

BAB II

TINJAUAN PUSTAKA

2.1. Penelitian Sebelumnya

Dalam penelitian terdahulu, *Particle Swarm Optimization* terhadap *K-Means* telah banyak digunakan dalam pengelompokan suatu data. Berdasarkan penelitian sebelumnya telah menunjukkan bahwa pengelompokan suatu data sangat dibutuhkan, untuk mencegah *K-Means* agar tidak berada di titik lokal optimum dan meningkatkan kemampuan algoritma untuk menemukan *clustering* yang lebih baik. Berikut adalah penelitian terdahulu:

Image Segmentation Algorithm Based on Particle Swarm Optimization with K-Means Optimization [11]. Tujuan dari penelitian ini untuk mencapai gambar yang akurat, serta dapat memecahkan masalah pada saat pemilihan pust awal pengelompokan k-means dan agar tidak berada di titik lokal optimal. Metode yang diterapkan pada penelitian ini adalah metode *K-Means Clustering*, dan *Particle Swarm Optimization*. Hasil dari penelitian memperoleh segmentasi gambar secara akurat, dan tingkat efisiensi segmentasi lebih tinggi jika dibandingkan dengan algoritma *K-Means* tunggal, oleh karena itu penelitian ini mengkombinasikan *K-Means* dengan *Particle Swarm Optimization*.

Clustering Pengunjung Mall Menggunakan Metode K-Means dan Particle Swarm Optimization [12] Tujuan dari penelitian ini mencluster pengunjung mall. Metode yang diterapkan pada penelitian ini adalah metode *K-Means Clustering*, *Particle Swarm Optimization* dan *Davies Buildin Index*. Hasil dari penelitian ini adalah *clustering* menghasilkan 5 *cluster* dan evaluasinya menunjukkan bahwa 5 *cluster* yang telah dioptimalkan lebih optimal daripada yang belum mengalami optimasi meskipun algoritamanya memiliki kinerja yang lebih rendah dan analisis tersebut mendapatkan 5 karakteristik.

Integrasi Particle Swam Optimization Menggunakan K-Means untuk Klasterisasi Pengangguran di kota Pagar Alam [13]. Tujuan dari penelitian ini mengelompokkan secara cluster para pengangguran di kota Pagar Alam. Metode yang diterapkan pada penelitian ini menggunakan metode *K-Means Clustering*, *Particle Swarm Optimization* dan *Davies Buildin Index*. Hasil dari penelitian ini adalah *Davies Bouldin index* menggunakan *K-Means* biasa didapatkan nilai 0,60. Sedangkan *K-Means* yang sudah di optimasi menggunakan *Particle Swam Optimization* didapatkan hasil 0,59.

Optimasi *K-Means Clustering* Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma *Particle Swarm Optimization* Di Kota Semarang [14]. Tujuan dari penelitian ini mengklasifikasi daerah-daerah endemik dengan tujuan untuk mencapai indentifikasi penyakit menular yang akurat dalam pengelompokkan. Metode yang diterapkan pada penelitian ini menggunakan metode *K-Means Clustering*, *Particle Swarm Optimization* dan *Davies Buildin Index*. Hasil dari penelitian ini adalah algoritma dengan iterasi 10 dan K-Fold 2 menghasilkan *Davies Buildin Index* sebesar 0,619, sementara pada *K-Means* dengan *Particle Swarm Optimization* iterasi 10 dan K-Fold 5 menghasilkan *Davies Bouildin Index* sebesar 0,113.

A Novel Approach of Data Clustering Using An Improved Particle Swarm Optimization Based K–Means Clustering Algorithm [15]. Tujuan dari penelitian untuk mengelompokkan data multidimensi. Metode yang diterapkan pada penelitian ini menggunakan metode *K-Means Clustering*, dan *Particle Swarm Optimization*. Hasil pada penelitian ini adalah menghasilkan *clustering* yang baik dan memberikan kinerja yang lebih baik.

A Novel Effective Distance Measure and a Relevant Algorithm for Optimizing the Intial Cluster Centroid [16]. Tujuan dari penelitian untuk mengelompokkan data multidimensi. Metode yang diterapkan pada penelitian ini menggunakan metode *K-Means Clustering*, dan *Particle Swarm Optimization*. Hasil pada penelitian ini adalah algoritma *K-Means* yang diusulkan memiliki kinerja terbaik diantara enam algoritma, pengelompokan data ini mencapai

peringkat pertama dalam perbandingan peringkat rata-rata, lebih tinggi daripada skor rata-rata dari tiga algoritma pengelompokan lainnya. Dalam hal kecepatan, algoritma yang diusulkan menunjukkan signifikan. Secara keseluruhan, hasil menunjukkan konvergensi dan akurasi yang baik.

Antlion Optimizer Algorithm Modification for Initial Centroid Determination in K-means Algorithm [17]. Tujuan dari penelitian ini menentukan pusat *cluster* awal memberikan peningkatan pengelompokan dibandingkan dengan metode *k-means* dan *k-means++* biasa. Metode yang diterapkan pada penelitian ini menggunakan metode *K-Means Clustering*, dan *Particle Swarm Optimization*. Hasil pada penelitian ini adalah nilai SICD terbaik untuk setiap metode dan dataset diukur dengan memberikan peringkat yang baik pada metode yang diusulkan pada dataset Iris, Wine dan Cancers mendapatkan peringkat pertama pada dataset Ecoli dan Glass, metode yang diusulkan dan metode *K-Means ++*, keduanya mendapatkan peringkat rangking pertama, yang memberikan bukti bahwa dapat meningkatkan hasil *clustering* dan dapat menjadi metode alternatif untuk menentukan pusat *cluster* awal dengan menggunakan metode *K-Means*.

Peningkatan kinerja Algoritma K-Means dengan Menggunakan Particle Swarm Optimization dalam Pengelompokkan Data Penyediaan Akses Sanitasi dan Air Bersih [18]. Tujuan pada penelitian ini menentukan wilayah yang akan dijadikan referensi dalam pengelompokkan wilayah pada data PAMSIMAS. Metode yang diterapkan pada penelitian ini menggunakan metode *K-Means Clustering*, *Particle Swarm Optimization* dan *Davies Buildin Index*. Hasil dari penelitian ini adalah *K-Means* dengan inisialisasi centroid melalui metode acak dengan nilai 0,208856082, sementara metode *K-Means* dengan inisialisasi centroid melalui metode sesuai standar pemerintah menghasilkan 0,280077. Metode terbaik dalam penelitian ini yaitu *K-Means* yang menggabungkan *Particle Swarm Optimization* dengan nilai *Davies Bouildin Index* sebesar 0,08383.

Implementasi Particle Swarm Optimization pada K-Means untuk Clustering Data Automatic Dependent Surveillance Broadcast [19]. Tujuan dari penelitian ini menggabungkan jalur perjalanan pesawat dalam data *Automatic Dependent Surveillance Broadcast* agar mendapatkan kemiripan antar rute penerbangan. Metode yang diterapkan pada penelitian ini menggunakan metode *K-Means Clustering* dan *Particle Swarm Optimization*. Hasil dari penelitian ini adalah Menghasilkan nilai 0,779.

Customer segmentation using K-means clustering and the adaptive particle swarm optimization algorithm [20]. Tujuan dari penelitian ini adalah untuk mengatasi masalah ketidaksamaan dalam mengelompokkan campuran data. Metode yang diterapkan pada penelitian ini menggunakan metode *K-Means Clustering* dan *Particle Swarm Optimization*. Hasil dari penelitian ini adalah Menunjukkan keefektifan dan kepraktisan IKM-ALPSO untuk segmentasi pelanggan.

Tabel 2.1 Ringkasan Penelitian Terdahulu

No	Literatur Review		Latar Belakang Penelitian	Desain Riset dan Metodologi	
	Penulis, Tahun	Judul	Masalah Penelitian	Metode	Hasil
1.	Xuexin Chen, Pu Miao, Qingkai Bu , 2019	Image Segmentation Algorithm Based on ParticleSwarm Optimization with K-Means Optimization	Terdapat masalah dalam segmentasi gambar pada pemilihan pusat awal pengelompokan <i>K-Means</i> sehingga memperbaiki kerugiannya karena terjerumus ke dalam lokal optimal.	Metode <i>K-Means Clustering</i> , dan <i>Particle Swarm Optimization</i> .	Segmentasi gambar yang dihasilkan akurat, dan memiliki efisiensi segmentasi yang lebih tinggi daripada algoritma <i>K-Means</i> sendiri.
2.	Teuku Muhammad Dista, Ferian Fauzi Abdulloh	Clustering Pengunjung Mall Menggunakan Metode <i>K-Means</i> dan <i>Particle Swarm Optimization</i>	Terjadinya pendapatan mall yang semakin menurun sejak pandemi maka, dilakukan <i>clustering</i> pengunjung mall.	Metode <i>K-Means Clustering</i> , <i>Particle Swarm Optimization</i> dan <i>Davies Bouldin Index</i> .	<i>Clustering</i> yang dihasilkan 5 <i>cluster</i> .
3.	Deka Agustina, Yogi Isro' Mukti, Siti Muntari, 2023	Integrasi <i>Particle Swam Optimization</i> Menggunakan <i>K-Means</i> untuk Klasterisasi Pengangguran di Kota Pagar Alam	Tingginya angka pengangguran di kota Pagar Alam menunjukkan adanya ketidakseimbangan antara jumlah pencari kerja dan peluang pekerjaan yang tersedia.	Metode <i>K-Means</i> , <i>Particle Swarm Optimization</i> dan <i>Davies Bouldin Index</i> .	<i>Davies Bouldin index</i> menggunakan <i>K-Means</i> biasa = 0,60. Sedangkan <i>K-Means</i> menggunakan <i>Particle Swarm Optimization</i> = 0,59.

No	Literatur Review		Latar Belakang Penelitian		Desain Riset dan Metodologi
	Penulis, Tahun	Judul	Masalah Penelitian	Metode	Hasil
4.	Suhardi Rustam, Heru Agus Santoso, dan Catur Supriyanto, 2018	Optimasi <i>K-Means Clustering</i> Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma <i>Particle Swarm Optimization</i> Di kota Semarang	Masalah identifikasi ketidak akuratan dalam menangani penyakit yang menular.	Metode <i>K-Means</i> , <i>Particle Swarm Optimization</i> dan <i>Davies Buildin Index</i> .	<i>K-Means</i> saja = 0,619, <i>K-Means</i> dengan <i>Particle Swarm Optimization</i> = 0,113.
5.	Shouvik Paul and Sourav De, 2020	A Novel Approach of Data Clustering Using An Improved Particle Swarm Optimization Baset K-Means Clustering Algorithm	Terdapat pemilihan cluster yang buruk di <i>K-Means</i> pada tahap awal dapat mempengaruhi hasil <i>clustering</i> dan terjebak pada minimum lokal .	Metode <i>K-Means Clustering</i> , dan <i>Particle Swarm Optimization</i> .	Menghasilkan <i>clustering</i> yang baik dan memberikan kinerja yang lebih baik.
6.	Yang Liu, Shuaifeng Ma, and Xinxin Du, 2020	A Novel Effective Distance Measure and a Relevant Algorithm for Optimizing the Intial Cluster Centroid	Algoritma <i>K-Means</i> tradisional sangat sensitif dalam memilih pusat pengelompokan dan menghitung dengan mudah konvergen ke solusi yang optimal secara lokal. Selain itu algoritma tradisional memiliki kecepatan pengelompokan yang rendah dan masalah	Metode <i>K-Means Clustering</i> , dan <i>Particle Swarm Optimization</i> .	Algoritma <i>K-Means</i> yang diusulkan memiliki kinerja terbaik diantara enam algoritma, pengelompokan data ini mencapai peringkat pertama dalam perbandingan peringkat rata-rata, lebih tinggi daripada skor rata-rata dari tiga algoritma

			kemacetan memori saat memproses data yang sangat besar.		pengelompokan lainnya. Dalam hal kecepatan, algoritma yang diusulkan menunjukkan signifikan. Secara keseluruhan, hasil menunjukkan konvergensi dan akurasi yang sangat baik.
7.	Nanang, Lestio Wibowo, Moch Arief Soeleman, Ahmad Zainul Fanani, 2021	Antlion Optimizer Algorithm Modification for Initial Centroid Determination in K-means Algorithm	Metode <i>K-Means</i> mengelompokan data berdasarkan jarak k dari menentukan <i>centroid</i> awal untuk di proses. Pemilihan <i>centroid</i> yang sembarangan dapat mengakibatkan proses pengelompokan terjebak pada optimum lokal.	Metode <i>K-Means Clustering</i> , dan <i>Particle Swarm Optimization</i> .	Nilai SICD terbaik untuk setiap metode dan dataset diukur dengan memberikan peringkat yang baik pada metode yang diusulkan pada dataset Iris, Wine dan Cancers mendapatkan peringkat pertama pada dataset Ecoli dan Glass, metode yang diusulkan dan metode <i>K-Means ++</i> , keduanya mendapatkan peringkat rangking pertama, yang memberikan bukti bahwa dapat meningkatkan hasil <i>clustering</i> dan dapat menjadi metode alternatif untuk menentukan

					pusat <i>cluster</i> awal dengan menggunakan metode <i>K-Means</i> .
8.	Ari Yunus Hendrawan, 2020	Peningkatan kinerja Algoritma <i>K-Means</i> dengan Menggunakan <i>Particle Swarm Optimization</i> dalam Pengelompokkan Data Penyediaan Akses Sanitasi dan Air Bersih	Kebingungan menentukan daerah yang akan dijadikan acuan dalam menentukan kelompok daerah pada data PAMSIMAS.	Metode <i>K-Means</i> , <i>Particle Swarm Optimization</i> dan <i>Davies Buildin Index</i> .	<i>K-Means</i> = 0,208856082. <i>K-Means</i> standar pemerintah = 0,280077. <i>K-Means</i> dengan <i>Particle Swarm Optimization</i> dengan <i>Davies Buildin Index</i> = 0,08383.
9.	Achmad Saiful, dan Joko Lianto Buliali, 2018	Implementasi Particle Swarm Optimization pada <i>K-Means</i> untuk Clustering Data Automatic Dependent Surveillance Broadcast	Adanya 212 penyelidikan kecelakaan penerbangan di Indonesia pada tahun 2010-2016.	Metode <i>K-Means</i> , <i>Particle Swarm Optimization</i> dan <i>Davies Buildin Index</i> .	Menghasilkan nilai 0,779.
10.	Yue Li, Xiaoquan Chu, Dong Tian, Jianying Feng, and Weisong Mu	Customer segmentation using <i>K-means</i> clustering and the adaptive particle swarm optimization algorithm	Adanya masalah ketidaksamaan untuk mengelompokkan campuran data.	Metode <i>K-Means</i> , <i>Particle Swarm Optimization</i> dan <i>Davies Buildin Index</i> .	Menunjukkan keefektifan dan kepraktisan IKM-ALPSO untuk segmentasi pelanggan.

Dari beberapa tinjauan pustaka penelitian sebelumnya, yang menjadi referensi utama adalah penelitian [21]. Penelitian ini berbeda dengan penelitian sebelumnya, terutama dalam optimasi *K-Means Clustering* untuk identifikasi daerah endemik penyakit menular menggunakan algoritma *Particle Swarm Optimization* di Kota Semarang. Masalah yang dihadapi adalah ketidakakuratan dalam penanganan penyakit menular. Metode yang digunakan meliputi *K-Means*, *Particle Swarm Optimization*, dan *Davies-Bouldin Index*. Hasilnya menunjukkan bahwa penggunaan *K-Means* saja menghasilkan nilai 0,619, sedangkan *K-Means* dengan *Particle Swarm Optimization* menghasilkan nilai 0,113.

2.2 Landasan Teori

2.2.1 Supermarket

Meningkatnya tingkat persaingan dalam dunia bisnis, terutama di sektor supermarket, para pengembang dituntut untuk mencari pola baru yang dapat meningkatkan penjualan dan strategi pemasaran produk di supermarket. Salah satunya adalah dengan memanfaatkan data transaksi. Supermarket merupakan salah satu jenis usaha ritel yang menawarkan berbagai macam produk kepada pelanggan mereka.

2.2.2 Strategi Penjualan

Strategi penjualan adalah rencana yang dilakukan oleh perusahaan untuk menentukan bagaimana cara agar dapat meningkatkan kapasitas penjualan produknya dan memastikan kepuasan dari permintaan konsumen. Pada strategi penjualan ini terdapat tujuan yaitu untuk meningkatkan pengelolaan stok, mengurangi pemborosan, mengoptimalkan promosi, dan lebih responsif terhadap perubahan permintaan pasar. Pada strategi penjualan, diperlukan adanya tools untuk menarik daya tarik konsumen salah satunya dengan menggunakan teknologi informasi seperti metode *clustering* dalam data mining.

2.2.3 Preprocessing

Sebelum melaksanakan tahapan data mining, diperlukan proses pembersihan data pada fokus *Knowledge Discover in Database* (KDD). Proses pembersihan ini melibatkan tindakan seperti menghapus duplikasi data, memeriksa konsistensi data, dan memperbaiki kesalahan data, termasuk kesalahan cetak seperti tipografi. Selain itu, dilakukan juga tahap *enrichment* yang melibatkan penambahan data atau informasi yang relevan dan diperlukan untuk proses *Knowledge Discoveri in Database* (KDD) termasuk data atau informasi eksternal.

2.2.4 Exploratory Data Analysis (EDA)

Exploratory Data Analysis (EDA) merupakan tahap awal yang digunakan untuk menggali struktur dan karakteristik dari data tersebut. *Exploratory Data Analysis* (EDA) ini memiliki tujuan untuk memperoleh wawasan dan informasi

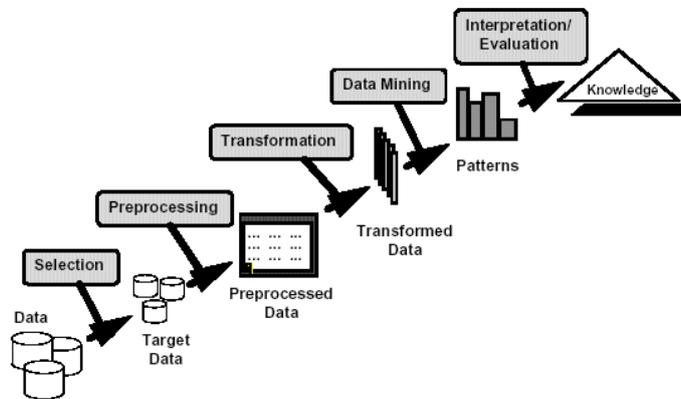
penting yang terdapat pada data sebelum melakukan analisis lebih lanjut atau membangun model. Pentingnya *Exploratory Data Analysis* (EDA) terletak pada kemampuannya membantu peneliti atau ilmuwan data merumuskan pertanyaan yang relevan, memastikan kualitas data, dan mengambil keputusan tentang pendekatan analisis yang akan dilakukan selanjutnya. Untuk mengeksplorasi data yang belum dikenal maka para ilmuwan data secara aktif melakukan serangkaian operasi analisis seperti filtering, agregasi dan filtering. Dalam proses *clustering* penghapusan *missing value*, penghapusan *duplicated* data bisa diterapkan pada *Exploratory Data Analysis* (EDA)

2.2.5 Data Mining

Data mining merupakan rangkaian tindakan dengan tujuan untuk mengungkapkan informasi baru yang sebelumnya tidak diketahui dari sejumlah data yang ada. Proses ini dilakukan dengan cara tertentu untuk mengolah data menjadi informasi yang lebih berarti, melalui ekstraksi dan pengidentifikasian pola-pola penting dalam dataset. Data mining melibatkan analisis mendalam terhadap data untuk menemukan pengetahuan baru dalam kumpulan data tersebut. Berikut langkah – langkah proses data mining: [22]

1. Tahap seleksi data melibatkan pengambilan data yang sesuai dari basis data untuk keperluan analisis.
2. *Preprocessing*, dimana data yang sudah dipilih mengalami pembersihan guna dapat menghilangkan duplikasi, mendeteksi inkonsistensi, serta memperbaiki kesalahan yang ada pada data.
3. Transformation, pada langkah ini diperlukan kreativitas dan sangat berguna pada jenis informasi yang dikehendaki dari sumber data. Tujuannya adalah mengubah data yang telah terpilih.
4. Data mining, melibatkan penerapan algoritma atau metode pintar untuk mengekstraksi informasi penting dari data.
5. Evaluasi, digunakan dalam mengidentifikasi pola atau informasi yang bisa dipahami oleh individu atau kelompok. Selanjutnya, tahap ini juga

dimanfaatkan untuk mendapatkan data yang mungkin tidak sejalan dengan informasi sebelumnya



Gambar 2.1 Proses Data Mining [22]

2.2.6 Clustering

Clustering merupakan proses pengelompokan data menjadi beberapa bagian sehingga data pada *cluster* memiliki kemiripan yang tinggi, tetapi sangat berbeda dengan data di *cluster* lainnya. Secara keseluruhan metode ini dapat diklasifikasi ke dalam beberapa kategori, termasuk diantaranya: [23]

1. Metode Partisi

Mengelompokkan data ke dalam kelompok sedemikian rupa sehingga setiap kelompok memuat minimal satu entitas. Metode ini membagi kumpulan data ke dalam bagian yang lebih spesifik dan pada dasarnya hanya menerapkan pengelompokan

2. Metode hirarki

Metode ini adalah proses pengaturan data dalam kelompok berdasarkan hierarki yang dihasilkan. Pendekatan ini dapat diinterpretasikan dari atas ke bawah, dimulai dengan semua objek dalam satu *cluster* tunggal, dan kemudian secara bertahap membagi *cluster* tersebut menjadi kelompok yang lebih kecil hingga setiap objek berada dalam *cluster* sendiri.

3. Metode berbasis kepadatan

Memisahkan kumpulan objek menjadi kelompok *eksklusif*. Umumnya, pendekatan ini dilakukan untuk memperhitungkan kelompok yang eksklusif dan tidak mempertimbangkan kelompok *fuzzy*.

4. Metode berbasis grid

Mengkuantisasi spasial objek dengan membaginya menjadi sel-sel terbatas yang membentuk suatu kerangka *grid*. Setelah itu, semua proses pengelompokan dilaksanakan pada kerangka grid tersebut, yakni di dalam ruang yang telah diukur secara terbatas.

2.2.7 *K-Means*

Algoritma *K-Means* adalah metode pengelompokan data, dimana algoritma ini adalah bentuk *clustering* non-hirarki yang bertujuan untuk membentuk beberapa kelompok dari data dengan karakteristik yang serupa. Data tersebut terdapat karakteristik yang berbeda jika dikelompokkan ke dalam kelompok yang sesuai dengan karakteristiknya, sehingga hasilnya adalah kelompok-kelompok yang memiliki perbedaan yang sangat kecil atau minimal.

Tahap-tahapan Algoritma *K-Means*: [24]

1. Menentukan *centroid* dari *cluster*

Menghitung jarak masing-masing data ke *centroid*

$$d(x_{ij}, c_{ij}) = \sqrt{\sum_{i=1}^{N_x} \sum_{j=1}^{N_f} (x_{ij} - c_{ij})^2} \dots \dots \dots (2.1)$$

Dengan keterangan sebagai berikut:

N_x : Jumlah data

N_f : Jumlah dimensi

x_{ij} : Data ke -i dengan atribut ke -j

c_{ij} : Data ke -i dengan *centroid* ke -j

2. Data tersebut dikelompokkan berdasarkan kedekatannya pada pusat *cluster* terdekat.
3. Pembaruan pusat *cluster* dilakukan berdasarkan data yang sudah dikelompokkan sebelumnya.
4. Lakukan perulangan langkah-langkah tersebut sampai tidak ada perubahan pada posisi pusat *cluster* dan jarak minimal atau maksimal data ke *cluster* tidak melebihi ambang batas yang ditetapkan.

2.2.8 Particle Swarm Optimization

Algoritma *Particle Swarm Optimization* dalam melakukan pencarian solusi dilakukan secara random dari suatu himpunan, yang masing-masing partikel dikaitkan terkait dengan letak serta kecepatannya dalam melakukan eksplorasi baru yang dinamis berdasarkan perilakunya. Tiap partikel memiliki skor kecocokan yang perlu dinilai pada setiap siklus generasi berdasarkan nilai terbaik lokal (*pbest*) dan nilai terbaik global (*gbest*) yang mencerminkan pengalaman tiap partikel dalam menghasilkan solusi terbaik.

Istilah umum pada algoritma *Particle Swarm Optimization* adalah sebagai berikut: [25]

- a. *Swarm* : Kelompok partikel.
- b. Partikel : Individu dalam sebuah kelompok yang menampilkan solusi dari masalah yang dihadapi.
- c. *Pbest* : Letak pada sebuah partikel.
- d. *Gbest* : Letak terbaik dari sekelompok partikel dalam *Pbest*.
- e. *Velocity* : Kecepatan partikel menunjukkan arah gerak dengan memperbaiki posisi sebelumnya.
- f. Bobot *Inersia* : Untuk mengatur perubahan kecepatan pada partikel.

Rumus bobot *inertia* (w) :

$$W = w_{max} - \frac{w_{max} - w_{min}}{iter_{max}} \times iter \dots \dots \dots (2.2)$$

Dengan keterangan sebagai berikut :

w_{max} : Bobot inersia maximal

w_{min} : Bobot inersia minimal

$iter_{max}$: Iterasi maksimal

$iter$: Iterasi saat ini

- g. Koefisien Akselerasi (c_1 dan c_2) : konstanta yang membantu perpindahan sebuah partikel dalam suatu kelompok.

Langkah – langkah *Particle Swarm Optimization*: [25]

1. Inisialisasi sejumlah partikel secara random.
2. Hitung nilai kecocokan (*fitness*) pada masing-masing partikel
3. Perbarui nilai *Pbest* sesuai dengan rumus:

$$Pbest_i = F(x_i) \geq F(Pbest_i) \dots \dots \dots (2.3)$$

Dengan keterangan sebagai berikut:

$F(x_i)$: Nilai *fitness* pada partikel ke – i

$F(Pbest_i)$: Nilai *fitness* pada partikel baik ke - i

4. Perbarui nilai *Gbest* dengan rumus:

$$best_i = MAX (F(Pbest_i), \dots , F(Pbest_n)) \dots \dots \dots (2.4)$$

Dengan keterangan sebagai berikut :

$MAX (F(Pbest_i), \dots , F(Pbest_n))$: *Fitness terbesar dari seluruh Pbest*

5. Perbarui nilai kecepatan dengan rumus:

$$v_{ij}(t + 1) = w \cdot v_{ij}(t) + c_1 \cdot r_1 [P_{ij}(t) - x_{ij}(t)] + c_2 \cdot r_2 [G_j(t) - x_{ij}(t)] \dots \dots \dots (2.5)$$

Dengan keterangan sebagai berikut:

w : Bobot kriteria

v_{ij} : Kecepatan iterasi

c_1 dan c_2 : Nilai konstanta akselerasi

r_1 dan r_2 : Nilai random dari 0 sampai

$x_{ij}(t)$: Partikel ke – i dengan atribut ke – j pada iterasi ke – t

$x_{ij}(t)$: Partikel balik ke – j dengan atribut ke – j

$x_{ij}(t)$: Atribut ke -j pada artikel terbalik

6. Perbarui nilai posisi dengan rumus:

$$x_{ij}(t + 1) = x_{ij}(t) + v_{ij}(t + 1) \dots \dots \dots (2.6)$$

Dengan keterangan sebagai berikut:

$v_{ij}(t + 1)$: Nilai kecepatan iterasi ke $-t + 1$

$x_{ij}(t)$: Partikel ke $-i$ dengan atribut ke $-j$ pada iterasi ke $-t$

7. Ulangi langkah 2, 3, dan 4 hingga iterasi maksimum.

2.2.9 Nilai fitness *Particle Swarm Optimization Clustering*

Nilai fitness ini diperoleh dengan menghitung rata-rata jarak antara data dan pusat *cluster*. Nilai fitness dapat diukur menggunakan persamaan di bawah ini: [25]

$$f = \frac{\sum_{i=1}^{N_c} \left(\frac{\sum_{j=1}^{N_x} d(x_j, c_i)}{N_x} \right)}{N_c} \dots\dots\dots(2.7)$$

Dengan keterangan:

N_c : Jumlah cluster

N_x : Jumlah data pada cluster ke $-i$

$d(x_j, c_i)$: Jarak antara data ke $-i$ dengan cluster ke $-i$

2.2.10 *Particle Swarm Optimization* dikombinasikan dengan *K-Means*

Algoritma ini menggabungkan dua metode, yaitu *Particle Swarm Optimization* dengan *K-Means*. Pada tahap implementasinya, *Particle Swarm Optimization* pertama-tama dijalankan sampai mencapai iterasi terakhir, kemudian *K-Means* dijalankan. Maka, *Particle Swarm Optimization* bertanggung jawab untuk menginisialisasi pusat-pusat *cluster* awal untuk *K-Means*, sehingga pendekatan ke hasil optimal dapat tercapai. Hasilnya, proses pengelompokan data oleh *K-Means* menjadi lebih efisien [25]

2.2.11 *Sum of Squared Error* (SSE)

Sum of Squared Error (SSE) adalah perbedaan kuadrat antara setiap pengamatan dan rata-rata (*mean*) kelompoknya. *Sum of Squared Error* (SSE) dapat digunakan sebagai ukuran variasi dalam sebuah cluster. Jika semua kasus dalam cluster identik, maka *Sum of Squared Error* (SSE) akan sama dengan 0.

Sum of Squared Error (SSE) berfungsi dalam mengukur kualitas yang merupakan bagian dari *cluster* serta kualitas hasil dari penjumlahan seluruh jarak masing-masing data dengan titik pusat *cluster* tersebut. Semakin kecil nilai dari *Sum of Squared Error* (SSE), maka semakin tinggi-tingkat similaritas data yang ada di dalam masing-masing *cluster* atau semakin baik *cluster* yang diperoleh. Rumus yang diterapkan agar mendapatkan nilai *Sum of Squared Error* (SSE) adalah sebagai berikut:

$$SSE = \sum_{i=1}^n (d_i - c_i)^2 \dots\dots\dots(2.8)$$

Dengan keterangan:

Sum of Squared Error (SSE): nilai kuadrat selisih antara koordinat *centroid* setiap data

n : jumlah data

d_i : nilai data ke- i

c_i : nilai *centroid* ke- i

2.2.12 *Quantization Error dan Quantization_mean*

Quantization Error adalah nilai tunggal yang mengukur total kesalahan kuantisasi untuk satu kali penerapan dalam *clustering*. Hal ini dihitung sebagai jumlah jarak kuadrat antara setiap titik data dan *centroid cluster* terdekatnya. Semakin kecil nilai *Quantization Error* maka, semakin baik representasi data oleh *centroid cluster* yang dipilih. Dengan kata lain, nilai *Quantization Error* yang lebih rendah menunjukkan bahwa *centroid cluster* mampu secara akurat merepresentasikan data dalam *cluster* tersebut. Dalam konteks evaluasi kualitas *cluster*, nilai *Quantization Error* yang rendah menunjukkan bahwa mampu menghasilkan *cluster* yang dimana data dalam setiap *cluster* memiliki kemiripan yang tinggi dengan *centroid cluster*. Sebaliknya, nilai *Quantization Error* yang tinggi menandakan bahwa *cluster* tidak kompak dan data di dalamnya tersebar secara luas dari *centroid cluster*. Sedangkan, *quantization_mean* adalah rata-rata dari nilai *Quantization Error* yang dihitung dari beberapa kali penerapan algoritma *clustering*. Misalnya, jika menjalankan algoritma sebanyak 20 kali,

quantization_mean adalah rata-rata dari 20 nilai *Quantization Error* yang diperoleh dari setiap iterasi tersebut

2.2.13 *Davies-Bouldin Index*

Davies-Bouldin Index adalah metode yang digunakan untuk mengukur akurasi *clustering*. Konsistensi mengukur jarak total data ke pusat *cluster* yang sudah diacak. Di sisi lain, separasi mengacu pada jarak antara pusat *cluster* dan *cluster* lainnya. Langkah awal dalam evaluasi adalah mengidentifikasi nilai *Sum of Square Within cluster* (SSW) dan *Sum of Square Between cluster* (SSB) pada titik fokus awal, kemudian menentukan *Davies-Bouldin Index*. Rumus yang diterapkan agar mendapatkan nilai *Sum of Square Within cluster* (SSW) adalah sebagai berikut: [26]

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \dots \dots \dots (2.9)$$

Dengan keterangan sebagai berikut:

M : jumlah data pada *cluster* ke i

x : data *cluster*

c : *centroid*

d(x, c) : jarak data dengan *centroid*

Sum of Square Between-cluster (SSB) berfungsi dalam menentukan pemisahan antar *cluster* satu dengan *cluster* lainnya.

$$SSB_{i,j} = d(x_j, c_i) \dots \dots \dots (2.10)$$

Dengan keterangan sebagai berikut:

C_i : *Cluster* satu

C_j : *Cluster* lain

d(x_j, c_i) : Jarak *centroid* satu dengan *centroid* lain

Selanjutnya untuk menghitung nilai skala yang dimiliki oleh setiap *cluster*, digunakan persamaan berikut:

$$R_{i,j} = \frac{SSW_i + SSW_j}{SSW_{i,j}} \dots \dots \dots (2.11)$$

Rasio $R_{i,j}$ adalah pengukuran yang menunjukkan seberapa baik nilai perbandingan antara *cluster* ke i dengan *cluster* ke j . *Cluster* yang optimal adalah *cluster* yang terdapat sejumlah kecil koherensi internal dan sejumlah besar separasi antar *cluster*.

Persamaan *Davies Buildin Index* adalah sebagai berikut :

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \dots \dots \dots (2.12)$$

Dengan keterangan sebagai berikut:

$R_{i,j}$: Nilai rasio perbandingan antara *cluster* ke- i dan *cluster* ke- j

K : jumlah *cluster* yang terbentuk