

Implementasi *Particle Swarm Optimization* pada *K-Means* untuk *Clustering Data Automatic Dependent Surveillance-Broadcast*

Achmad Saiful¹, Joko Lianto Buliali²

Departemen Informatika, Institut Teknologi Sepuluh Nopember Surabaya
Jl. Teknik Kimia - Gedung Informatika ITS Surabaya, Telp. 031-5939212
e-mail: ¹achmads123@gmail.com, ²joko@cs.its.ac.id

Abstrak

Investigasi kecelakaan penerbangan di Indonesia pada tahun 2010 sampai 2016 sebesar 212 investigasi. Hal tersebut dapat dihindari apabila ada sistem penerbangan yang dapat memastikan penerbangan berjalan aman, seperti sistem lalu lintas udara yang dapat mendeteksi apabila pesawat bergerak menuju ke arah yang salah. Penelitian ini bertujuan untuk mengelompokkan rute penerbangan pada data *Automatic Dependent Surveillance-Broadcast* menggunakan metode clustering untuk mendapatkan similaritas rute penerbangan. Penulis mengusulkan metode *particle swarm optimization* untuk mengoptimalkan metode *k-means*, yang berguna untuk menentukan titik centroid awal dengan *silhouette coefficient* sebagai *fitness function*. Hasil dari penelitian ini menghasilkan zona terbang berdasarkan kebiasaan sehingga dapat digunakan sebagai panduan penerbangan. Pengujian dilakukan dengan membandingkan nilai *Davies-Bouldin index* dengan metode *k-means*, *k-medoids* dan *fuzzy c-means*. Pada uji coba yang dilakukan, metode yang diusulkan menjadi kelompok metode terbaik pada lima dari enam segmen yang ada serta menghasilkan nilai *Davies-Bouldin index* lebih baik pada satu segmen sebesar 0,779.

Kata kunci: *Clustering, K-Means, Particle Swarm Optimization, Davies-Bouldin Index.*

Abstract

There are 212 investigations on aviation accident from 2010 to 2016 in Indonesia. The accidents can be avoided by providing a flight system that can ensure safety, such as air traffic systems that able to detect the movement of the plane in the wrong direction. This research clusters the data from the *Automatic Dependent Surveillance-Broadcast (ADS-B)* using clustering method to get the similarities of the flight route and proposes *particle swarm optimization* to determine the initial cluster in order to get the optimum of *k-means* and *k-medoids*. Output of this research produce flight zones based on habits that can be used as flight guides. The test is done by comparing *davies-bouldin index* values with *k-means*, *k-medoids* and *fuzzy c-means* method. Based on the experiments, the proposed method becomes the best group method on five of six existing segments and obtains better *davies-bouldin index* values (0,779) on one segment.

Keywords: *Clustering, K-Means, Particle Swarm Optimization, Davies-Bouldin Index.*

1. Pendahuluan

Berdasarkan investigasi yang dilakukan oleh Komite Nasional Keselamatan Kerja (KNKT), kecelakaan penerbangan dari tahun 2010 hingga 2016 sebesar 212 investigasi. Hal tersebut seharusnya tidak terjadi dikarenakan setiap tahapan yang dilakukan pada dunia penerbangan dilaporkan dengan baik. Untuk menghindari hal tersebut diperlukan suatu sistem penerbangan yang dapat mengurangi kecelakaan, seperti sistem yang dapat mendeteksi apabila pesawat bergerak menuju ke arah yang salah.

Tidak banyak penelitian pada sistem lalu lintas udara [1] terlebih fokus pada rute penerbangan atau kondisi ketika pesawat berada di udara dengan kata lain setelah *take-off* dan sebelum *landing*. Penelitian terkait rute penerbangan adalah karya James DeArmon, et all [2] yang bertujuan untuk mengurangi banyaknya rute penerbangan dengan cara mengelompokkan rute yang sama antar bandara dengan mengusulkan penggunaan *algoritma edit distance*, menghitung jarak antar rute berdasarkan perbedaan urutan sektor yang ditempuh oleh pesawat dalam satu kali penerbangan. Penelitian lainnya yaitu karya Omnia Osama, et all [3] yaitu penelitian tentang pengelompokan objek bergerak menggunakan *k-*

means dengan studi kasus pada penerbangan dan karya Pusadan, et all [4] meneliti tentang penentuan *waypoint* yang optimal berdasarkan kebiasaan data yang akan digunakan sebagai deteksi anomali suatu penerbangan.

Berdasarkan data dan penelitian di atas, penelitian ini melakukan pengelompokan rute penerbangan berdasarkan kebiasaan data dari penerbangan yang ada. Pengelompokan dilakukan dengan menggunakan metode *clustering* berbasis jarak yaitu *k-means clustering*. *K-means clustering* adalah suatu algoritma pengelompokan data berdasarkan kedekatan jarak dengan titik utamanya atau *centroid* sehingga menghasilkan beberapa kelompok data tertentu [5]. Kelemahan dari *k-means clustering* adalah penetapan titik *centroid* awal ditentukan secara acak. Untuk mengatasi kelemahan tersebut, digunakan *particle swarm optimization* [6] dengan *fitness function silhouette coefficient* untuk menentukan titik *centroid* awal. Hasil dari pengelompokan tersebut berupa zona terbang berdasarkan kebiasaan data yang ada sehingga dapat menjadi panduan penerbangan.

Data yang digunakan adalah rute Surabaya-Palu dengan kode penerbangan LNI860. Implementasi metode dilakukan pada tiap segmen yang terbagi berdasarkan *waypoint*, Evaluasi *cluster* berdasarkan pengukuran internal *validity cluster* menggunakan Davies-Bouldin index. Pengujian dilakukan dengan 10 variasi dari 70% data yang dipilih secara acak, setiap variasi akan dilakukan percobaan sebanyak 50 kali percobaan dan dibandingkan dengan metode *k-means*, *k-medoids* serta *fuzzy c-means*.

2. Metodologi Penelitian

2.1. Pre-Processing

Penelitian ini dimulai dengan memilih data *latitude* dan *longitude* dengan ketinggian yang tetap. Data yang digunakan adalah data penerbangan rute Surabaya-Palu pada bulan September, November dan Desember sebanyak 90 penerbangan. Pada Tabel 1 merupakan sampel data nomor 80-85 pada satu penerbangan pada ketinggian 33000, jumlah data dan ketinggian tetap pada tiap penerbangan berbeda-beda.

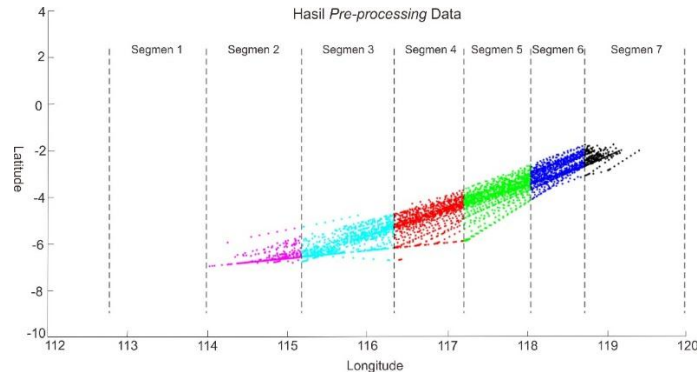
Tabel 1. Sampel data satu penerbangan pada ketinggian 33000.

1	<i>Timestamp, UTC, Callsign, Position, Altitude, Speed, Direction</i>
80	1481275496, 2016-12-09T09:24:56Z, LNI860, "-1.968554,118.71978", 33000, 447,35
81	1481275433, 2016-12-09T09:23:53Z, LNI860, "-2.074182,118.64505", 33000, 447,35
82	1481275344, 2016-12-09T09:22:24Z, LNI860, "-2.221473,118.541153", 33000, 448,35
83	1481275270, 2016-12-09T09:21:10Z, LNI860, "-2.349072,118.450943", 33000, 448,35
84	1481275076, 2016-12-09T09:17:56Z, LNI860, "-2.67424,118.221367", 33000, 444,35
85	1481275015, 2016-12-09T09:16:55Z, LNI860, "-2.778656,118.147591", 33000, 445,35

Kemudian data dibagi menjadi beberapa segmen sesuai dengan *waypoint*. Pada penerbangan Surabaya – Palu terdapat 8 *waypoint* (dapat dilihat pada Tabel 2) sehingga terbentuk 7 segmen. segmen yang digunakan hanya segmen 2 sampai 7 dikarenakan pada segmen 1 pesawat masih mengalami perubahan ketinggian sehingga tidak ada data pada segmen tersebut. Hasil dari *pre-processing* dapat dilihat pada Gambar 1.

Tabel 2. *Waypoint* Surabaya-Palu.

Waypoint	Latitude	Longitude
Surabaya	-7.37333	112.77167
Fando	-6.97333	113.98500
Kasol	-6.56833	115.17333
Dasty	-6.17333	116.33000
Endog	-5.87667	117.20167
Gamal	-2.86333	118.03833
Rudal	-2.66167	118.71167
Palu	-0.88500	119.96167



Gambar 1. Visualisasi data awal.

2.2. Implementasi *Particle Swarm Optimization* pada *K-means*

Implementasi *particle swarm optimization* [7] pada *k-means* berfungsi untuk menentukan titik centroid awal dari *k-means* yang sebelumnya ditentukan secara acak. *Fitness function* yang digunakan berdasarkan nilai *silhouette coefficient* [8].

Tahapan pertama yang dilakukan adalah menentukan titik *centroid* secara acak serta mengelompokkan data pada titik *centroid* terdekat. kemudian dalam setiap iterasinya, dihitung *fitness function* menggunakan dari *cluster* yang terbentuk tersebut menggunakan *silhouette coefficient* berikut:

$$b(i) = \min_d \frac{\sum_{j=1, j \neq i}^{N_{c_{not\ i}}} d(i, j)}{N_{c_i}} \quad (1)$$

$$a(i) = \mu_i = \frac{\sum_{j=1, j \neq i}^{N_{c_i}} d(i, j)}{N_{c_i}} \quad (2)$$

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}}, -1 \leq s(i) \leq 1 \quad (3)$$

Keterangan dari rumus tersebut adalah:

$d(i, j)$ = jarak *eulclidean distance* antara titik i dengan j
 N_{c_i} = jumlah data pada cluster i

Kemudian hitung *local best* menggunakan persamaan (4) dan hitung *global best* menggunakan persamaan (5).

$$localbest(c) = \max_{i=1,2,\dots,n_c} s(i) \quad (4)$$

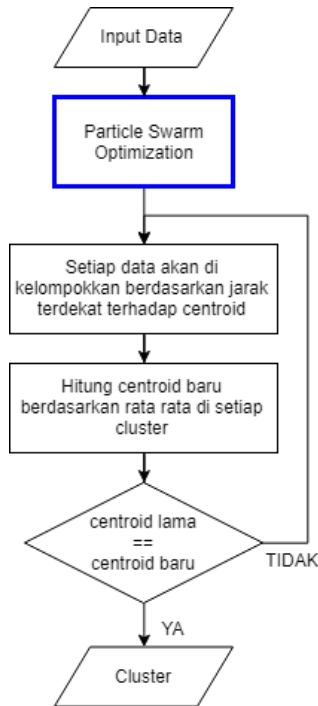
$$globalbest = \max_{i=1,2,\dots,c} localbest(i) \quad (5)$$

Setelah itu hitung kecepatan v menggunakan persamaan (6) dan hitung partikel x berdasarkan kecepatan menggunakan (7). Iterasi berhenti apabila jumlah iterasi sudah melalui batas maksimal iterasi, iterasi yang digunakan penulis sebesar 15 iterasi.

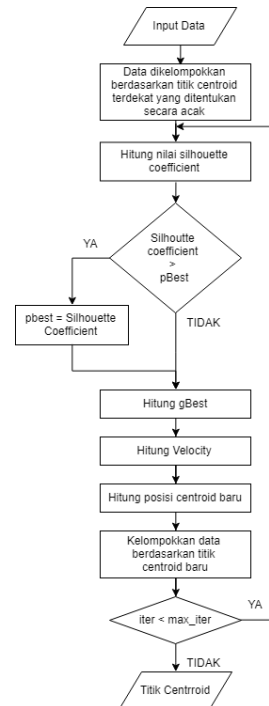
$$v_i^k = v_i^k + c_1 r_1 (localbest_i^k - x_i^k) + c_2 r_2 (globalbest_i^k), c_1 = 0.005, \quad c_2 = 0.005 \quad (6)$$

$$x_i^k = x_i^k + v_i^k \quad (7)$$

Hasil dari *particle swarm optimization* ini berupa titik *centroid* yang akan digunakan sebagai *centroid* awal dari *k-means*. Lebih jelasnya dapat dilihat pada Gambar 2 dan Gambar 3, dimana Gambar 2 adalah posisi urutan *particle swarm optimization* pada *k-means* serta pada Gambar 3 adalah tahapan yang dilakukan *particle swarm optimization*.



Gambar 2. Flowchart tradisional K-means.



Gambar 3. Flowchart Particle Swarm Optimization.

2.3. Evaluasi

Setelah proses pengelompokan selesai dilakukan kemudian dihitung nilai kualitas cluster yang terbentuk menggunakan pengukuran internal validitas cluster Davies-Bouldin [9] berikut:

$$DB(c) = \frac{1}{L} \sum_{i=1}^L \max_{i \neq j} \left\{ \frac{\Delta(c_i) + \Delta(c_j)}{\delta(c_i, c_j)} \right\} \tag{8}$$

Dimana $\Delta(c_i)$ adalah similaritas antar cluster/jarak rata rata semua titik ke centroid. $\delta(c_i, c_j)$ adalah perbedaan antar cluster atau jarak dari centroid cluster i ke centroid cluster j .

3. Hasil dan Analisis

3.1. Pengujian Metode

Pengujian dilakukan sebanyak 500 kali dengan 10 variasi data set. Variasi data set yang dimaksud adalah 70% dari data set yang diacak secara random pada setiap 50 kali pengujian. Metode yang diusulkan dibandingkan dengan metode tradisional k-means, tradisional k-medoids dan fuzzy c-means.

Pengujian dilakukan pada segmen dua sampai segmen tujuh karena pada segmen satu tidak dalam posisi crusing (mengalami perubahan ketinggian). Dalam pengujian ini didapatkan hasil evaluasi cluster berdasarkan nilai Davies-Bouldin index.

3.2. Analisis Hasil Evaluasi Cluster

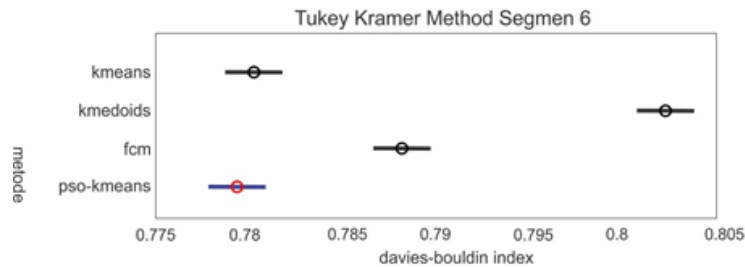
Analisis hasil evaluasi cluster dilakukan untuk mendapatkan metode yang terbaik di antara yang lain, metode yang digunakan untuk analisis adalah metode uji statistik one-way analysis of variance (one-way anova). Hipotesis awal dari one-way anova adalah ada perbedaan di antara metode yang digunakan. Contohnya pada segmen 6, hasil perhitungan anova pada Tabel 3 menghasilkan nilai p-value sebesar 2,51e-244, hal tersebut menunjukkan bahwa hipotesis bernilai benar, karena nilai yang dihasilkan di bawah threshold yaitu 0.05.

Tabel 3. Hasil Analysis of Variance pada segmen 6.

Source	SS	df	MS	F	p-value
Columns	0,3669	6	0,0611	225,87	2,51e-244

Error	0,9458	3493	0,0002
Total	1,3128	3499	

Setelah hipotesis awal memenuhi maka dilakukan uji statistik lanjut menggunakan *tukey-kramer method* dimana hasil dari metode tersebut berupa kelompok metode terbaik dan satu metode terbaik berdasarkan nilai *Davies-Bouldin index*. Hasil *tukey-kramer method* pada Gambar 4 menunjukkan bahwa pada segmen 6 terbentuk 3 kelompok metode terbaik yaitu metode *k-means* dan *pso-kmeans* sebagai kelompok metode terbaik, dengan metode yang diusulkan menjadi metode terbaik dengan nilai *Davies-Bouldin index* sebesar 0.779. Keseluruhan hasil analisis mulai dari segmen 2 sampai segmen 7 dapat dilihat pada Tabel 2.

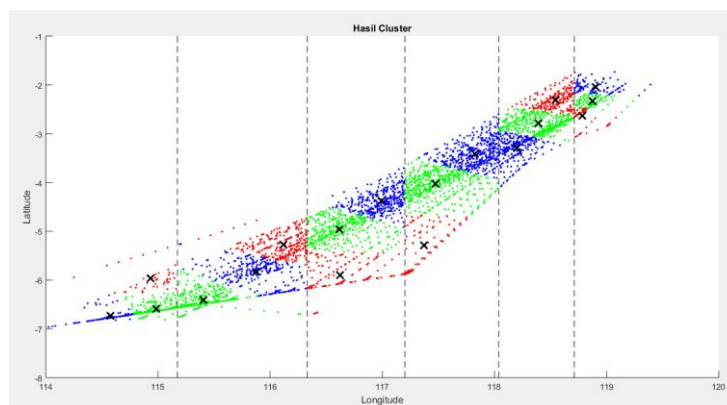


Gambar 4. Hasil *tukey-kramer method* pada segmen 6.

Hasil metode terbaik pada Tabel 4 dijadikan dasar untuk mengelompokkan rute penerbangan pada segmen 2 sampai segmen 7. Hasil *cluster* berupa zona terbang pada Gambar 5 yang membagi menjadi 3 zona yang dapat digunakan sebagai panduan penerbangan atau deteksi anomali. Pada Gambar 5 warna hijau menandakan zona tersebut paling banyak dilalui oleh pesawat, warna biru menandakan zona terbanyak kedua yang dilalui oleh pesawat dan yang terakhir yaitu warna merah, zona yang jarang dilalui oleh pesawat.

Tabel 4. Keseluruhan hasil analisis.

Segmen	Kelompok Metode Terbaik	Metode Terbaik	Davies-Bouldin index
Segmen 2	K-means Particle swarm optimization k-means Fuzzy c-means	K-means	0,668
Segmen 3	Fuzzy c-means	Fuzzy c-means	0,803
Segmen 4	Fuzzy c-means K-means Particle swarm optimization k-means	Fuzzy c-means	0,685
Segmen 5	Fuzzy c-means K-means Particle swarm optimization k-means	Fuzzy c-means	0,683
Segmen 6	Particle swarm optimization k-means K-means	Particle swarm optimization k-means	0,779
Segmen 7	K-means Particle swarm optimization k-means	K-means	0,850



Gambar 5. Zona terbang yang terbentuk berdasarkan metode clustering terbaik.

4. Kesimpulan

Berdasarkan pada uji coba dan analisis yang dilakukan, optimasi pada *k-means* menggunakan *particle swarm optimization* dengan *fitness function silhouette coefficient* berhasil dilakukan. Hal tersebut dapat dilihat pada pengelompokan pada *Automatic Dependent Surveillance-Broadcast (ADS-B)*, metode tersebut menjadi kelompok metode terbaik pada lima dari enam segmen yang ada serta menghasilkan nilai *Davies-Bouldin index* lebih baik pada satu segmen dibandingkan dengan metode *k-means*, *k-medoids* dan *fuzzy c-means* sebesar 0.779.

Daftar Pustaka

- [1] H. Wang and R. Wen, "Analysis of Air Traffic Network of China," in *24th Chinese Control and Decision Conference (CCDC)*, Taiyuan, China, 2012.
- [2] J. DeArmon, C. Taylor, T. Masek and C. Wanke, "Air Route Clustering for a Queuing Network Model of the," in *14th AIAA Aviation Technology, Integration, and Operations Conference*, Atlanta, GA, 2014.
- [3] O. Ossama, H. M. Mokhtar and M. E. El-Sharkawi, "An Extended K-means Technique for Clustering," *Egyptian Informatics Journal*, vol. 12, no. 1, pp. 45-51, 2011.
- [4] M. Y. Pusadan, J. L. Buliali and R. V. H. Ginardi, "Anomaly Detection of Flight Routes through Optimal Waypoint," *Journal of Physics: Conference Series*, vol. 801, no. 1, 2017.
- [5] P. Kaur, "Outlier Detection Using Kmeans and Fuzzy Min Max Neural Network in Network Data," in *International Conference on Computational Intelligence and Communication Networks (CICN)*, Tehri, India, 2016.
- [6] Tan, Long, "A Clustering K-means Algorithm Based on Imporved PSO Algorithm", in *International Conference on Communication Systems and Network Technologies*, Gwalior, India, 2015.
- [7] Dhote, C. A., Anuradha D. Thakare and Shruti M. Chaudhari, "Data Clustering using Particle Swarm Optimization and Bee Algorithm" in *International Conference on Computing, Communications and Networking Technologies (ICCCNT)*, Tiruchengode, India, 2013
- [8] Li, Ting, Peng Ye and Shikai Zheng, "State Grid Office System User Clustering Analysis based on K-means Algorithm", in *IEEE 3rd International Conference on Big Data Analysis (ICBDA)*, Shanghai, China, 2018.
- [9] Karo, Karo Ichwanul Muslim, Kiki Maulana Adhigraha, Arief Fatchul Huda, "A Cluster Validity for Spatial Clustering Based on Davies-Bouldin Index and Polygon Disiimilarity Function", *International Conference on Informatics and Computing (ICIC)*, Jayapura, Indonesia, 2017.