] Menurut penelitian yang pernah dilakukan oleh Ario, dkk bahwa metode ABCKM telah terbukti dapat menyelesaikan permasalahan pembagian data, dimana nilai silhouette yang diperoleh untuk metode ini selalu lebih tinggi bila dibandingkan dengan metode KM.[5] Permasalahan lain pada K-means adalah bahwa metode tersebut sangat bergantung pada penentuan titik awal centroid.[6]

Oleh karena itu diusulkan untuk mengimplementasikan algoritma *Bee Colony Optimization* (BCO) yang umum digunakan untuk pencarian global seperti masalah Traveling Salesman Problem (TSP) yang mencari rute termurah dan paling efisien untuk mencapai tujuan banyak tenaga penjualan. Cara kerja BCO yang menirukan lebah

mencari makan (*foregaging*) dari nektar ini, dimulai dengan terbangnya lebah pekerja mencari sumber makanan yang kemudian memberikan lokasinya kepada lebah *onlooker* dengan menari.[5]

Onlooker bee akan membandingkan nilai probabilitas setiap lokasi sumber makanannya dan mencari sumber makanan disekitar sumber yang dipilih, sehingga dengan segera sumber makanan yang ditemukan sekarang lebih banyak dari sumber sebelumnya, lebah akan melupakan informasi sebelumnya. sumber makanan dan untuk sumber makanan habis atau tidak berubah maka sumber makanan akan ditinggalkan dan dilupakan dan lebah akan menjadi lebah pramuka yang akan mengintai sumber makanan baru yang akan dibuat diruang pencarian. Langkah-langkah tersebut akan dilakukan hingga ditemukan solusi global yang optimal [2].

BCO merupakan bagian dari swarm intelligence, dimana Swarm intelligence merupakan sebuah metode penyelesaian masalah yang memanfaatkan perilaku sekumpulan agen yang saling bekerja sama. Setiap upaya untuk merancang algoritma atau didistribusikan pemecahan perangkat terinspirasi oleh perilaku kolektif dari serangga sosial koloni dan masyarakat hewan lainnya.[7]

Aplikasi lain dari BCO adalah penjadwalan produksi barang, yang akan digunakan untuk menjaga siklus produksi, yang merupakan rangkaian kegiatan bisnis seperti pencatatan pesanan dari pelanggan, pencatatan bahan baku, juga pencatatan gaji karyawan, dan kegiatan pengolahan data secara terus menerus. Peran BCO dalam mengolah data siklus bisnis akan sangat signifikan dimana akan membutuhkan perencanaan jumlah produksi dan persediaan, penjadwalan proses produksi dengan meminimalkan jumlah waktu proses yang dibutuhkan untuk menyelesaikan seluruh proses produksi, dan juga mengoptimalkan biaya produksi.

Metode pengelompokan yang digunakan dalam penelitian ini menggabungkan kedua algoritma tersebut menjadi Bee Colony Optimization K-means (BCO KM). Metode ini akan memperbaiki atau melengkapi cara kerja algoritma KM dalam pemilihan pusat kluster dan dengan menyisipkan sifat pencarian global BCO maka akan dipilih kumpulan sumber makanan yang disimpan, sehingga didapatkan sumber makanan yang optimal untuk membagi data. Sehingga dari solusi ini akan mampu memberikan solusi yang optimal atas kelemahan yang dimiliki oleh KM. mengingat teknik clustering dengan K-means cukup sederhana [8].

2. METODOLOGI PENELITIAN

Langkah pada penelitian ini menggabungkan langkah yang ada dalam algoritma K-Means dengan menyisipkan Bee Colony Optimization pada pemilihan kluster pusatnya.

2.1 Bee Colony Optimization

Bee Colony Optimization (BCO) adalah salah satu algoritma optimasi yang memulai pencarian makanan. Lebah merupakan serangga sosial yang sangat terorganisir. Koloni lebah buatan bersama-sama mencari solusi optimal dari masalah yang diberikan. Setiap lebah buatan menghasilkan satu solusi untuk masalah ini. Ada dua fase dalam satu langkah algoritma BCO yaitu fase maju (*forward pass*) dan fase mundur (*backward pass*)[9]

Setelah suatu area ditemukan, lebah pekerja akan menari (memberi tanda) agar nektar dapat diambil oleh lebah *onlooker*. Kemudian Bee *Onlooker* akan menentukan sumber makanan mana yang baik untuk dipanen, dan membiarkan sumber tersebut habis dan berubah menjadi pramuka lebah. Lebah pramuka bertugas mencari sumber baru yang dihasilkan secara acak di ruang pencarian, dimana informasi posisi nektar yang lebih sedikit akan dilupakan oleh lebah hingga sumber makanan terbaik ditemukan. [11].

Berikut langkah-langkah yang digunakan dalam BCO [11]:

1. Tentukan jumlah sumber makanan awal atau solusi potensial (SN) yang akan disebut X ij .
2. Mengevaluasi apakah f i (kualitas sumber makanan) dalam populasi karena semakin banyak nektar semakin besar probabilitas lebah pencari sumber yang dipilih.
3. Masuk ke fase *Employed Bee*, yaitu menghitung V ij yang merupakan tetangga di sekitar X ij.
4. Masuk ke fase *Onlooker Bee* yang akan melakukan langkah kedua dengan tujuan menggabungkannya menjadi sumber dengan probabilitas terbaik untuk masing-masing clusternya yang akan disebut X ij baru. Sumber akan diulang menggunakan langkah *Bee*, tetapi sumber terbaik akan disimpan dalam memori sebagai sumber terbaik dalam iterasi ini.
5. Memasuki fase *Lebah Pramuka*, fase ini menghasilkan sumber-sumber baru dengan syarat pengumpulan sumber makanan belum mencapai batas Jumlah Siklus Maksimum. Fase ini akan menjembatani iterasi pertama ke iterasi kedua dan berikutnya.
6. Pengumpulan sumber akan dibandingkan dengan probabilitas kompatibilitas sampai ditemukan sumber yang optimal.

$$x_{i,j} = l_{jk} + \text{rand}(0,1)(u_{jk} - l_{jk}) \quad (1)$$

Di mana:

SN = Jumlah sumber makanan

D = Jumlah dimensi data

K = Jumlah cluster

k = { 1,2 ..., K }

i = { 1,2, ..., SN }

$$j = \{1, 2, \dots, D\}$$

x_i, j = i-sumber makanan awal dalam dimensi ke- j

l_{jk} = nilai yang lebih rendah dari setiap nilai dalam dimensi- j untuk setiap cluster

u_{jk} = nilai teratas setiap dimensi- j untuk setiap cluster

$rand(0,1)$ = menghasilkan bilangan acak dengan distribusi biasa 0 sampai 1.

Hitung nilai fitness/kebugaran dengan :

$$fit_i = \frac{1}{1+f_i} \quad (2)$$

$$f_i = \frac{fit_i}{\sum_{n=1}^{SN} fit_n} \quad (3)$$

Memulai siklus iterasi sama dengan 1 hingga MCN. Untuk setiap solusi penghitungan lebah yang digunakan cara baru:

$$v_i = Z_{ij} + \phi_{ij}(Z_{ij} - Z_{kj}) \quad (4)$$

Di mana:

v_{ij} = Kandidat posisi makanan baru berdasarkan posisi lama di memori. $\{1, 2, \dots\} \{1, 2, \dots\}$ adalah indeks yang dipilih secara acak, tetapi k harus merupakan indeks yang berbeda dari i .

ϕ_{ij} = Angka acak antara $\{-1, 1\}$

Indeks acak dalam persamaan bertujuan untuk mengontrol produksi sumber makanan tetangga di sekitar, dan mewakili perbandingan dua posisi sumber makanan yang terlihat pada lebah. Langkah selanjutnya adalah menghitung nilai, dan melakukan proses seleksi sebagai sumber makanan yang rakus di setiap klaster-nya.

Untuk setiap lebah pengamat akan menentukan solusi z_i berdasarkan p_i . Kemudian hitung nilai v_i sesuai dengan langkah Employed Bee. Bandingkan dua sumber makanan yang rakus. Berdasarkan hasil probabilitas, solusi terbaik pada iterasi ini akan disimpan.

Berdasarkan hasil probabilitas, solusi terburuk yang tersisa adalah lebah onlooker dan diganti dengan solusi baru yang diperoleh secara acak dari :

$$Z_i^j = Z_{min}^j + rand(0,1)(Z_{max}^j - Z_{min}^j) \quad (5)$$

Dimana :

z_i = Sumber makanan meninggalkan lebah pengamat dan $\{1, 2, \dots\}$.

Ulangi siklus tersebut hingga Maximum Cycle Number (MCN). Jumlah Siklus Maksimum atau limit untuk BCO KM dihitung dengan persamaan berikut:

$$limit = SN * D \quad (6)$$

Sumber pangan yang optimal diperoleh dari perbandingan nilai probabilitas kesesuaian sumber pangan tersebut.

2.2 K-Means

K-means (KM) adalah algoritma data cluster yang mengelompokkan data berdasarkan data yang tidak berlabel kelas. Berikut langkah-langkah yang dilakukan dalam KM [4]:

1. Tentukan banyaknya k (cluster).
2. Hasilkan centroid awal (titik pusat cluster) secara acak.
3. Hitung jarak setiap data ke setiap centroid dengan menggunakan rumus korelasi antara dua objek yaitu Euclidean Distance di setiap cluster.
4. Kelompokkan setiap data menurut jarak terdekat antara data dan centroid.
5. Perbarui nilai centroid dengan nilai centroid
6. Ulangi langkah 3-5 sampai nilai centroid tidak berubah.

Jarak antara input dan centroid dihitung dengan:

$$d(X_i, Y_j) = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2} \quad (7)$$

Di mana:

X_i = vektor nilai masukan $\{x_1, x_2, \dots, x_n\}$

Y_j = Nilai pusat vektor $\{y_1, y_2, \dots, y_n\}$

Memperbarui nilai centroid dengan nilai centroid baru dilakukan dengan:

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \quad (8)$$

Di mana:

μ_k = Centroid baru dari k -cluster.

N_k = Jumlah data dalam k -cluster.

x_i = Vektor nilai input ke- i pada cluster ke- k .

Memilih titik pusat suatu cluster merupakan tugas yang rumit karena diatur atau diacak, jika nilai inisialisasi yang dilakukan buruk maka pengelompokan data dapat menjadi kurang optimal [3]. Pada BCOKM, hasil perhitungan

jarak Euclidean akan menggantikan fungsi biaya f_i pada BCO, dan pemilihan centroid dilakukan dengan menggunakan persamaan (1), dimana sumber makanan terdiri dari kombinasi centroid pada setiap cluster.

2.3 Optimasi BCOKM

Algoritma *Bee Colony Optimization* akan digunakan untuk menanggulangi kelemahan pada penentuan pusat kluster dalam K-means guna penentuan sumber makanan terbaik. Kelemahan tersebut adalah tingkat akurasi hasil kluster sangat tergantung dari penentuan titik awal pusat kluster sehingga sensitif terhadap penentuan titik awal dan memungkinkan hasil kluster konvergen pada lokal optimal [12]. Optimasi merupakan suatu proses yang memiliki tujuan untuk mencapai hasil yang optimal (nilai efektif yang dapat dicapai). Penggunaan KM seringkali menghasilkan solusi yang optimal. Kandidat solusi pada tetangga terdekat bukanlah keseluruhan dari semua solusi yang ada atau biasa disebut global optima. Jadi menggunakan BCO akan membantu untuk menghindari terjebak dalam local optima dalam KM.

Algoritma BCOKM akan menghasilkan sumber makanan terbaik yang telah disimpan dari fase onlooker pada setiap iterasi, kemudian sumber optimal akan dipilih dengan cara yang sama, berdasarkan nilai probabilitas fitness terbesar. Sehingga hasil terbaik akan menjadi KM terakhir kali untuk mendapatkan hasil data sharing/clustering. sumber makanan ini. Berikut ini akan dijelaskan metode BCOKM menggunakan flow chart yang akan menggambarkan langkah-langkah dari Bee Colony Optimization K-means secara umum, yang akan dijelaskan pada Gambar 1. Seperti yang telah diketahui, BCO KM memiliki 4 fase utama selain fase input dan output yaitu tahap inialisasi sumber makanan, *employed bee*, *onlooker bee* dan *scout bee*. Ketiga fase proses *employed*, *onlooker*, dan *scout bee* akan terus diulang hingga kriteria terminasi perulangan dipenuhi dan sumber makanan terbaik yang diperoleh akan dipilih sebagai solusi terakhir [12]. Sumber makanan terbaik yang didapatkan setelah itu akan dipilih menjadi solusi terakhir. K-means dilakukan kembali menggunakan sumber tersebut untuk membentuk kelompok cluster sebagai hasil uji coba untuk perhitungan akurasi [13].

Berdasarkan permasalahan tersebut, diperlukan metode yang dapat menghitung rute paling optimal dalam melakukan distribusi barang. Dalam penelitian ini akan digunakan metode Artificial Bee Colony (ABC) untuk menyelesaikan masalah di Nida Food. Algoritma ini terinspirasi dari perilaku sekumpulan lebah berkelompok untuk mencari sumber makanan. Setelah mereka menemukan sumber makanan kemudian mereka akan kembali ke sarang dan melakukan tarian lebah (*waggle dance*), dengan menggunakan *waggle dance* semua koloni saling berkomunikasi tentang sumber makanan yang mereka temukan, sehingga lebah-lebah yang lain akan mengetahui letak dari sumber makanan yang paling dekat dari sarang [14].

2.4 Pengujian Akurasi

Pengukuran akurasi digunakan untuk mengetahui seberapa baik hasil dari penelitian yang dilakukan. [15]. Akurasi didefinisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual. Pada information extraction terdapat beberapa istilah yaitu *precision*, *recall* dan *f-measure*. *Precision* adalah tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh sistem. *Recall* adalah tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi. *F-Measure* merupakan salah satu perhitungan evaluasi dalam informasi temu kembali yang mengkombinasikan *recall* dan *precision*. Nilai *recall* dan *precision* pada suatu keadaan dapat memiliki bobot yang berbeda. Ukuran yang menampilkan timbal balik antara *Recall* dan *Precision* adalah *F-Measure* yang merupakan bobot *harmonic mean*, *recall* dan *precision* [15].

Pada tahap ini, kinerja program diuji untuk masalah berbagi data. Uji performansi ini dilakukan dengan menentukan nilai akurasi dan fitness dari pengelompokan sumber makanan yang dihasilkan setelah menggunakan algoritma BCOKM dengan menggunakan jumlah sumber makanan dan variasi parameter iterasi ABC meningkat. Jumlah variabel dimensi dan cluster ditentukan dengan cara yang sama di semua percobaan. Artinya, jumlah dimensi yang dipakai adalah 5, dan jumlah clusternya adalah 10.

2.5 Tahapan Penelitian

Tahapan dari penelitian yang dilakukan untuk mengelompokkan data siswa dapat dilihat pada Gambar 1, dimana dalam tahapan ini sudah menerapkan gabungan antara metode K-Means yang didalamnya dimodifikasi dengan Bee Colony Optimization. Gambar 1 menunjukkan beberapa langkah untuk menganalisis data cluster mahasiswa, yaitu :

- a. Input data set mahasiswa dan inialisasi parameter awal
 - Inialisasi parameter awal dibagi menjadi parameter data mahasiswa dan parameter algoritma BCOKM.
 - 1. Data mahasiswa dibagi menjadi delapan variabel, yaitu Nomor Mahasiswa (SN), Nama, Jurusan (Dept), Lama Pengerjaan Skripsi (LT), Tanggal Registrasi (RD), Tanggal Wisuda (GD), Indeks Prestasi Kumulatif (GPA), dan Skor Kecakapan Bahasa Inggris (EP)
 - 2. Parameter BCOKM meliputi jumlah kelompok, jumlah iterasi/MCN.
- b. Proses tahapan BCOKM
 1. Fase Inialisasi K pusat secara random

Pada tahap ini dilakukan inisialisasi sumber makanan awal. Sumber makanan diinisialisasi secara acak dan masing-masing sumber makanan merepresentasikan satu *centroid*. Setiap sumber makanan nantinya akan diberi satu *Employee bee*. Menggunakan algoritma *K-Means* untuk *clustering* dan kemudian menghitung nilai rerata *fitness*

2. Fase Employee bee

Setiap *Employee bee* akan mencari sumber makanan baru untuk mendapatkan sumber makanan dengan jumlah nectar yang lebih banyak disekitar sumber makanan sekarang. Menggunakan algoritma *K-Means* untuk *clustering* dan menghitung nilai *fitness*. Kemudian melakukan seleksi *greedy* untuk evaluasi nilai fitness sumber makanan baru dan membandingkannya dengan sebelumnya. Sumber makanan yang lebih baik akan diterima oleh *onlooker bee*.

3. Fase Onlooker bee

Onlooker bee menunggu informasi sumber makanan dari Employee bee. Onlooker bee akan mengerjakan langkah kedua dengan tujuan menggabungkannya menjadi sumber makanan dengan nilai probabilitas tertinggi untuk setiap clusternya. Kemudian dilakukan perhitungan nilai probabilitas kumulatifnya menggunakan seleksi roulette wheel. Sumber makanan kemudian akan diulang langkah Employee bee, tetapi sumber makanan terbaik akan disimpan di memori sebagai sumber makanan terbaik di iterasi tersebut.

4. Fase Scout bee

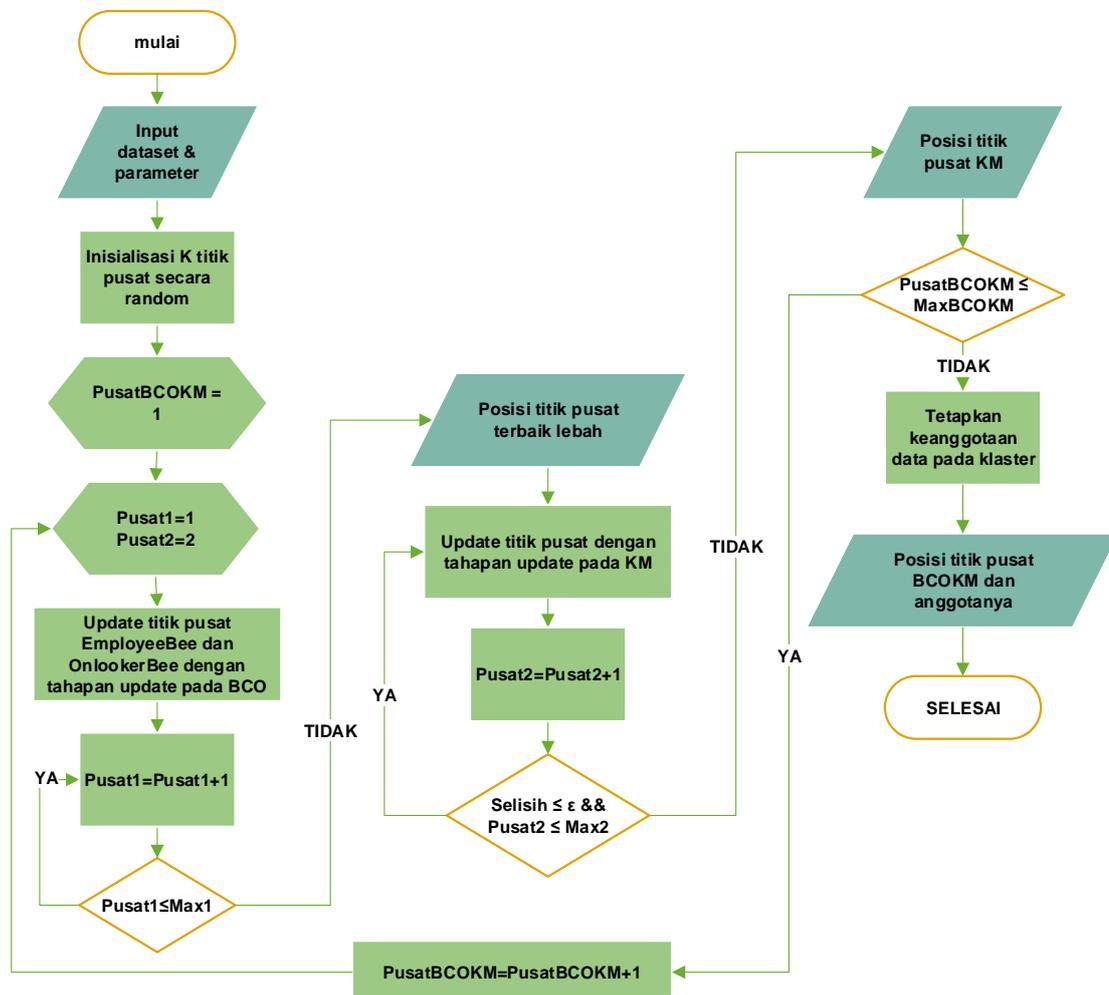
Pada fase ini, membangkitkan sumber makanan baru dengan kondisi kumpulan sumber makanan belum mencapai batas limit parameter atau MCN. Fase ini akan menjembatani iterasi pertama ke iterasi kedua dan seterusnya.

5. Pemilihan Sumber Makanan Terbaik

Ketiga fase (Employee bee, Onlooker bee dan Scout bee) akan terus menerus diulangi sampai ke batas iterasi. Sumber makanan terbaik dari semua iterasi akan dipilih sebagai solusi akhir kemudian dihitung nilai fitnessnya dan nilai fitness terbaik akan menjadi solusi sumber makanan terbaik oleh bee.

c. Hasil Pengelompokan

Hasil pengelompokan data mahasiswa ini berupa jumlah anggota pada setiap cluster dan nilai fitness beserta nilai silhouette untuk mengetahui kualitas pengelompokan.



Gambar 1. Flowchart Tahapan penelitian penerapan BCOKM

3. HASIL DAN PEMBAHASAN

Dataset dari sebuah lembaga pendidikan swasta berisi 724 baris data dengan delapan variabel, yaitu Nomor Mahasiswa (SN), Nama, Jurusan (Dept), Lama Pengerjaan Skripsi (LT), Tanggal Registrasi (RD), Tanggal Wisuda (GD), Indeks Prestasi Kumulatif (GPA), dan Skor Kecakapan Bahasa Inggris (EP). Tabel 1 menunjukkan kumpulan dataset.

Tabel 1. Dataset penelitian berupa data mahasiswa

Data Row	Sn	Dept	Lt (Month)	Rd	Gd	Gpa	Ep
1	08018001	Inf	13.6	2008-09-08	2014-02-17	2.92	406
2	08018003	Inf	6.8	2008-09-08	2012-12-22	3.53	400
3	08018004	Inf	4.83	2008-09-08	2013-01-05	3.44	403
...
723	1400018238	Inf	2.8	2014-09-01	2018-06-28	3.68	430
724	1400018243	Inf	2.8	2014-09-01	2018-06-28	3.63	413

Nilai fitness suatu sumber makanan dari lebah merupakan nilai total dari jarak sebuah node (data) menuju cluster yang terdekat darinya, yang dihitung dengan Sum of Squared Error (SSE). Semakin kecil nilai fitness yang dihasilkan, maka semakin baik cluster yang terbentuk. Hasil percobaan pada Tabel 2 menunjukkan bahwa secara keseluruhan nilai fitness yang dihasilkan hanya dengan menggunakan K-means yaitu 1251,42 lebih tinggi dibandingkan dengan menggunakan BCOKM, yaitu antara nilai terendah yang diperoleh dari pengujian dengan 100 SN dan 100 iterasi (1221,53) dan nilai tertinggi diperoleh dari pengujian menggunakan 50 SN dan 50 iterasi (1233,28), hal ini menunjukkan bahwa cluster yang terbentuk pada BCO lebih baik.

Tabel 2. Nilai Fitness

FITNESS	JUMLAH SUMBER MAKANAN						
	Kmeans	20	30	50	75	100	AVG
Kmeans	1251,42						
ITERASI	25	1224,57	1226,31	1226,25	1223,22	1221,96	1224,46
	50	1231,54	1227,52	1233,28	1225,37	1221,77	1227,9
	75	1225,46	1224,85	1225,73	1222,98	1223,01	1224,41
	100	1230,05	1225,41	1223,92	1223,47	1221,53	1224,88
AVG		1227,905	1227,4625	1225,855	1223,76	1222,07	

Hasil dari percobaan dengan 764 data pada 100 iterasi menunjukkan bahwa secara keseluruhan, nilai fitness yang dihasilkan :

- Jika hanya pakai K-means saja, yaitu **1251.42**, lebih tinggi daripada menggunakan BCOKM,
- BCOKM yaitu nilai terendah **1221.53** dan nilai tertinggi **1233.28** menunjukkan bahwa cluster yang terbentuk pada BCO lebih baik.

Hasil dari pengujian tingkat ketepatan antara informasi yang diminta oleh pengguna dengan jawaban yang diberikan oleh system atau tingkat presisi datanya disajikan dalam Tabel 4 bahwa Nilai precision dari K-means mampu mengungguli BCOKM yakni 11.09%, hanya saja pada pada iterasi sebanyak 25 kali dengan SN sebesar 30 dan 50 yang keduanya bernilai 11.48%, dan iterasi 75 dengan SN 30 dan 50 yang masing-masing bernilai 11.53% dan 11.03%.

Tabel 4. Nilai Precision

PRECISION	JUMLAH SUMBER MAKANAN						
	Kmeans	20	30	50	75	100	AVG
Kmeans	11.09%						
ITERASI	25	10.70%	11.48%	11.48%	11.05%	11.17%	11,18%
	50	11.16%	11.03%	11.05%	11.12%	11.07%	11,09%
	75	11.16%	11.53%	11.03%	10.97%	11.06%	11,15%
	100	11.07%	11.25%	11.04%	11.14%	11.12%	11,12%
AVG		11,02%	11,32%	11,15%	11,07%	11,11%	

Sedangkan untuk mengetahui tingkat keberhasilan sistem dalam menemukan kembali sebuah informasi atau Recall disajikan dalam Tabel 5 bahwa Nilai recall terbaik muncul ketika digunakan 100 SN dengan 25 kali iterasi (12.44%). Disini nilai recall algoritma K-means saja (11.98%) masih kalah dibandingkan sebagian besar penggunaan BCOKM. Meskipun demikian masih terdapat nilai recall lebih rendah dari K-means, yaitu pada penggunaan 20 SN dengan 25 kali iterasi (11.78%) dan 30 SN dengan 50 kali iterasi (11.62%).

Tabel 5. Nilai Recall

RECALL	JUMLAH SUMBER MAKANAN						AVG
	Kmeans	20	30	50	75	100	
Kmeans	11.98%						
ITERASI	25	11.78%	12.40%	12.47%	12.28%	12.44%	12.27%
	50	12.22%	11.62%	12.31%	12.39%	12.31%	12.17%
	75	12.33%	12.46%	12.36%	12.22%	12.16%	12.31%
	100	12.36%	12.36%	12.27%	12.29%	12.30%	12.32%
	AVG	12.17%	12.21%	12.35%	12.29%	12.30%	

Selanjutnya dilakukan perhitungan akurasi untuk mengetahui seberapa akurat pengelompokan yang dihasilkan melalui penggunaan BCOKM. Akurasi pengelompokan data akan dihitung berdasarkan kelas benchmark yang diperoleh dari dataset. Pada Tabel 3 diperoleh akurasi yang dihasilkan lebih baik bila jumlah iterasi yang digunakan adalah 25 (83,25%). Rata-rata akurasi terburuk justru didapat saat uji coba dilakukan menggunakan 100 SN dengan 100 kali iterasi (83,16%). Sedangkan akurasi yang dihasilkan hanya dengan menggunakan K-means (83,09%) masih lebih rendah dibandingkan dengan penggunaan metode BCOKM tanpa memperhatikan jumlah SN dan iterasi.

Tabel 3. Nilai Akurasi

RAND MEASURE	JUMLAH SUMBER MAKANAN						AVG
	Kmeans	20	30	50	75	100	
Kmeans	83.09%						
ITERASI	25	83.30%	83.20%	83.27%	83.24%	83.24%	83,25%
	50	83.19%	83.18%	83.23%	83.21%	83.24%	83,21%
	75	83.29%	83.21%	83.19%	83.20%	83.27%	83,23%
	100	83.23%	83.24%	83.21%	83.19%	83.16%	83,21%
	AVG	83,25%	83,21%	83,23%	83,21%	83,23%	

Hasil dari percobaan dengan 764 data pada 100 iterasi menunjukkan bahwa secara keseluruhan, nilai akurasi yang dihasilkan:

- Jika hanya pakai K-means saja, yaitu **83,09%**, lebih rendah daripada menggunakan BCOKM,
- BCOKM yaitu nilai terendah **83,16%** dan nilai tertinggi **83,30%** menunjukkan bahwa cluster yang terbentuk pada BCOKM lebih baik

4. KESIMPULAN

Berdasarkan hasil pengujian dan evaluasi yang telah dilakukan bahwa metode BCOKM telah terbukti dapat menyelesaikan permasalahan pembagian data, dimana metode BCOKM mampu membentuk cluster yang baik, ditunjukkan dari nilai fitness yang dihasilkan (nilai terendah 1221.53 dan nilai tertinggi 1233.28) semuanya lebih baik dari nilai fitness menggunakan K-means (1251.42). Begitu pula dalam hal akurasi, dimana penggunaan BCOKM seluruhnya menunjukkan hasil (83.16%-83.30%) lebih baik daripada penggunaan hanya K-means (83.09%). Nilai precision dan recall yang dihasilkan rendah, berkisar antara 11-12%. Hal ini kemungkinan disebabkan karena dataset yang digunakan sebenarnya lebih cocok digunakan untuk permasalahan klasifikasi yang merupakan metode pembelajaran dimana kelas awal data telah diketahui (supervised learning) dan bukan clustering pada algoritma K-means dan BCOKM yang merupakan metode unsupervised learning. Namun, penelitian ini juga menemukan bahwa hasil pengelompokan data yang telah dilakukan memiliki titik pusat yang berdekatan sehingga tidak dapat mewakili karakteristik siswa yang ingin dicari. Kami memperkirakan bahwa jumlah data yang kami gunakan terlalu kecil dan perlu mengevaluasi distribusi data. Peningkatan ini akan berfungsi untuk keberlanjutan penelitian ini.

REFERENCES

- Prasetyo, E., 2012. Data Mining Konsep dan Aplikasi Menggunakan Matlab, Andi Offset, Yogyakarta.
- Pham, D. T., Otri, S., Afify, A., Mahmuddin, M., & Al-Jabbouli, H. 2007. Data Clustering Using the Bees Algorithm.
- Nainggolan, Rena., 2014. Algoritma Modified K-Means Clustering Pada Penentuan Cluster Centre Berbasis Sum Of Squared Error (SSE). Thesis. Program Studi S2 Teknik Informatika Fakultas Ilmu Komputer Dan Teknologi Informasi, Universitas Sumatera Utara, Medan.
- Aggarwal, C. C. (2015). Data Mining: The Textbook. New York.
- Nugroho, B. Ario., Purwitasari, D., Fatichah, C., 2016. Implementasi Artificial Bee Colony Untuk Pemilihan Titik Pusat pada Algoritma K-means. Jurnal Teknik ITS, Vol. 5, No. 2, pp. A608-A613, ISSN:2337-3539
- Armano, G., & Farmani, M. R. 2014. Clustering Analysis with Combination of Artificial Bee Colony Algorithm and k-Means Technique. s.l. : International Journal of Computer Theory and Engineering.
- Bonabeau, E., Dorigo, M., dan Theraulaz, G., 1999, Swarm Intelligence from Natural to Artificial Systems, Oxford University Press, New York



- [8] P.-N. Tan, M. Steinbach, and V. Kumar. 2005. Introduction to Data Mining. Boston, MA, USA : Addison-Wesley Longman Publishing Co., Inc., 2005.
- [9] Pratama, J. Danar., 2019. Penerapan Algoritma Bee Colony Optimization Dalam Pencarian Rute Tercepat Rumah Sakit di Kota Bogor. Jurnal Riset Komputer (JURIKOM), Vol. 99 No. 99, ISSN 2407-389X (Media Cetak) Hal: 999-999.
- [10] Kaur, A., Goyal, S., 2011. A Survey on the Applications of Bee Colony Optimization Techniques, International Journal on Computer Science and Engineering(IJCSE), India, 8 Agustus 2011, vol. 3, 3037 –3046, ISSN : 0975 –3397.
- [11] Karaboga, D., & Ozturk, C. 2009. A Novel Clustering Approach: Bee Colony Optimization(BCO) algorithm.
- [12] R. E. Funderlic, M. T. Chu, N. Orlowski, D.Schlor, and J. Blevins, 2019. “Convergence and Other Aspects of the k-modes Algorithm for Clustering Categorical data.”
- [13] Novianto, Rizki., 2017. Optimasi K-Means Dengan Artificial Bee Colony Untuk Pengelompokan Resolusi Tahun Baru Dari Data Twitter. Undergraduate Thesis. Institute Teknologi Sepuluh November.
- [14] E. Rahmandha, R. Efendi, and D. Puspitaningrum, 2016. “Aplikasi Pencarian Lokasi Fasilitas Peayanan Umum Terdekat Menggunakan Metode Artificial Bee Colony Di Kota Bengkulu Berbasis Webview Android,” J. Teknol. Inf., vol. 12, pp. 141–154.
- [15] Nugroho, Fajar 2019 Sistem Rekomendasi Kata Kunci Untuk Website Menggunakan Hybrid Semantic Relatedness Dan Associative Neural Network. Masters thesis, Universitas Komputer Indonesia.