

SYSTEMIC: Information System and Informatics Journal

ISSN: 2460-8092, 2548-6551 (e)

Vol 4 No 2 - Desember 2018

Metode Hibridasi Artificial Bee Colony dan Fuzzy K-Modes untuk Klasterisasi Data Kategorikal**Khalid**

Universitas Islam Negeri Sunan Ampel Surabaya

Khalid@uinsby.ac.id

Kata Kunci*Data Kategorikal, klustering, Fuzzy K-Modes, Artificial Bee Colony***Abstrak**

Fuzzy K-Modes merupakan metode klasterisasi data yang efektif untuk data kategorikal. Metode ini perluasan fuzzy k-means dengan menggunakan modes (modus) dalam proses pencocokan ukuran ketidaksamaan (dissimilarity measure) untuk memutakhirkan titik pusat klaster dan mendapatkan solusi yang optimal. Meskipun demikian Fuzzy K-Modes memiliki kelemahan adanya kemungkinan berhenti dalam solusi lokal optimal. Artificial Bee Colony (ABC) merupakan metode optimasi yang sudah terbukti efektif dan memiliki kemampuan mendapatkan solusi global. Penelitian ini mengusulkan hibridasi algoritma Artificial Bee Colony dengan Fuzzy K-Modes untuk klasterisasi data kategorikal. Implementasi hibridasi Artificial Bee Colony dengan Fuzzy K-Modes (ABC-FKMO) terbukti mampu meningkatkan performa klasterisasi data kategorikal khususnya dalam aspek nilai Objective Function, F-Measure, dan Accuracy. Hasil pengujian dengan dataset Soybean Disease, Breast Cancer dan Congressional Voting Records dari UCI data repository, menunjukkan rata-rata Accuracy sebesar 0.991, 0.615, dan 0.867. Objective Function lebih baik rata-rata sebesar 2,73 %, F-Measure lebih baik rata-rata sebesar 4,31 % dan Accuracy lebih baik rata-rata sebesar 5,16 %.

Keywords*Categorical Data, Clustering, Fuzzy K-Modes, Artificial Bee Colony***Abstract**

Fuzzy K-Modes is an effective method for clustering categorical data. This method is an extensions of fuzzy k-means algorithm by using modes in the process of matching the dissimilarity measure to update centroid of the cluster and to obtain the optimal solution. Nevertheless, Fuzzy K-Modes has the disadvantage of the possibility of stopping in the optimal local solution. Artificial Bee Colony (ABC) is an optimization method that has been proven effective and has the ability to obtain global solutions. This study proposes a hybridization between the Artificial Bee Colony algorithm and Fuzzy K-Modes for clustering categorical data. The implementation of hybridization between Artificial Bee Colony and Fuzzy K-Modes (ABC-FKMO) has been proven to be able to improve the performance of categorical data clustering especially in the aspects of Objective Function, F-Measure, and Accuracy. The test results with datasets of the Soybean Disease, Breast Cancer and Congressional Voting Records from the UCI data repository, showed the Accuracy averages of 0.991, 0.615, and 0.867. Objective Function is better at an average of 2.73%, F-Measure is better at an average of 4.31% and Accuracy is better at an average of 5.16%.

1. Pendahuluan

Dewasa ini, penelitian di bidang klasterisasi data kategorikal sudah mulai berkembang, walaupun perkembangannya masih jauh lebih sedikit dibanding klasterisasi pada tipe data numerik. Data kategorikal secara alami tidak bisa diperlakukan sebagai data numerik karena ada beberapa operasi dalam data numerik yang tidak bisa dilakukan dalam data kategorikal seperti mean

dan median. Sebagai contoh atribut data kategorikal adalah atribut berdomain jenis kelamin (pria, wanita), domain agama (Islam, Kristen, Katolik, Hindu dan sebagainya), dan domain etnis (mongoloid, kaukasoid, negroid).

Klasterisasi merupakan metode utama dalam berbagai aplikasi dalam analisa data statistik, *data mining*, *information retrieval*, pengolahan citra dan sebagainya. Klasterisasi bertujuan melakukan pengelompokan obyek atau data ke dalam

beberapa klaster atau kelompok sehingga obyek atau data dalam satu klaster memiliki tingkat *similarity* (kesamaan) yang maksimum dan data antar klaster memiliki *similarity* yang minimum [1] [2].

Berdasar pendekatan dalam penetapan keanggotaan dalam klaster, Metode klasterisasi secara umum dapat dibagi menjadi dua yaitu hard clustering dan *fuzzy clustering* [3], [4]. Pada Hard clustering, tiap obyek/data hanya dialokasikan ke dalam satu satu klaster baik selama operasi klasterisasi maupun dalam output klasterisasi. Sedang pada *fuzzy clustering*, selama operasi klasterisasi tiap obyek/data dialokasikan ke dalam beberapa klaster dan diberi derajat keanggotaan dengan nilai antara 0 dan 1. Output *fuzzy clustering* dapat diubah menjadi hard clustering dengan memilih nilai keanggotaan tertinggi.

Terdapat berbagai macam metode klasterisasi data kategorikal, diantaranya ROCK [5], LIMBO [5], K-Modes, Fuzzy K-Modes, dan lain lain. Salah satu metode yang paling populer adalah K-Modes Clustering [6]. K-Modes termasuk dalam *Hard Clustering*. K-Modes merupakan pengembangan dari metode K-Means agar dapat digunakan untuk klasterisasi data kategorikal. K-Modes menggunakan sebuah ukuran jarak (*dissimilarity*) berupa kecocokan suatu nilai atribut tiap dimensi terhadap titik pusat klaster, menggantikan mean dengan modus, dan menggunakan metode berbasis frekuensi untuk memutakhirkan modus dalam proses meminimalkan jarak (*dissimilarity*) dari seluruh data ke pusat klaster masing masing. Karena K-Modes yang dikembangkan Huang merupakan pengembangan K-Means, maka K-Modes mempunyai karakteristik dan kelemahan yang sama dengan K-Means. Kelemahan tersebut adalah tingkat akurasi hasil klaster sangat tergantung dari penentuan titik awal pusat klaster sehingga sensitif terhadap penentuan titik awal. Dan memungkinkan hasil klaster konvergen pada lokal optimal [7], [8].

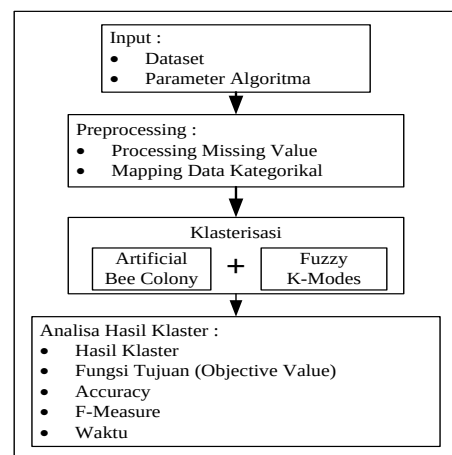
Untuk mengatasi masalah yang terjadi pada inialisasi pusat klaster dan keanggotaan klaster dalam metode K-Modes, Huang dan Ng mengusulkan sebuah metode baru yang diberi nama Fuzzy K-Modes [9]. Fuzzy K-Modes termasuk dalam *fuzzy clustering*. Merupakan pengembangan Fuzzy C-Means dengan menggunakan generate *fuzzy partition matrix* dari data kategorikal. Metode Fuzzy K-Modes juga masih mungkin terjadi permasalahan lokal optimal [8]. Untuk mengatasi masalah lokal optimal ini, dibutuhkan suatu metode yang mampu membantu mencari sebuah solusi global optimal dan menghindari kemungkinan solusi lokal optimal. Terdapat beberapa penelitian sebelumnya dalam rangka membantu mencari sebuah solusi global optima dari Fuzzy K-Modes diantaranya dengan pembobotan fitur secara otomatis [10] dan menggunakan koefisien jaccard untuk mengukur *dissimilarity* [11].

Artificial Bee Colony (ABC) merupakan suatu metode *metaheuristik* yang mengadopsi perilaku koloni lebah madu dalam mencari makan (*foraging behavior*). Metode ABC telah terbukti memiliki kemampuan untuk menangani permasalahan lokal optimal dan memiliki kualitas yang lebih baik dibandingkan dengan metode *metaheuristik* lainnya seperti Algoritma Genetika, Particle Swarm Optimization, dan Differential Evolution [12].

Dalam penelitian ini diusulkan sebuah metode baru untuk klasterisasi data kategorikal berbasis pada hibridasi metode Fuzzy K-Modes dan Artificial Bee Colony (ABC) yang diberi nama ABC-FKMO. ABC digunakan untuk membantu Fuzzy K-Modes agar dapat keluar dari permasalahan lokal optimal sehingga dengan metode ABC-FKMO ini diharapkan mampu mengoptimalkan posisi titik pusat klaster yang mengarah pada solusi global optimal.

2. Metodologi Penelitian

Pada Gambar 1 dijelaskan gambaran tentang metode yang diusulkan dalam penelitian. Dua tahap penting dalam penelitian ini adalah *Preprocessing*, dan klasterisasi. *Preprocessing* dilakukan untuk mengubah nilai missing value dengan modus tiap fitur. Berikutnya data yang sudah di *Preprocessing* diklasterisasi dengan Fuzzy K-modes dan dilakukan optimasi terhadap Fuzzy K-Modes menggunakan Artificial Bee Colony. Setelah tahap klasterisasi, parameter uji bisa dihitung untuk mengevaluasi kinerja metode yang diusulkan. Parameter yang digunakan adalah *Objective Function*, *Accuracy*, *F-Measure*, dan waktu komputasi.



Gambar 1. Metodologi Penelitian

3.1 Fuzzy K-Modes

Metode klasterisasi Fuzzy K-Modes merupakan versi fuzzy dari K-Modes dan juga merupakan pengembangan dan modifikasi dari metode Fuzzy C-Means agar dapat digunakan untuk klasterisasi data kategorikal. Fuzzy K-Modes pertama kali dikenalkan oleh Huang [6]. Secara

prinsip, perbedaan antara K-Modes dan Fuzzy K-Modes terletak pada proses penentuan pusat klaster dan Penentuan membership tiap titik.

Berikut ini akan diberikan gambaran metode Klasterisasi Fuzzy K-Modes. Anggap $D=\{X,Y\}$ adalah sebuah *dataset* kategorikal. X dan Y adalah dua obyek kategorikal yang dideskripsikan oleh $[x_1, x_2, x_3, \dots, x_m]$ dan $[y_1, y_2, y_3, \dots, y_m]$. Maka ukuran ketidaksamaan antara X dan Y dapat didefinisikan sebagai total ketidaksamaan pada atribut kategori yang berkorespondensi pada dua obyek. Semakin kecil jumlah ketidaksamaan, semakin mirip dua obyek tersebut. Ketidaksamaan tersebut dapat didefinisikan sebagai berikut :

$$d(X, Y) = \sum_{j=1}^m \delta(x_j, y_j) \quad (1)$$

Dimana

$$\delta(x_j, y_j) = \begin{cases} 0 & x_j = y_j \\ 1 & x_j \neq y_j \end{cases}$$

$d(X, Y)$ menggambarkan ketidaksamaan untuk tiap atribut.

Fungsi obyektif dari Fuzzy K-Modes adalah mencari W dan Z untuk meminimalisasi

$$F(W, Z) = \sum_{i=1}^k \sum_{l=1}^n w_{li}^\alpha d(Z_l, X_i) \quad (2)$$

Dimana $\alpha > 1$ adalah komponen pembobotan ($\alpha = 1$ digunakan untuk hard K-Modes). $W = (w_{li})$ merupakan matrik keanggotaan fuzzy $k \times n$. $Z = \{z_1, z_2, z_3, \dots, z_k\}$ merupakan himpunan pusat klaster dan z_i merupakan pusat klaster ke i dengan atribut kategorikal $A_1, A_2, A_3, \dots, A_m$.

Untuk melakukan update titik pusat klaster diberikan estimasi dari W , Huang [9] menyediakan teorema berikut :

Teorema 1 : Nilai $F_c(W, Z)$ dalam persamaan 2 diminimalisasi jika dan hanya jika $z_{ij} = a_{jr} \in \text{DOM}(A_j)$ dimana

$$r = \arg \max_{1 \leq t \leq n_j} \sum_{i, x_{ij}=a_{jr}} w_{li}^\alpha \quad (3)$$

Dimana

$$\sum_{i, x_{ij}=a_{jr}} w_{li}^\alpha \geq \sum_{i, x_{ij}=a_{jt}} w_{li}^\alpha, 1 \leq t \leq n_j$$

$$\text{untuk } 1 \leq j \leq d \text{ dan } 1 \leq t \leq k$$

Untuk update matrik keanggotaan fuzzy, diberikan estimasi dari Z . Huang [9] memberikan teorima berikut :

Teorema 2 : Jika $Z = \{z_1, z_2, \dots, z_k\}$ ditetapkan. Maka matrik keanggotaan fuzzy W yang meminimalisasi nilai $F_c(W, Z)$ yang didefinisikan di persamaan 2 dijabarkan berikut :

$$w_{li} = \begin{cases} 1 & \text{jika } x_i = z_l ; \\ 0 & \text{jika } x_i \neq z_l, h \neq l \\ \frac{1}{\sum_{h=1}^k \left[\frac{d(x_i, z_l)}{d(x_i, z_h)} \right]^{\frac{1}{\alpha-1}}} & \text{jika yang lainnya, } 1 \leq i \leq k, l \leq i \leq n \end{cases}$$

Berdasar dari dua teorema di atas, algoritma Fuzzy K-modes dapat diimplementasikan secara rekursif. Dengan asumsi bahwa r adalah jumlah iterasi maksimal, berikut adalah langkah langkah metode Fuzzy K-Modes [8]:

1. Inisialisasi : pilih inisial pusat klaster (Z_0) awal secara random.
2. Tentukan W_0 dimana fungsi biaya $F(W_0, Z_0)$ diminimalisasi.
3. For $t = 1$ to r do
4. Tentukan Z_1 dimana fungsi biaya $F(W_0, Z_1)$ diminimalisasi;
5. if $F(W_0, Z_1) = F(W_0, Z_0)$ then
6. Berhenti;
7. Else
8. Tentukan tentukan W_1 dimana fungsi biaya diminimalisasi
9. if $F(W_1, Z_1) = F(W_0, Z_1)$ then
10. Berhenti;
11. Else
12. $W_0 \leq W_1$
13. End if
14. End if
15. End for

3.2 Artificial Bee Colony

Metode ABC pertama kali dikenalkan oleh Karaboga [12]–[15]. Terinspirasi dari perilaku alami lebah madu dalam proses mendapatkan sumber makanan dengan mengetahui kualitas dan lokasi lebah madu.

Dalam model metode ABC ini, Ada tiga jenis artificial bee dalam koloni lebah yang digunakan yaitu: lebah pekerja (*employed bee*), lebah penunggu (*onlooker*), dan lebah scout. Setengah jumlah koloni terdiri atas lebah pekerja dan setengahnya adalah lebah penunggu. Jumlah lebah pekerja sama dengan jumlah sumber makanan di sekitar sarang karena dalam algoritma ini diasumsikan adanya satu lebah pekerja untuk satu sumber makanan. Lebah pekerja yang sudah meninggalkan sumber makanannya berubah menjadi lebah scout.

Sama seperti algoritma berbasis intelligent swarm yang lain, metode ABC adalah sebuah proses iteratif. Pendekatan dimulai dengan sebuah populasi solusi (atau sumber makanan) yang dihasilkan secara acak, diikuti langkah-langkah berikutnya yang diulang sampai kriteria penghentian terpenuhi. Langkah-langkah utama dari metode ABC diberikan di bawah ini.

1. Inisialisasi Populasi
2. Tempatkan lebah pekerja ke sumber makanan
3. Tempatkan lebah penunggu pada sumber makanan berdasarkan jumlah nektarnya
4. Kirim lebah scout ke daerah pencarian untuk menemukan sumber makanan baru
5. Mengingat sumber makanan terbaik yang telah ditemukan

Tabel 1. rata rata hasil uji coba dengan dataset soybean

Eksponen Fuzzy	Ukuran	Objective Function		F-Measure		Accuracy		Waktu	
		Fuzzy K-Modes	ABC-FKMO	Fuzzy K-Modes	ABC-FKMO	Fuzzy K-Modes	ABC-FKMO	Fuzzy K-Modes	ABC-FKMO
1.1	Rata	238,62	214,16	0,940	0,990	0,894	0,981	0,19	35,55
	stdev	34,72	0,29	0,063	0,003	0,112	0,007	0,15	4,36
1.2	Rata	216,63	210,35	0,976	1,000	0,955	1,000	0,15	30,89
	stdev	13,69	0,00	0,044	0,000	0,079	0,000	0,15	5,49
1.3	Rata	211,43	201,26	0,966	0,997	0,938	0,994	0,12	35,49
	stdev	24,58	0,13	0,053	0,005	0,095	0,010	0,14	4,52

Jika persyaratan penghentian tidak terpenuhi, balik ke langkah 2.

3. Skenario Uji Coba

3.1 Dataset

Dataset yang digunakan dalam penelitian ini terdiri dari dataset *Small Soybean Disease*, *Breast Cancer*, dan *Congressional Voting Record (Voting)*. Dataset tersebut diunduh dari UCI Machine Learning Repository [16].

Small Soybean Disease adalah dataset versi kecil dari dataset *Soybean Disease*. Dataset ini terdiri dari 35 atribut dan 47 data serta tidak terdapat *missing value*. Dataset ini bertujuan mengklasifikasi jenis penyakit yang diderita tumbuhan kacang kedelai. Semua atribut dataset ini bertipe kategorikal.

Breast Cancer adalah dataset tentang kanker payudara ini diperoleh dari University Medical Center, Institut Onkologi, Ljubljana, Yugoslavia. Terdiri dari 9 atribut dan data berjumlah 286.

Congressional Voting Record (Voting) adalah dataset yang berisi suara masing-masing Anggota Kongres DPR Amerika Serikat atas 16 voting pada tahun 1984. Dataset ini terdiri dari 435 data dan item *missing value* berjumlah 392.

Semua dataset tersebut memiliki atribut kategorikal. Pada Tabel 1 dapat di lihat deskripsi mengenai jumlah atribut, kelas, dan data dalam

Tabel 2. Deskripsi Dataset

Dataset	Jumlah atribut	Jumlah kelas	Jumlah Data
Soybean	35	4	47
Breast Cancer	9	2	286
Voting	16	2	435

penelitian ini.

3.2 Parameter Uji coba

Uji coba dilakukan untuk membandingkan pengaruh penggunaan ABC untuk optimasi Fuzzy K-Modes. Oleh karena itu akan dibandingkan metode dasar dan metode yang diusulkan dalam beberapa parameter uji yaitu : *Objective Function*, *Precision*, *Recall*, *F-Measure*, *Accuracy*, dan waktu komputasi. *Precision* adalah tingkat relevansi data yang dihasilkan sistem. *Precision* dihasilkan dari pembagian antara tingkat data hasil sistem yang relevan dengan seluruh hasil yang didapat. *Recall* adalah tingkat data yang dihasilkan sistem data yang relevan. *F-Measure* merupakan pengukuran yang menggabungkan *precision* dan *recall*. *Accuracy* adalah tingkat keakuratan / akurasi data yang dihasilkan sistem. Secara umum, parameter uji tersebut dapat dihitung dengan menggunakan rumus berikut :

$$Precision = \frac{TP}{TP + FP} \quad (5)$$

$$Recall = \frac{TP}{TP + FN} \quad (6)$$

$$FMeasure = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (7)$$

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (8)$$

Metode dasar dan metode yang diusulkan diuji coba dengan ketiga dataset. Parameter untuk Fuzzy K-Modes menggunakan eksponen Fuzzy 1.1, 1.2, dan 1.3, maksimum iterasi 100 dan threshold 0,001. Parameter ABC menggunakan *Max Cycle* 20,

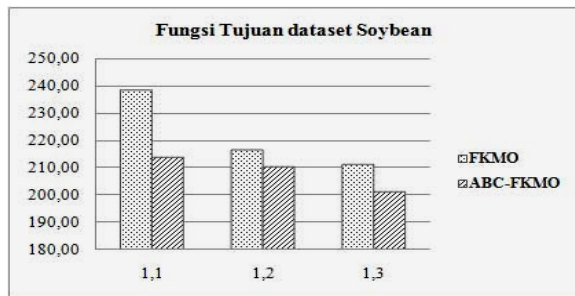
Tabel 3. rata rata hasil uji coba dengan Dataset Breast Cancer

Eksponen Fuzzy	Ukuran	Objective Function		F-Measure		Accuracy		Waktu	
		Fuzzy K-Modes	ABC-FKMO	Fuzzy K-Modes	ABC-FKMO	Fuzzy K-Modes	ABC-FKMO	Fuzzy K-Modes	ABC-FKMO
1.1	Rata	1.084,93	1.056,48	0,651	0,723	0,553	0,608	0,26	23,97
	stdev	32,85	6,45	0,058	0,015	0,047	0,017	0,27	4,17
1.2	Rata	1.051,86	1.022,59	0,667	0,732	0,572	0,619	0,11	17,75
	stdev	39,65	0,00	0,081	0,000	0,075	0,000	0,19	6,27
1.3	Rata	980,46	973,97	0,671	0,732	0,572	0,619	0,15	17,61
	stdev	8,28	0,00	0,059	0,000	0,043	0,000	0,21	4,38

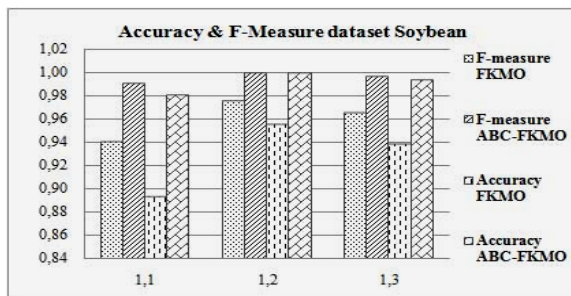
limit 10, jumlah koloni lebah 10. Tiap metode dan tiap parameter diuji sebanyak 10 kali.

4. Hasil Uji Coba

Hasil uji coba dengan menggunakan *dataset* soybean disease dapat dilihat pada Tabel 2. Dari Tabel 2 didapatkan hasil dari ketiga eksponen fuzzy, metode ABC-FKMO lebih unggul dengan menghasilkan rata nilai *Objective Function* yang lebih kecil, rata rata *F-Measure* yang lebih besar, rata rata *Accuracy* yang lebih besar, dan Standart devisasi yang lebih kecil dibanding Fuzzy K-Modes. ABC-FKMO juga menghasilkan nilai *Objective Function* paling baik (minimum). ABC-FKMO dan Fuzzy K-Modes dapat menghasilkan nilai *F-Measure* terbaik (maksimal) yang sama tetapi ABC-FKMO lebih stabil dengan menghasilkan standart deviasi yang lebih kecil. Pada pada Grafik 1 dan Grafik 2 dapat dilihat gambaran perbandingan performa dengan *dataset* soybean dalam tiga skenario pengujian.



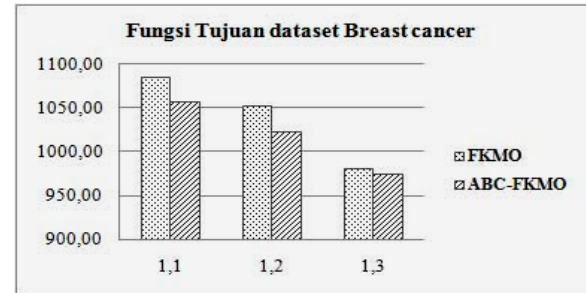
Grafik 1. Objective Function dataset Soybean



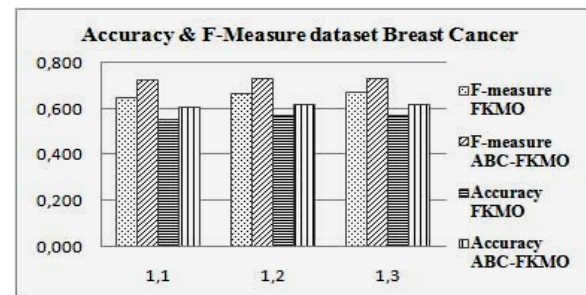
Grafik 2. Accuracy dan F-Measure dataset Soybean

Hasil uji coba dengan menggunakan *dataset* Breast cancer dapat dilihat pada Tabel 3. Dari tabel 3 terlihat bahwa nilai *Objective Function* klasterisasi data kategorikal dengan metode ABC-FKMO unggul disemua uji coba yang telah dilakukan. Hal ini berarti bahwa metode ABC-FKMO telah mampu mengoptimalkan posisi titik pusat klaster yang mengarah ke solusi global optimal. Dari Nilai *F-Measure* dan *Accuracy*, juga memperlihatkan metode ABC-FKMO unggul di semua uji coba yang dilakukan. Hal ini berarti bahwa metode ABC-FKMO juga mempunyai tingkat akurasi yang lebih baik dari Fuzzy K-Modes. Dari standart deviasi, juga memperlihatkan metode ABC-FKMO unggul di semua uji coba yang

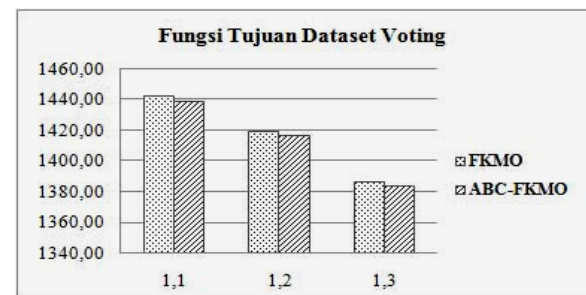
dilakukan. Hal ini berarti bahwa metode ABC-FKMO mempunyai performa yang lebih stabil. Dalam tolok ukur waktu, ABC-FKMO lebih lambat dari Fuzzy K-Modes. Pada Grafik 3 dan Grafik 4 dapat dilihat gambaran perbandingan performa dengan *dataset* Breast cancer dalam tiga skenario pengujian dapat dilihat.



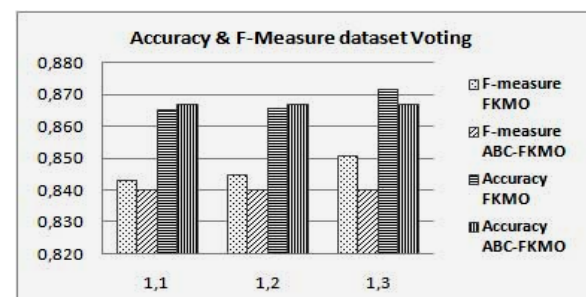
Grafik 3. Objective Function Dataset Breast Cancer



Grafik 4. Accuracy dan F-Measure Dataset Breast Cancer



Grafik 5. Objective Function Dataset voting



Grafik 6. Accuracy dan F-Measure Dataset voting

Hasil uji coba dengan menggunakan *dataset* Voting dapat dilihat pada Tabel 4. Dari tabel tersebut didapatkan hasil yang memperlihatkan metode ABC-FKMO menghasilkan *Accuracy* yang lebih tinggi dari Fuzzy K-Modes pada eksponen fuzzy 1.1 dan 1.2, tetapi menghasilkan *Accuracy* yang lebih rendah di eksponen fuzzy 1.3. Dari standart deviasi, memperlihatkan metode ABC-

Tabel 4. rata rata hasil uji coba dengan Dataset Voting

Eksponen Fuzzy	Ukuran	Objective Function		F-Measure		Accuracy		Waktu	
		Fuzzy K-Modes	ABC-FKMO	Fuzzy K-Modes	ABC-FKMO	Fuzzy K-Modes	ABC-FKMO	Fuzzy K-Modes	ABC-FKMO
1.1	Rata	1.442,28	1.439,33	0,843	0,840	0,865	0,867	0,11	155,64
	stdev	2,63	0,00	0,011	0,000	0,010	0,000	0,07	7,09
1.2	Rata	1419,77	1416,93	0,844	0,840	0,866	0,867	0,11	23,02
	stdev	2,19	0,00	0,011	0,000	0,011	0,000	0,06	3,23
1.3	Rata	1.386,79	1.384,21	0,851	0,840	0,871	0,867	0,17	13,95
	stdev	1,07	0,00	0,010	0,000	0,008	0,000	0,30	2,76

FKMO unggul di semua uji coba yang dilakukan. Dalam tolok ukur waktu, ABC-FKMO juga lebih lambat dari Fuzzy K-Modes. Dari hasil ujicoba dengan dataset Voting, juga dapat dilihat bahwa walau ABC-FKMO menghasilkan *Objective Function* yang lebih baik dan menghasilkan performa lebih stabil, tetapi belum tentu menghasilkan akurasi yang lebih baik dibanding Fuzzy K-Modes. Pada Grafik 5 dan Grafik 6 dapat dilihat gambaran perbandingan performa dengan dataset Breast cancer dalam tiga skenario pengujian.

5. Pembahasan Hasil

Hasil uji coba dengan menggunakan tiga dataset dan tiga parameter eksponen Fuzzy menunjukkan bahwa metode ABC-FKMO menghasilkan kinerja lebih baik dibanding Fuzzy K-Modes. Hal ini dapat dilihat dari hasil rata-rata *Objective Function*, *Accuracy*, *F-Measure*. Dari segi rata-rata *Objective Function*, ABC-FKMO lebih baik 2,73%, dari segi rata-rata *F-Measure* unggul 4,31%, dan dari segi rata-rata *Accuracy* unggul 5.16% dibanding Fuzzy K-Modes. Namun demikian waktu komputasi yang dibutuhkan untuk menjalankan ABC-FKMO lebih besar lama dari Fuzzy K-Modes.

Tabel 5 Rata-rata selisih kinerja ABC-FKMO dan Fuzzy K-Modes

Dataset	Exp Fuzzy	Obj Func (%)	F-Meas (%)	Accuracy (%)	Waktu (Kali)
Soybean Disease	1.1	10,25	5,3	9,76	82,01
	1.2	2,9	2,5	4,68	93,25
	1.3	4,81	3,21	5,9	91,2
Breast Cancer	1.1	2,62	11,13	9,99	123,95
	1.2	2,78	9,69	8,12	65,19
	1.3	0,66	9,13	8,19	88,77
Voting	1.1	0,2	-0,33	0,21	153,2
	1.2	0,2	-0,54	0,13	103,37
	1.3	0,19	-1,27	-0,53	83,81
rata-rata		2,73	4,31	5,16	98,31

Hasil dari rata rata selisih kinerja yang ditampilkan pada Tabel 5 menunjukkan bahwa nilai *Objective function* metode ABC-FKMO selalu lebih optimal, tetapi belum tentu menghasilkan nilai *F-Measure* atau *Accuracy* yang lebih baik. Hal ini dapat terjadi karena klasterisasi merupakan suatu proses Unsupervised learning. Klasterisasi hanya menggunakan nilai dari fitur untuk proses

klasterisasi tanpa melibatkan kelas label untuk mengarahkan proses. Hal ini menyebabkan nilai *Objective Function* klasterisasi yang lebih baik memungkinkan untuk memiliki *F-Measure* atau *Accuracy* yang lebih jelek.

Dari segi *Accuracy* dan *F-Measure*, ketika ABC-FKMO menghasilkan nilai *Accuracy* lebih baik, belum tentu juga menghasilkan *F-Measure* lebih baik atau sebaliknya. Hal ini disebabkan *Accuracy* hanya mempertimbangkan jumlah pelabelan kelas yang benar dan salah tanpa mempertimbangkan nilai *precision* klaster dan *recall* dari kelas klaster itu sendiri. Sedangkan *F-Measure* menggunakan pertimbangan *precision* dan *recall* dari pelabelan kelas klaster yang ada.

Dari segi waktu, ABC-FKMO memerlukan waktu yang lebih panjang dengan rata rata 98,31 kali lebih lama di banding Fuzzy K-Modes. Hal ini disebabkan oleh karena ABC sebagai metode optimasi untuk Fuzzy K-Modes, melakukan proses berulang untuk mencari solusi paling optimal. Banyaknya perulangan tergantung pada pemilihan inputan parameter. Parameter tersebut antara lain jumlah iterasi (*Cycle*) ABC, Jumlah koloni lebah, dan limit maksimal suatu solusi. Semakin banyak jumlah iterasi (*Cycle*) ABC dan Jumlah koloni lebah serta semakin kecil limit, maka akan makin banyak iterasi ABC-FKMO yang berimbas pada lamanya waktu *running*.

6. Kesimpulan

Dari seluruh hasil uji coba dan analisa dapat disimpulkan bahwa metode ABC-FKMO terbukti berhasil mengoptimalkan penempatan posisi titik pusat klaster dengan mengarahkan hasil klasterisasi data kategorikal menuju solusi global optimal. Hal ini dibuktikan dengan hasil penelitian yang menunjukkan nilai *Objective Function* selalu lebih kecil dari Fuzzy K-Modes dengan selisih rata rata 2,73%. Dalam hal pengukuran *F-Measure* dan *Accuracy*, metode ABC-FKMO juga menunjukkan keunggulan-nya dibandingkan dengan metode Fuzzy K-Modes. Hal ini dibuktikan dengan hasil penelitian yang menunjukkan bahwa *F-Measure* ABC-FKMO lebih baik 4.31% dan *Accuracy* ABC-FKMO lebih baik 5.16%. Dari sisi waktu komputasi, metode ABC-FKMO membutuhkan waktu komputasi yang lebih lama dibandingkan dengan metode Fuzzy K-Modes.

Daftar Pustaka

- [1] P.-N. Tan, M. Steinbach, and V. Kumar, "Introduction to Data Mining," 2005.
- [2] J. Han, M. Kamber, and J. Pei, "Data Mining: Concepts and Techniques," *Data Min. Concepts Tech.*, 2012.
- [3] A. K. Jain, M. N. Murty, and P. J. Flynn, "Data clustering: a review," *ACM Comput. Surv.*, vol. 31, no. 3, pp. 264–323, 1999.
- [4] A. K. Jain, "Data clustering: 50 years beyond K-means," *Pattern Recognit. Lett.*, vol. 31, no. 8, pp. 651–666, 2010.
- [5] S. Guha, R. Rastogi, and K. Shim, "Rock: a robust clustering algorithm for categorical attributes," *Inf. Syst.*, vol. 25, no. 5, pp. 345–366, 2000.
- [6] Z. Huang, "Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values," *Data Min. Knowl. Discov.*, vol. 2, no. 3, pp. 283–304, Sep. 1998.
- [7] R. E. Funderlic, M. T. Chu, N. Orlowski, D. Schlor, and J. Blevins, "Convergence and Other Aspects of the k-modes Algorithm for Clustering Categorical data," 2019.
- [8] G. Gan, J. Wu, and Z. Yang, "A genetic fuzzy k-Modes algorithm for clustering categorical data," *Expert Syst. Appl.*, vol. 36, no. 2 PART 1, pp. 1615–1620, 2009.
- [9] Z. Huang and M. K. Ng, "A Fuzzy K-modes Algorithm for Clustering Categorical Data," *Trans. Fuz Sys.*, vol. 7, no. 4, pp. 446–452, Aug. 1999.
- [10] A. Saha and S. Das, "Categorical fuzzy k-modes clustering with automated feature weight learning," *Neurocomputing*, vol. 166, pp. 422–435, 2015.
- [11] F. Cao, J. Z. Huang, and J. Liang, "A fuzzy SV-k-modes algorithm for clustering categorical data with set-valued attributes," *Appl. Math. Comput.*, vol. 295, pp. 1–15, 2017.
- [12] D. Karaboga and B. Akay, "A comparative study of Artificial Bee Colony algorithm," *Appl. Math. Comput.*, vol. 214, no. 1, pp. 108–132, 2009.
- [13] D. Karaboga, "An Idea Based on Honey Bee Swarm for Numerical Optimization, Technical Report - TR06," *Tech. Report, Erciyes Univ.*, 2005.
- [14] D. Karaboga and B. Basturk, "A powerful and efficient algorithm for numerical function optimization: Artificial bee colony (ABC) algorithm," *J. Glob. Optim.*, vol. 39, no. 3, pp. 459–471, 2007.
- [15] D. Karaboga and B. Basturk, "On the performance of artificial bee colony (ABC) algorithm," *Appl. Soft Comput.*, vol. 8, no. 1, pp. 687–697, 2008.
- [16] D. Dheeru and E. Karra Taniskidou, "{UCI} Machine Learning Repository." 2017.