

Penerapan Algoritma Invasive Weed Optimnization untuk Penentuan Titik Pusat Klaster pada K-Means

I Putu Adi Pratama*¹, Agus Harjoko²

¹Program Studi S2 Ilmu Komputer, FMIPA UGM, Yogyakarta

²Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta

e-mail: *¹putudipa@gmail.com, ²aharjoko@ugm.ac.id

Abstrak

K-means merupakan salah satu algoritma clustering yang paling populer. Salah satu alasan dari kepopuleran K-means adalah karena mudah dan sederhana ketika diimplementasikan. Namun hasil klaster dari K-means sangat sensitif terhadap pemilihan titik pusat awalnya. K-means seringkali terjebak pada solusi lokal optima. Hasil klaster yang lebih baik seringkali baru bisa didapatkan setelah dilakukan beberapa kali percobaan. Penyebab lain seringkali K-means terjebak pada solusi lokal optima adalah karena cara penentuan titik pusat baru untuk setiap iterasi dalam K-means dilakukan dengan menggunakan nilai mean dari data-data yang ada pada klaster bersangkutan. Hal tersebut menyebabkan K-means hanya akan melakukan pencarian calon titik pusat baru disekitar titik pusat awal. Untuk mengatasi permasalahan tersebut, penerapan metode yang memiliki kemampuan untuk melakukan pencarian global akan mampu membantu K-means untuk dapat menemukan titik pusat klaster yang lebih baik. Invasive Weed Optimization merupakan algoritma pencarian global yang terinspirasi oleh proses kolonisasi rumput liar. Pada penelitian ini diusulkan sebuah metode yang merupakan hasil hibridasi dari metode K-means dan algoritma Invasive Weed Optimization (IWOKM). Kinerja dari metode IWOKM telah dicobakan pada data bunga Iris kemudian hasilnya dibandingkan dengan K-means. Dari pengujian yang dilakukan, didapat hasil bahwa metode IWOKM mampu menghasilkan hasil klaster yang lebih baik dari K-means.

Kata kunci—K-means, IWO, IWOKM, analisa klaster

Abstract

K-means is one of the most popular clustering algorithm. One reason for the popularity of K-means is it is easy and simple when implemented. However, the results of K-means is very sensitive to the selection of initial centroid. The results are often better after several experiment. Another reason why K-means stuck in local optima is due to the method of determining the new center point for each iteration that is performed using the mean value of the data that exist on the cluster. This causes the algorithm will do search for the centroid candidates around the center point. To overcome this, implement a method that is able to do a global search to determine the center point on K-means may be able to assist K-means in finding better cluster center. Invasive Weed Optimization (IWO) is a global search algorithm inspired by weed colonization process. In this study proposed a method which is the result of hybridization of K-means and IWO (IWOKM). Performance of the method has been tested on flower Iris dataset. The results are then compared with the result from K-means. The result show that IWOKM able to produce better cluster center than K-means.

Keywords—K-means, IWO, IWOKM, cluster analysis

1. PENDAHULUAN

K-means merupakan salah satu algoritma *clustering* yang paling populer. Salah satu alasan dari kepopuleran *K-means* adalah karena mudah dan sederhana ketika diimplementasikan [1]. Beberapa literatur [1,2,3] menyatakan bahwa hasil kluster dari *K-means* sangat sensitif terhadap pemilihan titik pusat awalnya. Untuk setiap percobaan yang dilakukan, *K-means* cenderung menghasilkan hasil kluster yang berbeda-beda. Hasil kluster yang lebih baik seringkali didapatkan setelah dilakukan beberapa kali percobaan, tapi sulit untuk menentukan batasan eksperimen agar *K-means* selalu mendapatkan hasil yang baik [2]. Keadaan belum mampunya *K-means* untuk menghasilkan hasil kluster terbaik diistilahkan dengan terjebaknya *K-means* pada solusi lokal optima [4]. Solusi lokal optima merupakan suatu kondisi dimana sudah ditemukannya puncak dari fungsi tujuan padahal pencarian baru dilakukan pada sebagian kecil ruang pencarian. Fungsi tujuan dari *K-means* adalah menemukan solusi dimana tidak ada solusi lain yang memiliki nilai *SSE* (*Sum Squared Error*) lebih kecil dari solusi yang ditemukan [5]. *SSE* merupakan hasil penjumlahan dari seluruh jarak masing-masing data dengan titik pusat klasternya. Semakin kecil nilai *SSE* yang didapat, semakin seragam data yang ada didalam masing-masing kluster, semakin baik kluster yang dihasilkan.

Alasan lain kurang mampunya *K-means* didalam menemukan solusi global adalah karena cara penentuan titik pusat baru untuk setiap iterasi dalam *K-means* dilakukan dengan menggunakan nilai *mean* dari data-data yang ada pada kluster bersangkutan. Cara tersebut hanya mampu membuat *K-means* untuk melakukan penelusuran calon titik pusat baru pada setiap iterasinya dalam wilayah yang sempit disekitar titik pusat awal yang ditentukan secara acak [6]. Sekiranya penerapan algoritma yang dapat melakukan pencarian calon solusi pada area yang lebih luas untuk penentuan titik pusat baru pada *K-means* akan dapat membantu *K-means* untuk menemukan titik pusat kluster yang lebih baik.

Algoritma *Invasive Weed Optimization* yang diusulkan oleh [7] merupakan algoritma pencarian global yang terinspirasi oleh proses kolonisasi rumput liar. Rumput liar merupakan tumbuhan yang kuat didalam penyebarannya juga adaptif terhadap perubahan lingkungan sehingga dapat menjadi ancaman bagi tumbuhan budidaya. *Algoritma Invasive Weed Optimization (IWO)* mencoba menirukan sifat acak dan adaptif dari penyebaran rumput liar didalam membangun koloni. Ide dari algoritma *IWO* adalah menyebarkan rumput pada area yang luas dan sempit sekaligus, yaitu dengan memanfaatkan angka acak yang penyebarannya mengikuti distribusi normal.

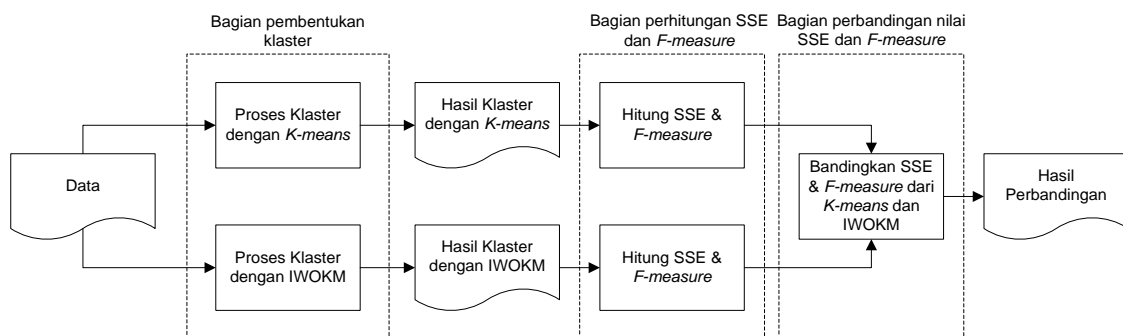
Pada penelitian ini diusulkan sebuah metode klasterisasi data yang merupakan hasil hibridasi dari algoritma *IWO* dengan algoritma *K-means*. Diharapkan dengan menerapkan karakteristik pertumbuhan dan penyebaran benih dari algoritma *Invasive Weed Optimization* untuk menentukan titik pusat baru pada *K-means* nantinya dapat membantu *K-means* untuk tidak terjebak pada solusi lokal optima dan mampu mengarahkan hasil kluster menuju solusi global optima.

2. METODE PENELITIAN

Pada penelitian ini dilakukan hibridasi antara *K-means* dan algoritma *Invasive Weed Optimization*, yang selanjutnya akan disebut *IWOKM*. *IWO* digunakan didalam penentuan titik pusat baru untuk setiap iterasi yang terjadi pada *K-means*. Titik pusat awal kluster pada *K-means* akan menjadi rumput induk pada *IWO* yang selanjutnya akan menghasilkan rumput anak dan disebar pada area pencarian. Sebelum dapat menghasilkan rumput anak, perlu dicari anggota dari masing-masing rumput induk untuk selanjutnya dihitung nilai *fitness*-nya. Selanjutnya nilai *fitness* dari rumput induk akan digunakan didalam penentuan jumlah rumput anak yang dapat dihasilkan oleh masing-masing rumput induk. Proses penyebaran rumput anak pada area pencarian dilakukan dengan memanfaatkan angka acak yang penyebarannya mengikuti distribusi normal dengan nilai *mean* sama dengan posisi dari rumput induk dan nilai standar

deviasi yang digunakan adalah berubah-ubah untuk setiap iterasi yang terjadi. Dengan mekanisme tersebut, dimungkinkan untuk menyebar titik pusat kluster pada area yang luas dan sempit sekaligus sehingga memberikan kesempatan lebih besar pada *K-means* untuk dapat menemukan titik pusat kluster terbaiknya.

Untuk keperluan pengukuran tingkat keberhasilan metode *IWOKM* dalam hal pencapaian solusi global, maka dilakukan perbandingan nilai fungsi objektif (SSE) dari hasil kluster yang diperoleh oleh metode *IWOKM* dengan *K-means*. Selain pengujian terhadap nilai SSE, juga dilakukan pengujian terhadap nilai *F-measure* yang merupakan ukuran kemampuan dari algoritma untuk menemukan kembali kelas label dari masing-masing data. Tahapan-tahapan yang perlu dilakukan untuk mendapatkan hasil perbandingan nilai SSE dan *F-measure* antara *IWOKM* dan *K-means* adalah seperti terlihat pada Gambar 1.



Gambar 1 Alur sistem pembentukan dan perbandingan hasil kluster *K-means* dan *IWOKM*

Gambar 1 menggambarkan hubungan antara bagian pembentukan kluster, proses perhitungan nilai SSE, *F-measure* dan proses perbandingan nilai SSE & *F-measure*. Proses dimulai dengan menerima data uji, selanjutnya data diproses pada bagian pembentukan kluster dengan masing-masing metode *K-means* dan *IWOKM*. Hasil kluster dari masing-masing metode akan menjadi data masukan pada bagian perhitungan SSE & *F-measure*. Pada tahap akhir, dilakukan perbandingan hasil antara nilai SSE & *F-measure* dari algoritma *K-means* dan *IWOKM*.

Secara garis besarnya *IWOKM* dapat dibagi menjadi 6 fase, yaitu fase inialisasi, fase penentuan keanggotaan kluster, fase reproduksi dan penyebaran rumput anak, fase penentuan kombinasi calon solusi (koloni rumput), fase penentuan anggota masing-masing rumput pada masing-masing koloni rumput, dan terakhir fase pemilihan koloni rumput yang memiliki nilai fitness terbaik.

1. Fase Inialisasi

Fase inialisasi merupakan fase awal dari proses *IWOKM*. Terdapat tiga tahap yang masuk dalam fase inialisasi, yaitu menentukan nilai awal parameter *IWOKM*, menerima data masukan dan terakhir menentukan titik pusat awal secara acak. Parameter-parameter yang perlu diberi nilai awal adalah seperti terlihat pada Tabel 1.

Penentuan titik pusat awal pada *IWOKM* adalah sama dengan cara penentuan titik pusat awal pada *K-means*, yaitu dengan membangkitkan bilangan acak pada area pencarian sebanyak k jumlah kluster yang diinginkan. Titik pusat awal yang sekaligus merupakan calon solusi awal yang ditentukan secara acak ini akan menjadi rumput induk di dalam proses *IWO* yang selanjutnya akan menghasilkan rumput anak dan disebar pada area pencarian.

2. Fase Penentuan Anggota dari Masing-masing Rumput Induk

Fase ini masih sama dengan proses penentuan keanggotaan pada *K-means*, yaitu terdiri dari dua tahap yaitu: mencari jarak masing-masing data dengan seluruh titik pusat yang ada dan menentukan keanggotaan masing-masing kluster.

Perhitungan jarak antara masing-masing data dengan seluruh titik pusat yang ada dilakukan dengan menggunakan rumus *Euclidean Distance*. Untuk dua buah data dalam ruang berdimensi p , misalkan $x = (x_1, x_2, \dots, x_p)$ dan $y = (y_1, y_2, \dots, y_p)$. Jarak *Euclidean* antara x dan y didefinisikan sebagai persamaan (1).

$$d_{euc}(x, y) = \left[\sum_{j=1}^p (x_j - y_j)^2 \right]^{\frac{1}{2}} \quad (1)$$

Dimana x_j dan y_j adalah nilai dari x dan y pada atribut ke- j .

Penentuan keanggotaan kluster dilakukan dengan cara mengalokasikan data menjadi anggota kluster dimana jarak data ke titik pusat kluster bersangkutan adalah terpendek.

Tabel 1 Parameter-parameter *IWOKM*

Parameter	Keterangan
k	jumlah kluster
$iter_{max}$	jumlah iterasi maksimum
k_{max}	jumlah maksimum koloni
S_{max}	jumlah maksimum benih
S_{min}	jumlah minimum benih
n	<i>nonlinear modulation index</i>
$\sigma_{initial}$	nilai standar deviasi awal
σ_{final}	nilai standar deviasi akhir

3. Fase Reproduksi dan Penyebaran Rumput Anak

Fase ini akan terdiri dari 3 tahap, yaitu menghitung nilai *fitness* dari masing-masing kluster, menentukan jumlah rumput anak yang dapat dihasilkan oleh masing-masing rumput induk dan terakhir menentukan posisi persebaran rumput anak.

Fungsi *fitness* yang digunakan untuk menentukan jumlah rumput anak yang dapat dihasilkan oleh masing-masing rumput induk adalah nilai *Mean Squared Error* dari masing-masing kluster.

Sebuah tanaman diijinkan untuk menghasilkan bibit tergantung pada nilai *fitness*-nya. Jumlah bibit yang dapat dihasilkan oleh masing-masing tanaman meningkat secara linear dari kemungkinan terkecil sampai kemungkinan terbesar. Persamaan untuk menentukan jumlah rumput anak yang akan dihasilkan oleh masing-masing rumput induk seperti yang disampaikan oleh [8] adalah seperti terlihat pada persamaan (2).

$$n_{seed} = \frac{F_i - F_{worst}}{F_{best} - F_{worst}} (S_{max} - S_{min}) + S_{min} \quad (2)$$

Dimana,

n_{seed} = jumlah rumput anak yang akan dihasilkan

F_i = nilai *fitness* dari rumput ke i

F_{worst} = nilai *fitness* terburuk dalam koloni rumput

F_{best} = nilai *fitness* terbaik dalam koloni rumput

S_{max} = jumlah maksimum rumput anak yang bisa dihasilkan rumput induk

S_{min} = jumlah minimum rumput anak yang bisa dihasilkan rumput induk

Keacakan penyebaran rumput anak pada area pencarian terdapat pada bagian ini. Bibit yang telah dihasilkan selanjutnya disebar secara acak pada area pencarian dengan angka acak yang penyebarannya mengikuti distribusi normal dimana nilai *mean*-nya sama dengan

posisi dari rumput induk dan nilai standar deviasinya adalah berubah-ubah untuk setiap iterasi yang terjadi dengan mengikuti persamaan (3).

$$\sigma_{iter} = \frac{(iter_{max} - iter)^n}{(iter_{max})^n} (\sigma_{initial} - \sigma_{final}) + \sigma_{final} \quad (3)$$

Dimana σ_{iter} adalah nilai standar deviasi pada waktu sekarang, $iter_{max}$ merupakan jumlah maksimum iterasi, n merupakan *nonlinear modulation index*, $\sigma_{initial}$ nilai standar deviasi awal, dan σ_{final} nilai standar deviasi akhir.

4. Fase Penentuan Kombinasi Calon Solusi (Koloni Rumput)

Tahap ini terdiri dari dua proses utama yaitu menggabungkan nilai *mean* dari klaster induk dengan rumput anaknya, kemudian menentukan kombinasi calon solusi sebanyak k_{max} .

Setelah posisi persebaran rumput anak didapat, tahap selanjutnya yaitu menggabungkan rumput anak dengan *mean* dari klaster induknya. Hal ini dilakukan untuk memberikan kesempatan pada nilai *mean* dari klaster induk ikut bersaing dengan rumput anak untuk terpilih sebagai rumput induk pada generasi berikutnya.

Setelah nilai *meandari* klaster induk dan rumput anak digabung, tahap selanjutnya adalah menentukan kombinasi calon solusi sebanyak k_{max} . Kombinasi calon solusi ditentukan secara acak dengan ketentuan selalu menyertakan anggota salah satu rumput anak/*mean* dari klaster induk pada kombinasi yang terpilih.

5. Fase Penentuan Anggota Masing-masing Rumput pada Masing-masing Koloni Rumput

Sama seperti tahap penentuan keanggotaan dari rumput induk, fase penentuan anggota dari masing-masing rumput pada masing-masing koloni rumput juga terdiri dari dua tahap yaitu menentukan jarak masing-masing data dengan titik pusat yang ada pada masing-masing calon solusi kemudian menentukan keanggotaan masing-masing rumput pada masing-masing koloni rumput.

6. Fase Pemilihan Koloni Rumput yang Memiliki Nilai Fitness Terbaik

Fase pemilihan koloni rumput yang memiliki nilai fitness terbaik terdiri dari tahapan-tahapan sebagai berikut:

1. Menghitung nilai *fitness* total (SSE) dari masing-masing koloni rumput.
2. Memilih koloni rumput yang memiliki nilai SSE terbaik.

Koloni rumput dengan nilai *fitness* terbaik akan menjadi rumput induk untuk generasi berikutnya.

Secara umum, tahapan-tahapan didalam penerapan IWO untuk penentuan titik pusat klaster pada *K-means* adalah seperti terlihat pada Gambar 2.

Pengujian

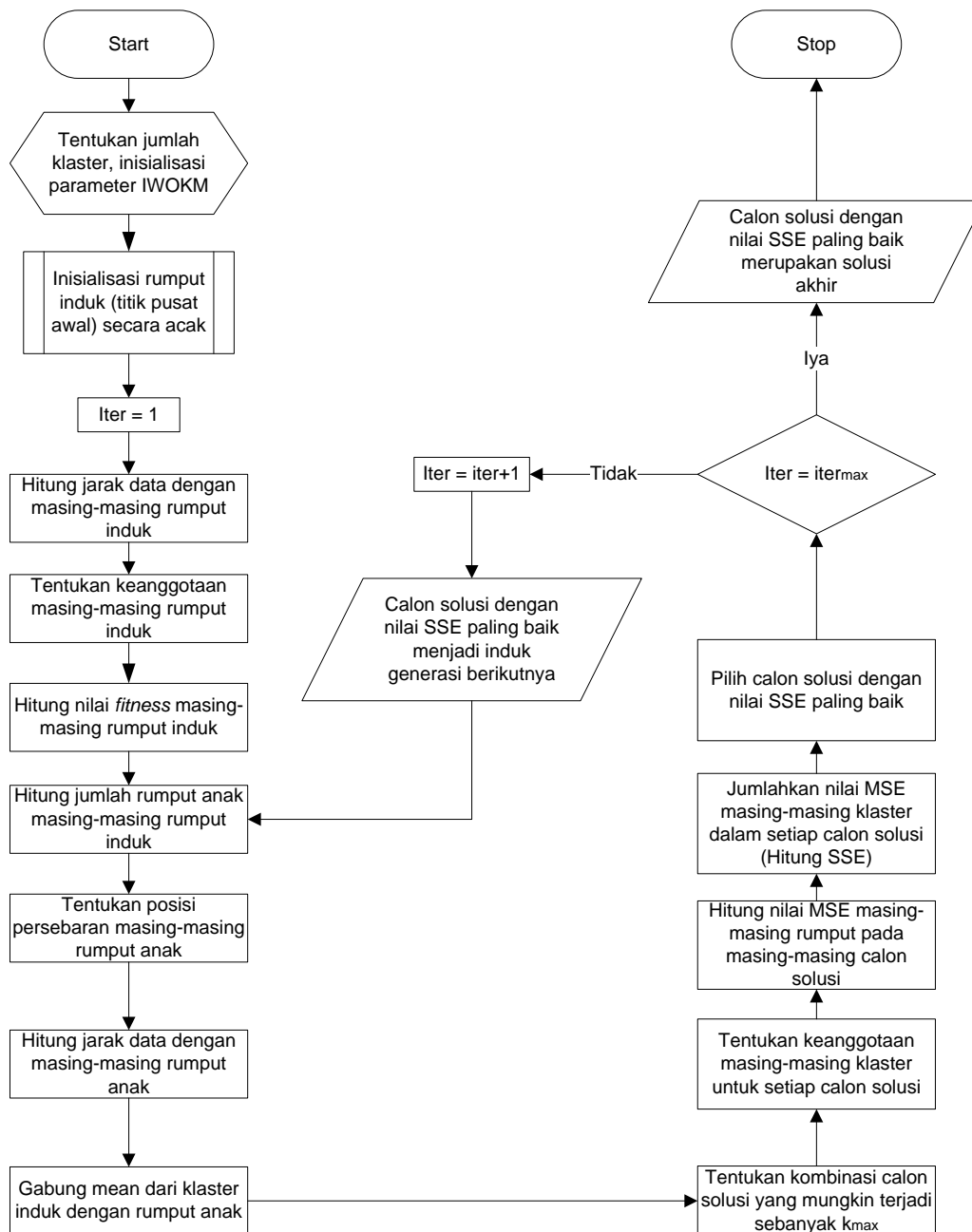
Sum Squared Error

Tujuan dari *K-means* adalah untuk meminimalkan hasil jumlah *squared error* dari seluruh k klaster [5]. Misalkan $X = \{x_i\}$, $i = 1, \dots, n$ merupakan kumpulan dari nd -*dimensional point* yang akan diklaster menjadi sebuah set k klaster, $C = \{c_k, k = 1, \dots, k\}$. Algoritma *K-means* nantinya akan membentuk sebuah partisi agar *squared error* antara *empirical mean* dari sebuah klaster dan *point* dalam klaster bernilai minimal. Katakan μ_k merupakan *mean* dari klaster c_k . Definisi dari *squared error* antara μ_k dan *point* dalam klaster c_k adalah seperti terlihat pada persamaan (4).

$$J(c_k) = \sum_{x_i \in c_k} (x_i - \mu_k)^2 \quad (4)$$

Persamaan (5) menunjukkan tujuan dari *K-means*, yaitu untuk meminimalkan hasil jumlah *squared error* dari seluruh k kluster.

$$J(C) = \sum_{k=1}^K \sum_{x_i \in c_k} (x_i - \mu_k)^2 \quad (5)$$



Gambar 2 Tahapan IWOKM

F-measure

Selain pengujian terhadap fungsi tujuan, juga dilakukan uji *F-Measure*. *F-Measure* merupakan nilai yang didapatkan dari pengukuran *Precision* dan *Recall* antara kelas hasil kluster dengan kelas sebenarnya yang terdapat pada data masukan [9]. *Precision* dan *Recall* bisa didapatkan dengan rumus seperti terlihat pada persamaan (6) dan (7).

$$p(i, j) = \frac{n_{ij}}{n_j} \quad (6)$$

$$r(i, j) = \frac{n_{ij}}{n_i} \quad (7)$$

Dimana $p(i, j)$ merupakan nilai *Precision* dari kelas i kluster ke j dan $r(i, j)$ merupakan nilai *Recall* dari kelas i kluster ke j , kemudian n_i adalah jumlah data pada kelas i yang diharapkan sebagai hasil *query*, n_j adalah jumlah data pada kluster j yang dihasilkan oleh *query*, dan n_{ij} adalah jumlah elemen dari kelas i yang masuk di kluster j .

Sedangkan rumus untuk menghitung nilai *F-Measure* kelas i kluster ke j adalah seperti terlihat pada persamaan (8).

$$F(i, j) = \frac{2 \cdot (p(i, j) \cdot r(i, j))}{p(i, j) + r(i, j)} \quad (8)$$

Untuk mendapatkan nilai *F-measure* dari data set dengan jumlah data n , maka rumus yang digunakan adalah seperti terlihat pada persamaan (9).

$$F = \sum_i \frac{n_i}{n} \max_j \{F(i, j)\} \quad (9)$$

Semakin besar nilai *F-measure*, semakin mampu algoritma menemukan kembali kelas label dari masing-masing *record* data [9].

3. HASIL DAN PEMBAHASAN

Pengujian dilakukan pada dataset bunga Iris. Dataset Iris terdiri dari 150 *instance* yang terdistribusi pada 3 kelompok, yaitu Setosa, Versicolor dan Virginica. Ada empat fitur penyusun data ini, yaitu sepal length, sepal width, petal length, petal width, semuanya dalam Cm. Skenario pengujian yaitu dengan mencoba 10 nilai titik pusat awal yang berbeda untuk 10 nilai k yang berbeda. Setelah itu dihitung nilai rata-rata SSE untuk masing-masing nilai k baik itu yang diperoleh oleh *K-means* maupun *IWOKM*. Parameter *IWOKM* yang digunakan dalam pengujian adalah seperti terlihat pada Tabel 2.

Hasil pengujian seperti terlihat pada Tabel 3 menunjukkan dari 100 percobaan yang dilakukan, metode *IWOKM* lebih sering menghasilkan nilai SSE yang lebih baik dari *K-means* yaitu sebanyak 69 kali. Sementara 31 percobaan lainnya baik metode *IWOKM* maupun *K-means* menghasilkan nilai SSE yang sama. Hal tersebut menunjukkan *IWOKM* memiliki kemampuan yang lebih baik dari *K-means* didalam menemukan titik pusat kluster yang lebih baik. Jika dilihat nilai rata-rata SSE dari 100 percobaan yang dilakukan seperti terdapat pada Tabel 4, rata-rata SSE hasil kluster dari *IWOKM* untuk masing-masing nilai k adalah selalu lebih baik jika dibandingkan dengan nilai rata-rata SSE *K-means*, dimana selisih rata-rata nilai SSE untuk 100 percobaan yang dilakukan adalah 4.5503. Untuk grafik perbandingan nilai rata-rata SSE dari *K-means* dan *IWOKM* untuk beberapa nilai k dapat dilihat pada Gambar 3.

Tabel 2 Nilai-nilai parameter *IWOKM* yang digunakan dalam pengujian

Parameter	Keterangan	Nilai
k	Jumlah kluster	6 – 15
k_{\max}	Jumlah koloni rumput	10
s_{\min}	Jumlah maksimum rumput anak	5
s_{\max}	Jumlah minimum rumput anak	50
σ_{initial}	Standar deviasi awal	3
σ_{final}	Standar deviasi akhir	0.2
iter_{\max}	Jumlah generasi	30
n	Nonlinear modulation index	3

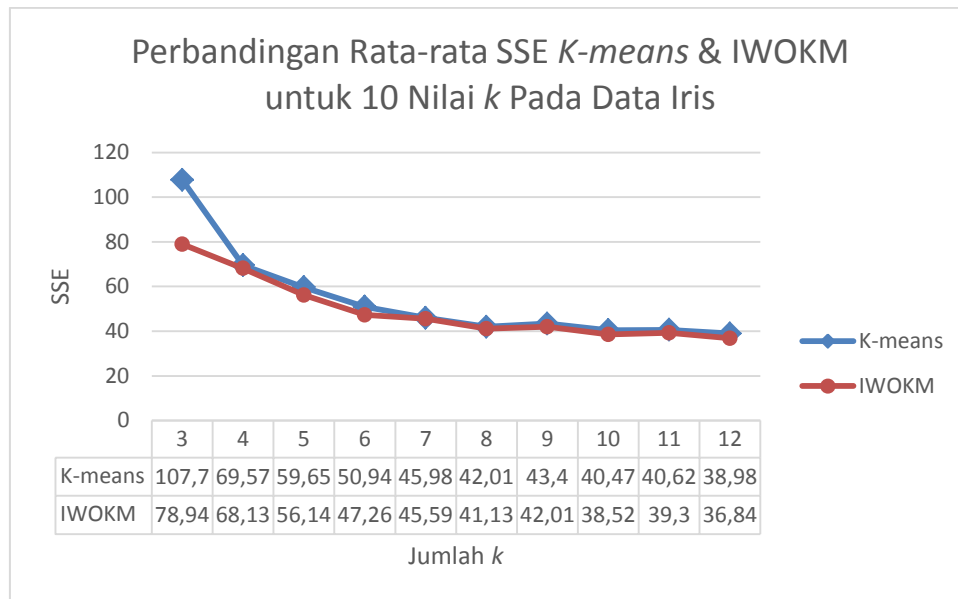
Untuk waktu proses pembentukan kluster, seperti terlihat pada Tabel 4, metode *IWOKM* memerlukan waktu yang sedikit lebih lama jika dibandingkan dengan algoritma *K-means*. Hal tersebut disebabkan karena dalam setiap iterasinya algoritma *IWOKM* melakukan penelusuran calon solusi lebih banyak jika dibandingkan dengan *K-means*. Selisih rata-rata waktu proses pembentukan kluster antara algoritma *K-means* dan *IWOKM* adalah 3 detik 533. Grafik perbandingan pengaruh perubahan nilai k terhadap waktu proses pembentukan kluster dapat dilihat pada Gambar 4.

Tabel 3 Perbandingan frekuensi penemuan titik pusat kluster yang lebih baik pada data Iris

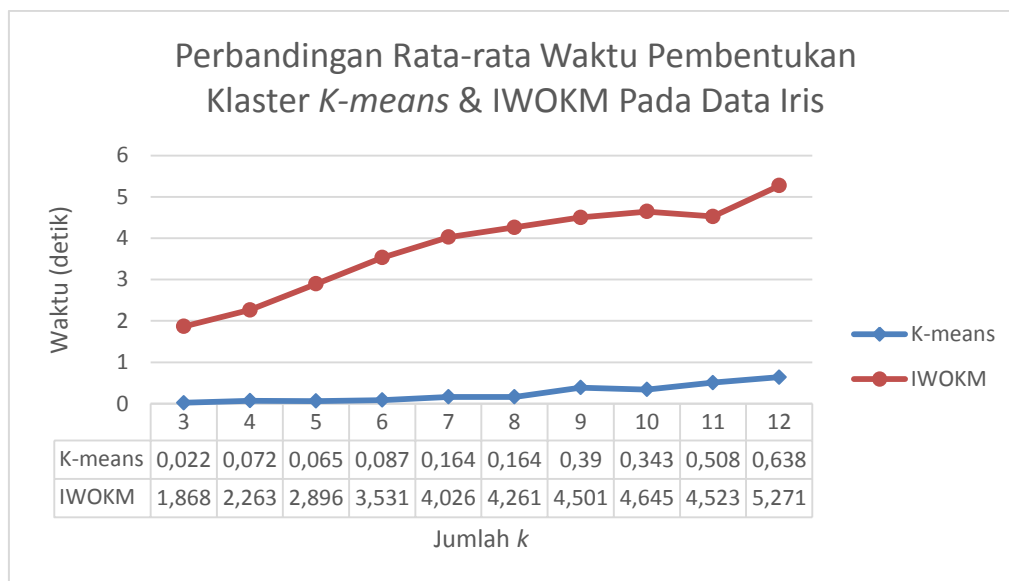
k	Nilai SSE				
	Lebih Baik		Sama	Terbaik	
	K-means	IWOKM		K-means	IWOKM
3	-	9	1	1	10
4	-	10	-	-	5
5	-	5	5	1	2
6	-	8	2	-	1
7	-	6	4	-	1
8	-	9	1	1	1
9	-	7	3	-	1
10	-	6	4	1	1
11	-	5	5	1	1
12	-	4	6	-	1
Total	0	69	31	5	24

Tabel 4 Perbandingan Hasil K-means dan IWOKM untuk Beberapa nilai k

k	SSE		Waktu	
	K-means	IWOKM	K-means	IWOKM
3	107.737	78.94084	0.022	1.868
4	69.57163	68.12936	0.072	2.263
5	59.64794	56.13528	0.065	2.896
6	50.94094	47.25963	0.087	3.531
7	45.98417	45.59263	0.164	4.026
8	42.0063	41.12745	0.164	4.261
9	43.39852	42.00716	0.390	4.501
10	40.46991	38.51888	0.343	4.645
11	40.61501	39.296	0.508	4.523
12	38.97918	36.8405	0.638	5.271
Rata-rata	53.9351	49.3848	0.245	3.778



Gambar 3 Perbandingan rata-rata nilai SSE *K-means* & *IWOKM* untuk beberapa nilai *k* pada dataset bunga Iris



Gambar 4 Perbandingan waktu proses pembentukan kluster algoritma *K-means* & *IWOKM* pada dataset bunga Iris

Untuk mengukur kemampuan algoritma didalam menemukan kembali kelas dari data dilakukan pengujian terhadap nilai *F-measure*. Hasil pengujian nilai *F-measure* untuk data Iris yang memiliki 3 kelas asli adalah seperti terlihat pada Tabel 5.

Pada Tabel 5 dapat dilihat dari sepuluh kali percobaan yang dilakukan, nilai *F-measure* yang didapat oleh algoritma *IWOKM* adalah selalu sama. Dari sepuluh percobaan yang dilakukan, algoritma *IWOKM* mampu mendapatkan nilai *F-measure* yang lebih baik sebanyak 9 kali, sementara 1 percobaan lainnya baik *K-means* maupun *IWOKM* menghasilkan nilai *F-measure* yang sama. Rata-rata selisih nilai *F-measure* dari 10 percobaan tersebut adalah 0.058709. Jika dilihat jumlah data yang tidak dikelompokkan dengan benar ke kelas aslinya seperti terlihat pada Tabel 6, dari 10 percobaan yang dilakukan, rata-rata kesalahan pengelompokan yang dilakukan *K-means* sebanyak 26 data sementara *IWOKM* 12 data. Nilai a, b

dan c pada Tabel 6 secara berturut-turut adalah menunjukkan nama kelas yang ada pada dataset bunga Iris yaitu Iris Setosa, Iris Virginica dan Iris Versicolor.

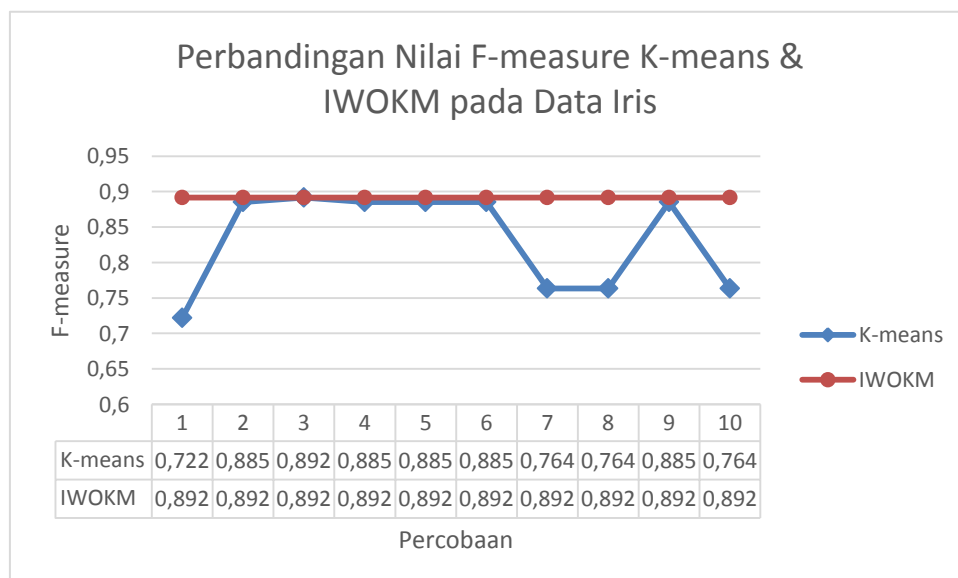
Tabel 5 Perbandingan nilai *F-measure* dari *K-means* dan *IWOKM* pada dataset bunga Iris

Perc	<i>F-measure</i>	
	K-means	IWOKM
1	0.721881	0.891775
2	0.885279	0.891775
3	0.891775	0.891775
4	0.885279	0.891775
5	0.885279	0.891775
6	0.885279	0.891775
7	0.763534	0.891775
8	0.763534	0.891775
9	0.885279	0.891775
10	0.763534	0.891775
Rata-rata	0.833065	0.891774

Grafik perbandingan nilai *F-measure* untuk *K-means* dan *IWOKM* untuk dataset bunga Iris dapat dilihat pada Gambar 5.

Tabel 6 Kesalahan pengelompokan hasil kluster dataset bunga Iris

Perc	K-means				IWOKM			
	a	b	c	total	a	b	c	total
1	-	46	-	46	-	-	12	12
2	11	-	-	11	12	-	-	12
3	-	12	-	12	-	12	-	12
4	-	11	-	11	-	12	-	12
5	11	-	-	11	12	-	-	12
6	-	-	11	11	-	-	12	12
7	3	-	47	50	-	12	-	12
8	-	3	47	50	-	12	-	12
9	11	-	-	11	12	-	-	12
10	3	-	47	50	12	-	-	12
Rata-rata				26.3				12



Gambar 5 Perbandingan rata-rata nilai *F-measure* *K-means* & *IWOKM* pada dataset bunga Iris

4. KESIMPULAN

Berdasarkan implementasi dan pengujian yang telah dilakukan, dapat ditarik beberapa kesimpulan sebagai berikut :

1. Algoritma *Invasive Weed Optimization* dapat diterapkan untuk penentuan titik pusat kluster pada *K-means*.
2. Tingkat keberhasilan *IWOKM* didalam menemukan titik pusat yang lebih baik dari *K-means* ketika dicobakan pada data bunga Iris adalah sebesar 69%, dimana rata-rata selisih nilai SSE dari 100 percobaan yang dilakukan adalah 4.55.
3. Algoritma *IWOKM* memerlukan waktu proses pembentukan kluster yang sedikit lebih lama jika dibandingkan dengan algoritma *K-means*. Untuk 100 percobaan yang dilakukan, rata-rata selisih waktu proses pembentukan kluster antara algoritma *K-means* dan *IWOKM* ketika dicobakan pada data bunga Iris adalah 3 detik 533 milidetik.
4. Akurasi hasil pengelompokan data bunga Iris dengan *K-means* adalah sebesar 82.47%, sementara *IWOKM* sebesar 92%, dimana selisih nilai *F-measure*-nya adalah 0.06.

5. SARAN

Metode yang diusulkan dalam penelitian ini masih memiliki kekurangan dalam hal waktu proses pembentukan kluster. Hal ini disebabkan oleh besarnya kemungkinan calon solusi yang ditelusuri oleh metode usulan dalam setiap iterasinya. Pada penelitian selanjutnya sekiranya dapat dilakukan perbaikan didalam cara pemilihan rumput yang akan menjadi rumput induk pada setiap generasi *IWOKM*, sehingga diharapkan dapat mempercepat waktu proses pembentukan kluster dari metode *IWOKM*.

DAFTAR PUSTAKA

- [1] Shen, H., 2010, Hybridization of Particle Swarm Optimization with The K-Means Algorithm for Clustering Analysis, *IEEE Fifth International Conference on Bio-Inspired Computing: Theories and Applications (BIC-TA)*, pp : 531–535.
- [2] Arai, K. and Ridho, A., 2007, Hierarchical K-means: an Algorithm for Centroids Initialization for K-means, *Reports of the faculty of science and engineering*, Saga University, Vol. 36, No. 1, pp : 25–31.
- [3] Redmond, S. J. and Heneghan, C., 2007, A method for initialising the K-means clustering algorithm using kd-trees, *Pattern Recognition Letters* 28, pp : 965-973
- [4] Steinley, D., 2003, Local Optima in K-Means Clustering: What You Don't Know May Hurt You, *Psychological Methods* 2003, Vol. 8, No. 3, pp : 294-304.
- [5] Jain, A. K., 2010, Data Clustering: 50 Years Beyond K-means, *Pattern Recognition Letters*, Vol. 31 Issue 8, pp : 651-666.
- [6] Cui, X. and Potok, T. E., 2005, Document Clustering using Particle Swarm Optimization, *Proceedings 2005 IEEE Swarm Intelligence Symposium (SIS 2005)*, pp : 191-197.
- [7] Mehrabian, A. R. and Lucas, C., 2006, A Novel Numerical Optimization Algorithm Inspired from Weed Colonization, *Ecological Informatics*, Vol. 1, pp : 355-366.

-
- [8] Zhang, X., Wang, Y., Cui, G., Niu, Y. and Xu, J.,2009, Application of a novel IWO to the design of encoding sequences for DNA computing, *Computers and Mathematics with Applications* 57, pp : 2001–2008
- [9] Yang, F., Sun, T. and Zhang, C., 2009, An efficient hybrid data clustering method based on K-harmonic means and Particle Swarm Optimization, *Expert Systems with Application* 36, pp : 9847-9852.