

ID 1090 JTIK

by 1090 Jtiik

Submission date: 02-Nov-2018 02:14PM (UTC+0700)

Submission ID: 1031536898

File name: 1090-3299-1-SM.docx (73.63K)

Word count: 3434

Character count: 21992

PENINGKATAN PERFORMA CLUSTER FUZZY C-MEANS PADA KLASTERING SENTIMEN MENGGUNAKAN PARTICLE SWARM OTIMIZATION

Rimbun Siringoringo¹, Jamaluddin Jamaluddin²

^{1,2} Manajemen Informatika, Universitas Methodist Indonesia

Email: ¹rimbun.ringo@gmail.com, ²jac.satuno@gmail.com

Abstrak

Fuzzy C-Means (FCM) merupakan algoritma klastering yang sangat baik dan lebih fleksibel dari algoritma klustering konvensional. Selain kelebihan tersebut, kelemahan utama algoritma ini adalah sensitif terhadap pusat klaster. Pusat klaster yang sensitif mengakibatkan hasil akhir sulit di kontrol dan FCM mudah terjebak pada optimum lokal. Untuk mengatasi masalah tersebut, penelitian ini memperbaiki kinerja FCM dengan menerapkan Particle Swarm Optimization (PSO) untuk menentukan pusat klaster yang lebih baik. Penelitian ini diterapkan pada klastering sentimen dengan menggunakan data berdimensi tinggi yaitu ulasan produk yang dikumpulkan dari beberapa situs toko *online* di Indonesia. Hasil penelitian menunjukkan bahwa penerapan PSO pada pembangkitan pusat klaster FCM dapat memperbaiki performa FCM serta memberikan luaran yang lebih sesuai. Performa klastering yang menjadi acuan adalah *Rand Index*, *F-Measure* dan *Objective Function Value* (OFV). Untuk keseluruhan performa tersebut, FCM-PSO memberikan hasil yang lebih baik dari FCM. Nilai OFV yang lebih baik menunjukkan bahwa FCM-PSO tersebut membutuhkan waktu konvergensi yang lebih cepat serta penanganan *noise* yang lebih baik.

2
Kata kunci: *fuzzy c means, particle swarm optimization, klastering sentimen, ulasan produk*

AN IMPROVED FUZZY C-MEANS FOR SENTIMENT CLUSTERING BASED ON PARTICLE SWARM OPTIMIZATION

Abstract

23
Fuzzy C-Means (FCM) is one of the best-known clustering algorithm and more flexible than conventional clustering algorithms. However, FCM is significantly sensitive to the initial cluster center and easily trapped in a local optimum. To overcome this problem, this study proposes an improved FCM with Particle Swarm Optimization (PSO) algorithm to determine a better cluster center for high dimensional and unstructured sentiment clustering. This study uses product review data collected from several online shopping websites in Indonesia. Initial processing product review data consists of Case Folding, Non Alpha Numeric Removal, Stop Word Removal, and Stemming. PSO is applied for the determination of suite cluster center. Clustering performance criteria are Rand Index, F-Measure and Objective Function Value (OFV). The results showed that FCM-PSO can provide better performance compared to the conventional FCM in terms of Rand Index, F-measure and Objective Function Values (OFV). The better OFV value indicates that FCM-PSO requires faster convergence time and better noise handling

Keywords : *fuzzy c means, particle swarm optimization, sentiment clustering, product review*

1. PENDAHULUAN

Pendekatan *data mining* dalam analisis sentimen menjadi pendekatan yang paling populer saat ini untuk menggali opini publik (Amolik, *et al*, 2016). Beberapa penelitian terkait penerapan *data mining* pada sentimen analisis adalah penerapan *Naive Bayes Classifier* (Fang & Zhang, 2015) pada analisis sentimen terhadap produk buku dan elektronik berdasarkan ulasan produk yang dikumpulkan dari situs jual beli *Amazon.com*. Penerapan metode *K-*

Means Clustering (KMC) dan *Decision Tree* pada analisis sentimen untuk mengidentifikasi opini konsumen terhadap produk baru (Soni & Mathai, 2016). Metode *Support Vector Machine* (SVM) diterapkan pada analisis sentimen terhadap *film box-office* berdasarkan ulasan produk pada situs *imdb.com* (Nagamma, *et al*, 2015).

Algoritma klastering bertujuan untuk mengelompokkan data, dimana distribusi dari data tersebut belum diketahui dan untuk menemukan struktur yang tersembunyi pada data. Metode klastering

yang konvensional termasuk metode yang kaku dan sensitif. Pada masalah pertama, sebuah objek data hanya dapat dikelompokkan dengan satu kluster saja secara eksklusif, hubungan objek data dengan kluster lain tidak ada sama sekali, oleh karenanya disebut metode klustering yang tegas atau *hard clustering* (Zhang and Shen, 2014). Pada klustering kabur atau *fuzzy clustering*, seperti *Fuzzy C-Means* (FCM), sifat kekakuan ini dapat dihilangkan, sebuah objek data dapat memiliki hubungan dengan banyak kluster berdasarkan nilai keanggotaan data tersebut. Dalam banyak situasi, FCM menghasilkan kinerja dan luaran yang lebih baik dari metode klustering konvensional (Cebeci and Yildiz, 2015).

Masalah sensitifitas terhadap pusat kluster awal adalah kelemahan utama FCM (Ma et al., 2015). Sensitifitas terhadap pusat kluster dapat menghasilkan iterasi proses yang sangat rumit dan dapat mengakibatkan proses klustering terjebak pada kondisi optimum lokal (Ye and Jin, 2016a). Kondisi ini tentu saja menjadi lebih buruk jika FCM bekerja pada data yang berdimensi tinggi seperti ulasan produk. Sensitifitas tersebut mengakibatkan hasil klasterisasi sulit dikontrol, hasil akhir sangat tergantung pada inisialisasi centroid (Patil et al., 2014).

Pendekatan yang paling populer untuk menangani pusat kluster awal adalah cara random yaitu membangkitkan bilangan acak (Gayatri and Vasanthi, 2017). Pendekatan tersebut membutuhkan waktu yang lama dalam proses pencarian solusi dan luaran sulit dikontrol, dimana dengan input data yang sama, luaran selalu berubah dari satu pengujian ke pengujian selanjutnya (Kumar et al., 2015).

Beberapa penelitian terbaru berhasil membuktikan bahwa pendekatan teknik optimalisasi seperti *Genetic Algorithm* (GA) (Wikaisuksakul, 2014, Ye and Jin, 2016b), *Ant Colony Optimization* (ACO) (Raghtate and Salankar, 2015) berhasil mengatasi kelemahan FCM. Hasil penelitian menunjukkan bahwa dengan menerapkan metode-metode tersebut, diperoleh algoritma yang lebih efisien dan performa *klustering* yang lebih baik serta peningkatan stabilitas dan akurasi klustering.

Untuk mencari solusi terbaik terhadap masalah sensitifitas pusat kluster, penelitian ini difokuskan pada penerapan *Particle Swarm Optimization* (PSO). Ada dua alasan mendasar pada penerapan PSO. Pertama, PSO lebih sederhana dari pada GA serta mudah diterapkan karena PSO tidak memiliki banyak prosedur seperti seleksi, mutasi maupun pindah silang atau *crossover* (Sun and Xu, 2017). Kedua, PSO berhasil dikombinasikan dengan metode-metode *machine learning* seperti *Support Vector Machine* (SVM) pada pengujian usia tulang menggunakan dataset *bone age data*. Hasil yang dicapai adalah sebuah model SVM-PSO yang lebih efektif (Güraksin et al., 2014). PSO berhasil mendesain jumlah layer, jumlah neuron dan jumlah

bias pada Jaringan Syaraf Tiruan (Garro and Vázquez, 2015), kombinasi PSO dengan SOM (*Self Organizing Map*) telah menghasilkan metode baru yaitu SOSwarm (O'Neill and Brabazon, 2008). Kombinasi PSO pada metode-metode *machine learning* tersebut menghasilkan perbaikan kinerja yang signifikan.

Pada penelitian ini PSO (*Particle Swarm Optimization*) diterapkan pada inisialisasi pusat kluster FCM. Penelitian ini membangun sebuah model klasterisasi FCM-PSO untuk mengevaluasi sekaligus menganalisis sentimen masyarakat terhadap produk toko *online*. Pengoptimalan pusat kluster FCM dengan PSO diharapkan meningkatkan performa FCM pada tugas klasterisasi (klustering).

2. Fuzzy C-Means Clustering Algorithm

FCM merupakan salah satu algoritma klustering yang mengadaptasi konsep himpunan *fuzzy*. FCM merupakan algoritma *fuzzy clustering* yang paling banyak digunakan. FCM adalah suatu teknik pengelompokan data dimana keberadaan atau letak suatu data dalam kluster tergantung pada derajat keanggotaan data tersebut. Nilai derajat keanggotaan berada pada rentang 0 dan 1. Semakin tinggi nilai derajat keanggotaan maka semakin besar kemiripan antara suatu data dengan kelompok yang ada (Ye and Jin, 2016a). FCM memungkinkan setiap objek data dimiliki oleh lebih dari satu kelompok atau kluster tergantung pada nilai keanggotaan data tersebut. Jika diasumsikan bahwa kita hendak mengelompokkan sebanyak n buah objek data ke dalam c grup atau kluster, maka fungsi objektif yang dapat diterapkan pada FCM menggunakan persamaan (1) (Ma et al., 2015).

$$(U, V) = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m (d_{ik})^2 \quad (1)$$

$U = [\mu_{ik}]$ adalah matrik derajat keanggotaan dengan dimensi $c \times n$, μ_{ik} adalah derajat keanggotaan antara suatu data k ke kelompok i . Nilai derajat keanggotaan berada pada rentang 0 dan 1. Semakin tinggi nilai μ_{ik} maka semakin besar kepemilikan data k terhadap kelompok i . $d_{ik} = \|x_k - v_i\|$ adalah merupakan jarak *euclidean* antara k dengan pusat kluster i , m adalah indeks *fuzzy* yang memiliki nilai pada rentang $[1, \infty]$.

Algoritma FCM dapat diterapkan dengan langkah-langkah sebagai berikut (Ye and Jin, 2016a).

Langkah 1 : Inisialisasi parameter c , m dan T (jumlah iterasi maksimum). Inisialisasi μ_{ik} secara acak ($i=1, \dots, c$ dan $k=1, \dots, n$) untuk objek k ke kelompok i .

Langkah 2 : Tentukan pusat kluster dengan persamaan (2) berikut :

$$v_i = \frac{\sum_{k=1}^n (\mu_{ik})^m x_k}{\sum_{k=1}^n (\mu_{ik})^m} \quad (2)$$

Langkah 3 : Tentukan perubahan matriks fungsi keanggotaan dengan persamaan (3).

$$u_{ik} = \frac{1}{\sum_{k=1}^n \frac{d_{ik}^2}{d_{jk}^{m-1}}} \quad (3)$$

Langkah 4 : Hitung nilai obyektif $J(U,V)$ menggunakan persamaan (2) dan perhatikan kriteria berhenti.

Langkah 5, Cek kondisi berhenti . Jika $|J_t - J_{t+1}| < \text{error}$ terkecil (ϵ) atau ($t > T$) maka berhenti, Jika tidak $t = t + 1$; kembali ke langkah 2.

22

3. Particle Swarm Optimization

Particle Swarm Optimization (PSO) adalah teknik optimasi stokastik berbasis populasi. PSO dapat digunakan untuk pencarian solusi yang optimal di dalam ruang pencarian yang luas. Dalam proses pencarian lokasi optimal, partikel-partikel bergerak dengan kecepatan tertentu dan senantiasa mengubah posisi sampai akhirnya menemukan lokasi optimal. Suatu partikel bergerak di ruang pencarian n -dimensi, vektor posisi partikel adalah $X_i = (x_{1i}, \dots, x_{ni})$ dan vektor kecepatan $V_i = (v_{1i}, \dots, v_{ni})$. Untuk menemukan lokasi optimumnya, partikel senantiasa memperbarui posisi dan kecepatan menggunakan Persamaan (4) digunakan untuk menghitung kecepatan baru tiap partikel dan persamaan (5) digunakan untuk memperbarui posisi tiap partikel pada ruang solusi (Olusanya et al., 2015).

$$V_i(t+1) = \omega V_i(t) + c_1 r_1 (P_i - X_i) + c_2 r_2 (P_g - X_i) \quad (4)$$

$$X_i(t+1) = X_i(t) + V_i(t+1) \quad (5)$$

P_i adalah posisi terbaik pada tiap partikel dan P_g adalah posisi terbaik *swarm*, r_1 dan r_2 adalah bilangan acak dengan interval $[0,1]$, c_1 dan c_2 adalah faktor pembelajaran atau *learning factor* yang secara berturut turut merepresentasikan sebuah komponen kognitif dan sebuah komponen sosial. Parameter t merupakan indeks iterasi, ω merupakan parameter bobot inersia digunakan untuk menyeimbangkan antara kemampuan pencarian global dan pencarian lokal (Zhang et al., 2015). Biasanya, bobot inersia yang bagus adalah kurang sedikit dari satu.

4. Pengukuran Performa

Rand Index merupakan kriteria yang paling umum untuk mengukur kinerja klustering. *Rand Index* (RI) menyatakan persentase keputusan yang tepat yang dihasilkan oleh algoritma klustering serta seberapa mirip hasil klustering dengan data yang sebenarnya RI dapat ditentukan menggunakan persamaan (6). *F-measure* merupakan ukuran yang mengkombinasikan nilai *Precision* (P) dan nilai *Recall* (R). Nilai *Recall* dan *Precision* pada persamaan (7) dan (8) terkadang dapat memiliki bobot atau nilai yang berbeda. Ukuran yang menampilkan timbal balik antara R dan P adalah *F-measure* yang merupakan bobot *harmonic mean* pada *recall* dan *precision*. *F-measure* dapat ditentukan menggunakan persamaan (9)

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

$$\text{Precision } (P) = \frac{TP}{TP+FP} \quad (7)$$

$$\text{Rand Index } (RI) = \frac{TP+TN}{TP+TN+FP+FN} \quad (8)$$

$$F - \text{Measure} = \frac{2PR}{P+R} \quad (9)$$

Pada persamaan (6), *True positif* (TP) adalah keputusan menempatkan dua data yang identik ke kelompok yang sama, *True Negative* (TN) adalah keputusan menempatkan dua data yang tidak identik ke kelompok berbeda. Ada dua jenis kesalahan yang dapat terjadi pada proses klustering yaitu *False positif* (FP) dan *False negative* (FN). FP adalah keputusan menempatkan dua data yang tidak identik ke kelompok yang sama. FN adalah keputusan menempatkan dua data yang identik ke kelompok yang berbeda.

24

5. Term Frequency - Inverse Document Frequency (TF-IDF)

Term Frequency (TF) merupakan banyaknya kata atau *token* tertentu yang terdapat dalam suatu dokumen. *Inverse Document Frequency* (IDF) adalah frekuensi kemunculan kata atau *term* pada keseluruhan dokumen (Khoirudin et al., 2018). TF-IDF dapat ditentukan menggunakan persamaan (10).

$$W_{dt} = tf_{dt} * idf_t = tf_{dt} * \log \left(\frac{N}{df_t} \right) \quad (10)$$

W_{dt} merupakan bobot *term* ke- t terhadap dokumen d , tf_d merupakan jumlah kemunculan *term* t dalam dokumen- d , N adalah jumlah dokumen secara keseluruhan dan df_t adalah jumlah dokumen yang mengandung *term* t . Pada penelitian ini, TF digunakan untuk menentukan banyaknya kata atau *term* pada ulasan tertentu, N merupakan banyaknya ulasan produk keseluruhan, sementara IDF digunakan untuk menentukan frekuensi kemunculan kata atau *term* tertentu pada keseluruhan ulasan produk.

6. Fuzzy C Means berbasis PSO

Langkah 1 : Penentuan parameter PSO yang terdiri dari jumlah *swarm*, jumlah partikel, W_{up} dan W_{low} . Penentuan parameter FCM yang terdiri dari Iterasi Maksimum (T), *error* terkecil (ϵ), Jumlah klaster (c), bobot (w), Fungsi objektif awal (p_0), dan Iterasi awal (t). Pada tabel 1 dan tabel 2 dijabarkan parameter PSO dan FCM yang digunakan pada penelitian ini.

Tabel 1. Parameter PSO

No	Parameter	Nilai
1	Jumlah partikel	10
2	Jumlah iterasi/swarm	500
3	Velocity maksimum	0,01
4	W_{up}	2
5	W_{low}	1

Tabel 2. Parameter FCM

No	Parameter	Nilai
1	Iterasi maksimum	500
2	Error terkecil (ϵ)	10^{-6}
3	Jumlah kluster (c)	2
4	Pangkat (w)	2
5	Fungsi objektif awal (p_o)	0
6	Iterasi awal (t)	1

Langkah 2 : Inisialisasi pusat kluster. Dalam penerapan FCM, pusat kluster harus diinisialisasi terlebih dahulu. Inisialisasi pusat kluster ditampilkan pada gambar 1.

1	2	3		c
C_1	C_2	C_3		C_c

Gambar 1. Skema Inisialisasi pusat kluster

Langkah 3 : Bangkitkan Partikel awal secara random. Pusat kluster awal dibangkitkan secara paralel sebanyak n partikel. Skema pembangkitan partikel disajikan pada gambar 2.

P_1	1	2	3		c
	C_{11}	C_{12}	C_{13}		C_{1c}
P_2	1	2	3		c
	C_{21}	C_{22}	C_{23}		C_{2c}
P_n	1	2	3		c
	C_{n1}	C_{n2}	C_{n3}		C_{nc}

Gambar 2. Skema pembangkitan partikel

Langkah 4 : Setiap partikel dievaluasi berdasarkan nilai *fitness*. Fungsi *fitness* (*fitness function*) merupakan fungsi yang digunakan untuk menghitung *fitness* atau tingkat kebaikan suatu partikel untuk bertahan hidup. Fungsi ini mengambil parameter berupa individu dan menghasilkan keluaran nilai *fitness* dari individu tersebut. Untuk setiap permasalahan yang akan diselesaikan dengan algoritma PSO harus terdefinisi suatu *fitness function*. Pada penelitian ini, fungsi *fitness* ditentukan menggunakan persamaan (11) dan persamaan (12)

$$F(P) = \frac{1}{J_{FCM}} \quad (11)$$

$$J_{FCM} = \sum_{k=1}^n \sum_{i=1}^c (\mu_{ik})^m (d_{ik})^2 \quad (12)$$

Langkah 5 : Update posisi dan kecepatan partikel. Prosedur PSO mengharuskan posisi dan kecepatan setiap partikel dievaluasi berdasarkan persamaan (4) dan (5).

Langkah 6 : Pengujian model FCM untuk menemukan partikel terbaik. Partikel terbaik yang dievaluasi berdasarkan nilai *fitness* pada langkah 4 diterapkan untuk melatih data latih kembali dan dihitung nilai *fitness* nya

29

Langkah 7 : Perbaharui *personal optimal fitness* (*pfit*) dan *personal optimal position* (*pbest*). Sejauh ini semua partikel pada iterasi awal (pertama) telah dievaluasi, iterasi pertama menghasilkan sebuah partikel terbaiknya (*pbest*).

Langkah 8 : Pada tahap ini dievaluasi apakah seluruh iterasi telah selesai dikerjakan. Jika belum akan dilanjutkan pada **langkah 3**, jika telah selesai maka dilanjutkan pada **langkah 10**.

Langkah 9 : Setiap iterasi memiliki partikel terbaik. Pada tahap ini partikel terbaik dari semua iterasi akan dievaluasi untuk menentukan partikel terbaik *global optimal position* (*gbest*).

Langkah 10 : Jika nilai *gbest* pada **langkah 9** merupakan nilai yang diharapkan maka pada saat tersebut telah diperoleh inisial pusat kluster yang optimal. Jika nilai *gbest* belum sesuai dengan kriteria yang diharapkan maka kembali dibangkitkan populasi baru dengan kembali pada **langkah 3**.

7. Pengujian dan analisis komparatif

28

Penelitian ini menggunakan data ulasan produk yang dikumpulkan dari beberapa situs jual beli online di Indonesia. Data ulasan produk terdiri dari 1073 data dengan 627 sentimen positif dan 445 sentimen negatif. Gambar 3 berikut ini adalah sampel ulasan produk yang digunakan

1. Sepatunya bagus. tetapi ukuran nya kecil dari ukuran biasanya. Jd mesti pesan ukuran lebih besar dari ukuran biasanya
2. kiriman cepat sampai,BARANG SESUAI
3. Anak saya pas dipesan tidak mau...tapi ketika barang nya datang, dia suka banget.
4. beda sama yang digambar..!
5. warna kurang sesuai dengan gambar
6. pengirimannya lama banget ᄇᄇᄇᄇ

Gambar 3. Sampel ulasan produk

7.1 Pemrosesan awal data

Setelah data ulasan produk berhasil di kumpulkan, maka tahap selanjutnya adalah tahap pemrosesan awal (*pre-processing*) agar data ulasan produk dapat diterapkan pada algoritma klustering. Tahapan *pre-processing* yang diterapkan adalah *Case Folding*, *Non Alpha Numeric Removal*, *Stop words Removal*, dan *Stemming*. Daftar *stop words* untuk Bahasa Indonesia terdiri atas 760 kata (Tala, 2003). **27** algoritma *Stemming* yang diterapkan adalah algoritma *stemming* khusus untuk bahasa Indonesia yaitu algoritma *nazief*-Andriani (Adriani et al., 2007).

1. sepatu bagus ukur ukur mesti pesan ukur besar ukur
2. kirim cepat barang sesuai
3. anak pas pesan barang datang suka
4. beda gambar
5. warna kurang sesuai gambar
6. kirim lama
7. kecewa

Gambar 4. Hasil pemrosesan awal

Untuk menentukan bobot setiap fitur pada data ulasan diterapkan algoritma TF-IDF. Hasil penerapan TF-IDF menghasilkan matriks data dengan dimensi 86 atribut x 1073 data. Dimensi data di atas masing sangat besar dan tidak efektif, sehingga atribut yang ada dievaluasi dan difilter. Seleksi atribut dilakukan dengan algoritma *Correlation-based Feature Selection* (CFS) yang tersedia pada WEKA. Seleksi atribut menghasilkan sebanyak 18 fitur. Berikut ini adalah top 5 fitur berdasarkan bobot TF-IDF.

Tabel 3. Top 5 fitur data ulasan produk sepatu

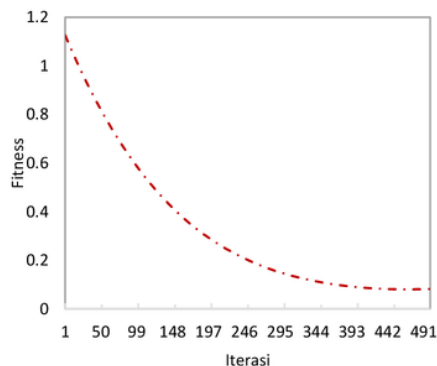
No	Fitur	TF-IDF
1	Bagus	0,520
2	Sesuai	1,030

Tabel 4. Pengkodean partikel

	Matriks partikel	Fitness
Particle best	1.84, 4.67, 0.41, 1.31, 1.07, 4.80, 2.91, 4.80, 3.78, 0.97, 4.47, 4.66, 0.08, 2.79, 3.86, 0.67, 3.68, 3.40	1,125
(pbest)	4.22, 1.97, 2.04, 1.21, 4.85, 4.83, 1.01, 0.61, 4.25, 0.33, 3.12, 4.52, 2.58, 1.87, 3.54, 2.58, 3.04, 4.72	
Global best	0.01, 0.00, 0.00, 4.93, 0.00, 4.27, 0.00, 2.31, 0.00, 2.26, 0.00, 4.78, 4.77, 0.64, 0.00, 0.10, 4.91, 3.55	0,065
(gbest)	0.00, 3.44, 0.00, 2.69, 0.00, 0.00, 0.00, 0.00, 0.02, 4.15, 0.90, 0.00, 0.00, 2.62, 4.71, 4.69, 0.67, 4.61	

7.3 Nilai fitness

Setiap partikel dievaluasi melalui nilai *fitness*. Proses penelusuran nilai fitness terbaik ditampilkan pada grafik. Nilai *fitness* yang terkecil menggambarkan pusat kluster yang menghasilkan *error* terkecil.

Gambar 5. Penelusuran *gbest*

Pekerjaan klastering menggunakan FCM membutuhkan inisial pusat kluster yang baik. Metode pembangkitan pusat kluster secara random digantikan dengan PSO untuk membangkitkan pusat kluster. Dengan percobaan sebanyak 10 kali, maka pada tabel 5 ditampilkan indeks rata-rata efektifitas metode klastering yang diuji pada penelitian ini, dimana performa terbaik ditandai dengan huruf tebal.

3	Terimakasih	1,236
4	Suka	1,464
5	Cepat	1,588

7.2 Pengkodean PSO

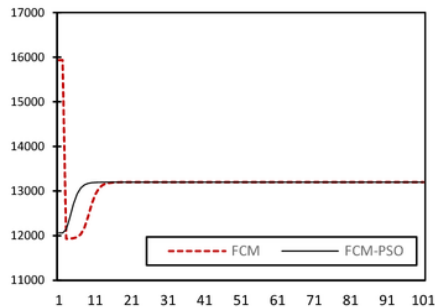
Setiap partikel pada PSO mewakili pusat kluster FCM. Partikel dikodekan dengan bilangan *real* dengan rentang 0 sampai 5. Untuk jumlah kluster = 2, setiap partikel terdiri dari matriks dua dimensi. Pada tabel 2 ditampilkan bentuk pengkodean partikel serta nilai *fitness* yang dihasilkan. Partikel terbaik pada iterasi pertama (*pbest*) memiliki nilai fitness 1, 125, sedangkan partikel terbaik secara global (*gbest*) memiliki nilai *fitness* 0,065

Tabel 5. Perbandingan Performa FCM vs FCM-PSO

	FCM-PSO	FCM
TP	609	221
TN	272	430
FP	18	406
FN	174	16
Recal (R)	0.778	0.932
Precision (P)	0.971	0.352
Rand Index (RI)	0.821	0.607
F-Measure	0.864	0.512
OFV	13070.68	13136.61

Pada tabel 26 ditunjukkan bahwa perbaikan metode FCM-PSO menunjukkan kinerja yang lebih baik dibandingkan dengan metode sebelumnya. FCM-PSO memiliki nilai RI yang lebih baik dari FCM. Hal ini mengindikasikan bahwa pembangkitan pusat kluster FCM dengan PSO meningkatkan kemampuan FCM dalam mengenal pola sentimen yang tersembunyi di dalam data. Hal tersebut juga menjadi indikator bahwa FCM-PSO dapat bekerja dengan baik pada data berdimensi tinggi dan tidak terstruktur seperti data ulasan produk. Model FCM-PSO juga memiliki *F-Measure* yang lebih baik, hal tersebut menunjukkan bahwa FCM-PSO lebih efektif. Dari sisi OFV, metode FCM-PSO memiliki nilai yang lebih baik. Nilai OFV yang besar menunjukkan bahwa metode tersebut membutuhkan waktu yang lama untuk mencapai konvergen. Pada gambar 6 ditunjukkan bahwa FCM memiliki waktu yang lebih lama untuk mencapai konvergen. Dengan nilai epsilon 10^{-6} dan iterasi 100, FCM-PSO konvergen pada iterasi ke 26, sementara FCM konvergen pada

iterasi ke 32. FCM-PSO memiliki harmonisa yang lebih baik daripada FCM sehingga FCM-PSO lebih baik dalam menangani *noise* pada data.



Gambar 6. Grafik waktu konvergensi

8. Kesimpulan

FCM merupakan algoritma klastering yang sangat baik, tetapi algoritma ini sensitif terhadap pusat klaster. Penelitian ini menguji efektifitas penerapan *Particle Swarm Optimization* pada optimalisasi hasil klastering sentimen pada FCM. *Particle Swarm Optimization* diterapkan untuk menentukan pusat klaster yang lebih baik. Penelitian ini diterapkan pada data berdimensi tinggi yaitu ulasan produk toko *online*. Hasil penelitian menunjukkan bahwa PSO dapat memperbaiki kinerja FCM atau dengan kata lain Model FCM-PSO lebih baik jika dibandingkan dengan model FCM saja. Hal tersebut terbukti melalui perbaikan performa klastering setelah menerapkan PSO. Penelitian ini menunjukkan bahwa FCM-PSO sangat cocok digunakan untuk tugas klastering dalam rangka menemukan pola pada data berdimensi tinggi.

7 DAFTAR PUSTAKA

- ADRIANI, M., ASIAN, J., NAZIEF, B., TAHAGHOGHI, S.M., WILLIAMS, H.E., 2007. Stemming Indonesian: A confix-stripping approach. *ACM Trans. Asian Lang. Inf. Process. TALIP* 6, 1–33.
- CEBECI, Z., YILDIZ, F., 2015. Comparison of K-means and Fuzzy C-means algorithms on different cluster structures. *Agrár. Agric. Inform.* 6, 13–23.
- GARRO, B.A., VÁZQUEZ, R.A., 2015. Designing artificial neural networks using particle swarm optimization algorithms. *Comput. Intell. Neurosci.* 2015, 61.
- GAYAH, I., D., VASANTHI, K., 2017. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology* 2, 4.
- GÜRAKSIN, G.E., HAKLI, H., UĞUZ, H., 2014. Support vector machines classification based on particle swarm optimization for

- bone age determination. *Appl. Soft Comput.* 24, 597–602.
- KHOIRUDIN, M.M., WIRANTO, WINARNO, 2018. News Opinion Mining around Universitas Sebelas Maret Using Naive Bayes Algorithm 7, 7.
- KUMAR, B.V., KARPAGAM, G.R., REKHA, N.V., 2015. Performance analysis of deterministic centroid initialization method for partitional algorithms in image block clustering. *Indian J. Sci. Technol.* 8, 63–73.
- MA, L., LI, Y., FAN, S., FAN, R., 2015. A Hybrid Method for Image Segmentation Based on Artificial Fish Swarm Algorithm and Fuzzy-Means Clustering. *Comput. Math. Methods Med.* 2015.
- OLUSANYA, M.O., ARASOMWAN, M.A., ADEWUMI, A.O., 2015. Particle swarm optimization algorithm for optimizing assignment of blood in blood banking system. *Comput. Math. Methods Med.* 2015.
- O'NEILL, M., BRABAZON, A., 2008. Self-organizing swarm (SOSwarm) for financial credit-risk assessment. *IEEE*, pp. 3087–3093.
- PATIL, A.J., PATIL, C.S., KARHE, R.R., AHER, M.A., 2014. Comparative study of different clustering algorithms. *Int J Adv Res Electr Electron Instrum Eng* 3, 10490–10497.
- RAGHT, E., G.S., SALANKAR, S.S., 2015. Modified fuzzy C means with optimized ant colony algorithm for image segmentation, in: *Computational Intelligence and Communication Networks (CICN), 2015 International Conference On. IEEE*, pp. 1283–1288.
- SUN, T., XU, M., 2017. A swarm optimization genetic algorithm based on quantum-behaved particle swarm optimization. *Comput. Intell. Neurosci.* 2017.
- TALA, F.Z., 2003. A study of stemming effects on information retrieval in Bahasa Indonesia. *Inst. Log. Lang. Comput. Univ. Van Amst. Neth.*
- WIKAISUKSAKUL, S., 2014. A multi-objective genetic algorithm with fuzzy c-means for automatic data clustering. *Appl. Soft Comput.* 24, 679–691.
- YE, A.-X., JIN, Y.-X., 2016a. A fuzzy c-means clustering algorithm based on improved quantum genetic algorithm. *simulation* 9.
- YE, A.-X., JIN, Y.-X., 2016b. A fuzzy c-means clustering algorithm based on improved quantum genetic algorithm. *simulation* 9.
- Zhang, J., Shen, L., 2014. An improved fuzzy c-means clustering algorithm based on shadowed sets and PSO. *Comput. Intell. Neurosci.* 2014, 22.

12

Zhang, Y., Wang, S., Ji, G., 2015. A comprehensive survey on particle swarm optimization algorithm and its applications. Math. Probl. Eng. 2015.

ORIGINALITY REPORT

15%

SIMILARITY INDEX

14%

INTERNET SOURCES

11%

PUBLICATIONS

%

STUDENT PAPERS

PRIMARY SOURCES

1

repository.usu.ac.id

Internet Source

2%

2

Lilin Jie, Weidong Liu, Zheng Sun, Shasha Teng. "Hybrid fuzzy clustering methods based on improved self-adaptive cellular genetic algorithm and optimal-selection-based fuzzy c-means", Neurocomputing, 2017

Publication

1%

3

jtiik.ub.ac.id

Internet Source

1%

4

Dey, Nilanjan, Soumyo Bose, Achintya Das, Sheli Sinha Chaudhuri, Luca Saba, Shoaib Shafique, Andrew Nicolaides, and Jasjit S. Suri. "Effect of Watermarking on Diagnostic Preservation of Atherosclerotic Ultrasound Video in Stroke Telemedicine", Journal of Medical Systems, 2016.

Publication

1%

5

"Population Size Reduction in Particle Swarm Optimization Using Product Graphs", Advances

1%

in Intelligent Systems and Computing, 2015.

Publication

6	www.springerprofessional.de Internet Source	1 %
7	publikasiilmiah.ums.ac.id Internet Source	1 %
8	journal.uii.ac.id Internet Source	1 %
9	M. M. Gowthul Alam, S. Baulkani. "Geometric structure information based multi-objective function to increase fuzzy clustering performance with artificial and real-life data", Soft Computing, 2018 Publication	1 %
10	Hayat Al-Dmour, Ahmed Al-Ani. "Chapter 21 MR Brain Tissue Segmentation Based on Clustering Techniques and Neural Network", Springer Nature, 2017 Publication	1 %
11	ladangbelajar.blogspot.com Internet Source	1 %
12	link.springer.com Internet Source	1 %
13	B. Srinivas, G. Sasibhusana Rao. "Unsupervised learning algorithms for MRI	1 %

brain tumor segmentation", 2018 Conference on Signal Processing And Communication Engineering Systems (SPACES), 2018

Publication

14

Kumar, B. Vinoth, and G.R. Karpagam. "Knowledge-based differential evolution approach to quantisation table generation for the JPEG baseline algorithm", International Journal of Advanced Intelligence Paradigms, 2016.

Publication

1%

15

Youqun Zhao, Han Xu, Yaoji Deng, Qiuwei Wang. "Multi-objective optimization for ride comfort of hydro-pneumatic suspension vehicles with mechanical elastic wheel", Proceedings of the Institution of Mechanical Engineers, Part D: Journal of Automobile Engineering, 2018

Publication

1%

16

www.mdpi.com

Internet Source

<1%

17

kuscholarworks.ku.edu

Internet Source

<1%

18

Melva Hermayanty Saragih, Abba Suganda Girsang. "Sentiment analysis of customer engagement on social media in transport online", 2017 International Conference on

<1%

Sustainable Information Engineering and Technology (SIET), 2017

Publication

19	www.smurfitschool.ie Internet Source	<1 %
20	m.scirp.org Internet Source	<1 %
21	dblp.dagstuhl.de Internet Source	<1 %
22	www.researchgate.net Internet Source	<1 %
23	Staiano, A.. "Improving RBF networks performance in regression tasks by means of a supervised fuzzy clustering", Neurocomputing, 200608 Publication	<1 %
24	www1.ics.uci.edu Internet Source	<1 %
25	www.science.gov Internet Source	<1 %
26	repository.its.ac.id Internet Source	<1 %
27	lintaka.com Internet Source	<1 %

Jamaluddin, Rimbun Siringoringo. "Improved Fuzzy K-Nearest Neighbor Using Modified Particle Swarm Optimization", Journal of Physics: Conference Series, 2017

Publication

Exclude quotes Off

Exclude matches Off

Exclude bibliography Off