

Optimasi K-Means untuk Pengelompokan Jumlah Sekolah di Riau Menggunakan Particle Swarm Optimization

Dicky Surya¹, Taslim², Dafwen Toresa³, Fajrizal⁴, Susi Handayani⁵

³Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Lancang Kuning
^{1,2,3}Jl. Yos Sudarso KM. 8 Rumbai, Pekanbaru, Riau, telp. 0811 753 2015

e-mail: dickysurya081@gmail.com, taslim@unilak.ac.id, dafwen@unilak.ac.id,
fajrizal@unilak.ac.id, susi@unilak.ac.id

Abstrak

Pendidikan merupakan aspek yang sangat penting dalam menciptakan sumber daya manusia yang unggul dan mampu bersaing. Jarak tempuh yang jauh antara sekolah dan tempat tinggal mempengaruhi murid melanjutkan pendidikannya yang diakibatkan oleh ketidakmerataan distribusi sekolah. Peristiwa ini dapat mempengaruhi tingkat indeks pendidikan di provinsi Riau dan kualitas sumber daya manusianya. Penelitian ini bertujuan untuk mengelompokkan jumlah sekolah di provinsi Riau berdasarkan kecamatan menggunakan algoritma K-Means yang dioptimalkan dengan algoritma Particle Swarm Optimization untuk menentukan centroid awal dan algoritma Elbow dalam menentukan jumlah kluster yang optimal. Hasil penelitian menunjukkan bahwa optimasi yang dilakukan pada metode K-means memberikan hasil clustering yang optimum yang dilihat dari Davies-Bouldin Index (DBI), Silhouette Score, Sum of Square Error (SSE), dan Quantization error daripada tidak dilakukannya optimasi.

Kata Kunci: K-Means, PSO, elbow, klasterisasi, jumlah sekolah.

Abstract

Education is a very important aspect in creating superior and competitive human resources. The long distance between school and residence influences students to continue their education due to the unequal distribution of schools. This event can affect the level of education index in Riau province and the quality of its human resources. This research aims to group the number of schools in Riau province based on sub-districts using the K-Means algorithm which is optimized with the Particle Swarm Optimization algorithm to determine the initial centroid and the Elbow algorithm to determine the optimal number of clusters. The results of the research show that optimization carried out using the K-means method provides optimum clustering results as seen from the Davies-Bouldin Index (DBI), Silhouette Score, Sum of Square Error (SSE), and Quantization error compared to not optimizing

Keywords: K-Means, PSO, elbow, clustering, number of schools.

1. PENDAHULUAN

Pendidikan adalah aspek yang sangat penting dalam kehidupan karena keterkaitannya dengan mutu sumber daya manusia sebagai kunci pertumbuhan negara. Untuk mencapai kesuksesan dalam pembangunan nasional, perlu dilakukan pengembangan di sektor pendidikan guna menciptakan sumber daya manusia (SDM) yang berkualitas [1]. Pendidikan berkualitas menjadi investasi untuk menciptakan sumber daya manusia yang unggul dan mampu bersaing [2].

Meskipun demikian, Indonesia masih menghadapi tantangan besar dalam meningkatkan kualitas Pendidikan. Berdasarkan laporan, Indonesia menduduki peringkat keenam terbawah dengan kualitas pendidikan yang rendah, berada di peringkat 52 dari 73 negara di ASEAN. Pendidikan menjadi fokus utama dalam mencapai pembangunan unggul di Indonesia. Salah satu tantangan utama adalah ketidakmerataan akses pendidikan, khususnya di Provinsi Riau. Di Provinsi Riau sendiri berdasarkan data BPS tahun 2020, provinsi Riau pernah menempati peringkat keempat sebagai provinsi dengan tingkat melek huruf tertinggi untuk usia di atas 15 tahun di Indonesia. Secara bersamaan, pencapaian ini

juga merupakan yang tertinggi di Pulau Sumatera. Dengan merujuk pada data Badan Kependudukan dan Keluarga Berencana Nasional (BKKBN), Provinsi Riau memiliki sekitar 167.000 anak usia sekolah yang tercatat sebagai anak putus sekolah. Salah satu faktor penyebabnya adalah distribusi sekolah yang tidak merata, di mana sekolah-sekolah lebih banyak terkonsentrasi di wilayah perkotaan daripada di daerah pedesaan. Hal ini dikarenakan adanya perbedaan pada kepadatan penduduk antar daerah, kesenjangan sarana pendidikan dan jarak antara sekolah dan tempat tinggal sehingga mempengaruhi murid melanjutkan pendidikannya [3]. Dampak dari ketidakmerataan akses terhadap pendidikan menjadi terbatas di daerah-daerah tertentu, yang pada gilirannya berdampak pada kualitas SDM dan peningkatan angka kemiskinan di Provinsi Riau. Sehingga perlu dilakukan klasterisasi jumlah sekolah di Provinsi Riau untuk mengatasi masalah ini. Pada penelitian ini dilakukan klasterisasi data jumlah sekolah di Provinsi Riau berdasarkan Kecamatan.

Klasterisasi adalah metode yang memisahkan data menjadi beberapa kelompok yang memiliki makna. Salah satu algoritma pengelompokan yang paling populer adalah *k-means*. Algoritma *K-Means* memiliki ketelitian cukup tinggi pada ukuran objek, sehingga algoritma ini relatif terukur dan efisien untuk pengolahan objek yang jumlahnya besar dan juga *K-Means* tidak terpengaruh terhadap urutan objek [3]. Metode ini bekerja dengan membagi data menjadi beberapa kelompok berdasarkan jarak antara data dengan *centroid* kelompok tersebut [4][5].

Namun, *K-Means Clustering* memiliki kelemahan yaitu sensitifitasnya terhadap inisialisasi *centroid* awal dan pemilihan jumlah kluster [6][7]. Oleh karena itu perlu dilakukannya optimasi untuk *K-Means clustering* menggunakan algoritma PSO (*Particle Swarm Optimization*) untuk mendapatkan nilai *centroid* awal yang optimal [8]. PSO merupakan metode optimasi berdasarkan perilaku sekelompok burung atau ikan yang meniru perilaku sosial organisme tersebut [9][10]. Konsep dasar dari PSO adalah bahwa setiap individu dalam kawanan, yang disebut sebagai partikel, mewakili solusi potensial dari masalah yang sedang dipecahkan. Partikel-partikel ini bergerak dalam ruang pencarian dengan kecepatan tertentu, yang diperbarui secara dinamis berdasarkan pengalaman terbaik mereka sendiri dan juga pengalaman terbaik dari seluruh kawanan [11]. Penelitian ini bertujuan untuk melihat sejauh mana tingkat optimasi dari klasterisasi data jumlah sekolah di Provinsi Riau menggunakan algoritma optimasi PSO pada penentuan nilai *centroid* k-means.

2. METODE PENELITIAN

2.1. Studi Literatur

1. *K-Means*

K-Means merupakan salah satu metode *cluster analysis non hirarki* yang berusaha untuk mempartisi objek yang ada ke dalam satu atau lebih *cluster* atau kelompok objek berdasarkan karakteristiknya, sehingga objek yang mempunyai karakteristik yang sama dikelompokkan dalam satu *cluster* yang sama dan objek yang mempunyai karakteristik yang berbeda dikelompokkan ke dalam *cluster* yang lain [12]

$$D(x_1, x_2) = \sqrt{\sum_i^n 1(x_1 - x_2)^2} \quad (1)$$

Pada tahap awal, nilai *centroid* akan dipilih secara acak. Kemudian, setiap data yang ada akan ditempatkan pada salah satu pusat kelompok berdasarkan jarak terdekat antara data tersebut dengan pusat kelompok. Posisi pusat kelompok akan dihitung kembali hingga semua data dikelompokkan ke pusat kluster masing-masing dan posisi pusat kelompok yang baru terbentuk. Langkah ini akan berulang terus menerus sampai kondisi konvergen tercapai [13]. Salah satu kelemahan *K-Means* ialah yaitu sensitifitasnya terhadap inisialisasi *centroid* awal dan sensitif terhadap pemilihan jumlah *cluster* (*k*) yang dapat menghasilkan

kelompok yang tidak optimal [6][7]. Oleh karena itu perlu dilakukannya optimasi untuk *K-Means clustering* menggunakan algoritma *Elbow* untuk mengatasi pemilihan jumlah *cluster* (*k*) dan algoritma *PSO (Particle Swarm Optimization)* untuk mendapatkan nilai *centroid* awal yang optimal.

2. Algoritma Elbow

Algoritma *Elbow* merupakan metode yang digunakan untuk mencari jumlah *cluster* yang optimal pada suatu dataset pada analisis *clustering*. Algoritma ini memperhatikan nilai perbandingan *Sum of Square Error* pada setiap nilai *cluster* dimana antara jumlah *cluster* yang akan membentuk siku pada suatu titik [13].

$$SSE = \sum_{k=1}^k \sum_{xi \in Sk} \|X_i - C_k\|_2^2 \quad (2)$$

Dimana *SSE* merupakan nilai yang akan dicari, *k* adalah jumlah kluster, *xi* yaitu jumlah data dan *C_k* yaitu banyaknya kluster *i* pada kluster ke *k* [7].

3. Algoritma Particle Swarm Optimization

Particle Swarm Optimization merupakan metode optimasi berdasarkan perilaku sekelompok burung atau ikan yang meniru perilaku sosial organisme tersebut [9]. *PSO* merupakan metode berbasis populasi yang menggunakan cara untuk memanfaatkan setiap individu dalam menyelesaikan suatu masalah. Dimana individu yang bisa disebut dengan *particle*, tiap *particle* dapat berpindah menyesuaikan dengan daerah pencarian dan menyimpan posisi yang pernah dilewati [14][15].

Algoritma *PSO* dapat dijelaskan sebagai berikut:

- b. Inisialisasi Jumlah Partikel
- c. Inisialisasi kecepatan awal partikel secara acak
- d. Menghitung *fitness*, lalu hitung *Pbest* dan *Gbest*. Dalam menghitung *fitness* dapat menggunakan rumus *Sum of Square Error (SSE)*. Selanjutnya ialah perbarui *pBest* dan *gBest* dari perhitungan sebelumnya. *gBest* merupakan nilai *global* terbaik dari setiap *pBest* yang merupakan nilai terkecil.
- e. Perbarui kecepatan dan posisi. Dalam memperbarui kecepatan menggunakan rumus :

$$vi(t+1) = w \cdot vi(t) + c1 \cdot r1 \cdot (pi_best - xi(t)) + c2 \cdot r2 \cdot (gbest - xi(t)) \quad (3)$$

Dimana *w* adalah bobot kriteria, *V_{ij}* adalah kecepatan iterasi ke -*t*, *c₁* dan *c₂* adalah nilai konstanta akselerasi, *r₁* dan *r₂* adalah nilai random dari 0 sampai 1, *P_{ij}(t)* adalah partikel baik ke -*i* dengan atribut ke -*j*, *G_j(t)* adalah atribut ke -*j* pada *particle* terbaik, dan *x_{ij}(t)* adalah *particle* ke -*i* dengan atribut ke -*j* pada iterasi ke -*t*. Setelah mendapatkan nilai kecepatannya kita perlu memperbarui nilai posisi menggunakan rumus:

$$x3(t+1) = x3(t) + v3(t+1) \quad (4)$$

- f. Lakukan langkah evaluasi dan update posisi-kecepatan berulang kali hingga mencapai jumlah iterasi maksimum atau konvergensi (ketika perubahan dalam *gBest* menjadi sangat kecil).
- g. Integrasi dengan *K-means*. Setelah *PSO* menemukan *gBest*, yaitu sekumpulan *centroid* yang optimal, gunakan *centroid* ini sebagai inisialisasi dalam algoritma *K-means*.

4. Davies-Bouldin Index

Davies-Bouldin Index merupakan sebuah metrik yang digunakan untuk menilai kualitas klasterisasi pada analisis data. Metrik ini bertujuan untuk mengukur sejauh mana klaster yang dihasilkan oleh algoritma klasterisasi memisahkan kelompok data yang berbeda dan seberapa dekat data tersebut dengan pusat klasternya. Semakin rendah nilai *DBI*, semakin baik kualitas klasterisasi yang dihasilkan [16].

$$R_{ij} = \frac{S_i + S_j}{d_{ij}} \quad (5)$$

Dimana R_{ij} merupakan nilai DBI antara klaster C_i dan C_j , S_i menggambarkan dispersi atau penyebaran data dalam klaster C_i , yang dihitung menggunakan variasi atau rata-rata jarak antara titik-titik data dalam klaster dengan pusat klaster M_i , S_j menggambarkan dispersi atau penyebaran data dalam klaster C_j , yang dihitung menggunakan variasi atau rata-rata jarak antara titik-titik data dalam klaster dengan pusat klaster M_j , d_{ij} merupakan jarak antara pusat klaster m_i dan m_j .

5. Silhouette Score

Silhouette Score adalah sebuah teknik yang dapat dilakukan dalam mengukur seberapa baik suatu objek atau data dalam suatu *cluster*. Proses ini melibatkan perhitungan *silhouette score* untuk setiap titik data, yang menunjukkan seberapa baik titik tersebut terintegrasi dalam kelompoknya sendiri dibandingkan dengan kelompok lainnya. Nilai perhitungan yang dihasilkan oleh *Silhouette Coefficient* adalah -1 hingga 1 dimana nilai yang mendekati merupakan yang terbaik [17].

$$S_i = \frac{(b_i - a_i)}{\max(b_i - a_i)} \quad (6)$$

Dimana a_i adalah rata-rata jarak objek i dengan seluruh objek dalam sebuah *cluster* dan b_i adalah nilai terkecil rata-rata objek i dengan objek pada *cluster* yang berbeda.

6. Sum of Square Error (SSE)

Sum of Square Error ialah salah satu cara dalam melakukan validasi *cluster* melalui jumlah kuadrat setiap anggota *cluster* menuju pusatnya [18]. Nilai SSE terbaik di ukur dari nilai yang paling rendah.

$$SSE = \sum_{k=1}^K \sum_{x_i \in S_k} \|X_i - C_k\|_2^2 \quad (7)$$

Dimana K adalah jumlah *cluster*, X_i adalah attribut data, X_{ci} adalah attribut *centroid*, dan C_i adalah *centroid*.

7. Quantization Error

Quantization Error merupakan salah satu cara yang dapat digunakan untuk melakukan pengukuran kualitas dari algoritma klasterisasi. Nilai *Quantization Error* lebih rendah menunjukkan bahwa hasil yang terbaik [19].

$$QE = \frac{\sum_{j=1}^k \left\{ \frac{\sum_{i=1}^{N_j} \|x_i - c_j\|^2}{N_j} \right\}}{k} \quad (7)$$

Dimana k = jumlah *cluster*, c_j = *centroid cluster j*, N_j = jumlah titik data dalam *cluster-j* dan $\|x_i - c_j\|^2$ = jarak antara titik data dan *centroid*.

2.2. Data

Data yang digunakan ialah data jumlah sekolah di Riau berdasarkan kecamatan dengan jumlah 172 kecamatan. Data jumlah sekolah tersebut meliputi jumlah Taman Kanak-kanak (TK), Kelompok Bermain (KB), Tempat penitipan anak (TPA), Satuan PAUD Sejenis (SPS), Pusat Kegiatan Belajar Masyarakat (PKBM), Sanggar Kegiatan Belajar (SKB), Sekolah Dasar (SD), Sekolah Menengah Pertama (SMP), Sekolah Menengah Atas (SMA), Sekolah Menengah Kejuruan (SMK), dan Sekolah Luar Biasa (SLB). Data diperoleh dari website Data Pokok Pendidikan Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (<https://dapo.kemdikbud.go.id/>)

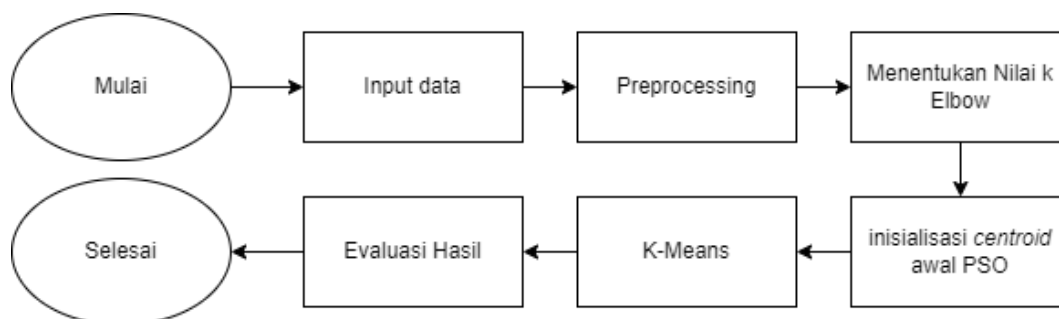
2.3. Lokasi Penelitian

Lokasi penelitian yaitu Jumlah Sekolah di Riau berdasarkan kecamatan. Data tersebut diperoleh dari website Data Pokok Pendidikan Kementerian Pendidikan, Kebudayaan, Riset dan Teknologi (<https://dapo.kemdikbud.go.id/>).

2.4. Metode Penelitian dan Evaluasi

Metode yang digunakan pada penelitian yakni *K-Means Clustering* yang di optimasikan oleh algoritma *Particle Swarm Optimization* dalam mencari *centroid* awal terbaik dan optimasi dalam menentukan jumlah kluster optimal menggunakan algoritma *Elbow*.

Ada beberapa proses pada penelitian ini yaitu data yang di input akan mengalami proses *preprocessing* terlebih dahulu sebelum diolah. Data yang telah di *preprocessing* akan masuk ke proses selanjutnya yaitu menentukan nilai k optimal menggunakan algoritma *Elbow* dan selanjutnya melakukan proses PSO untuk mendapatkan *centroid* yang optimal dilanjutkan proses *K-Means Clustering* hingga tidak ada perubahan *cluster*, dan terakhir proses evaluasi dengan cara di uji dari nilai *Davies-Bouldin index* (DBI), *Silhouette Score*, *Sum of Square Error* dan *Quantization error*.



Gambar 1. Flowchart Metode Penelitian

Ada beberapa proses pada penelitian ini yaitu data yang di input akan mengalami proses *preprocessing* terlebih dahulu sebelum diolah. Data yang telah di *preprocessing* akan masuk ke proses selanjutnya yaitu menentukan nilai k optimal menggunakan algoritma *Elbow* dan selanjutnya melakukan proses PSO untuk mendapatkan *centroid* yang optimal dilanjutkan proses *K-Means Clustering* hingga tidak ada perubahan *cluster*, dan terakhir proses evaluasi dengan cara di uji dari nilai *Davies-Bouldin index* (DBI), *Silhouette Score*, *Sum of Square Error* dan *Quantization error*.

3. HASIL DAN PEMBAHASAN

3.1. Normalisasi

Sebelum mengolah dataset perlu dilakukan normalisasi data karena merupakan langkah penting agar data yang diolah dapat bekerja dengan baik. Normalisasi data adalah proses mengubah nilai-nilai dalam dataset sehingga dapat disesuaikan atau dibawa ke skala yang seragam. Normalisasi menggunakan *MinMaxScaler* bertujuan agar mengubah skala fitur ke kisaran 0 hingga 1.

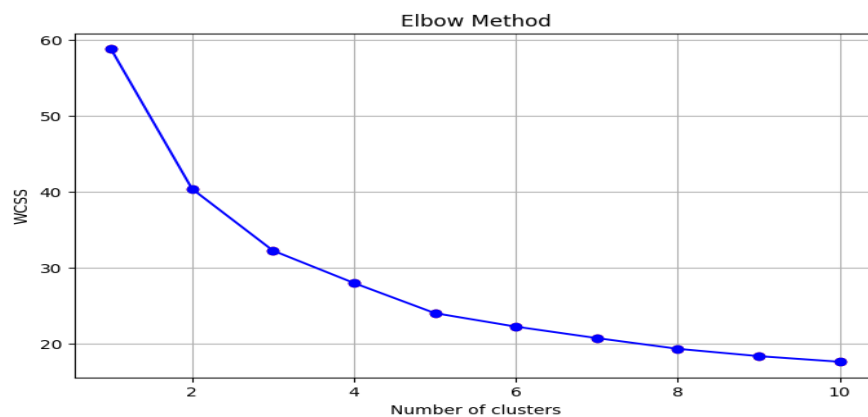
Tabel 1. Hasil Normalisasi

No	TK	KB	TPA	SPS	PKB M	SK B	SD	SMP	SMA	SMK	SL B
1	0,62	0,546	0,72 7	0	0,857	0	0,607	0,85 7	0,471	0,75	0,4
2	0,69	0,273	0,27 3	0,28 6	1	0	0,574	0,81	0,471	0,916 7	0,6

3	0,38	0,409	1	0	0,857	0	0,344 2	0,95 2	0,353	0,916 7	0,2
4	0,493	0,409	0,27 3	0	0,714	0	0,393	0,90 5	0,588 2	0,333	0,2
...	...	...	...	...	...	...	...	...	...	...	..
17 2	0,042 3	0,113 6	0	0	0	0	0,164	0,14 3	0,059	0,166 7	0

3.2 Optimasi Jumlah kluster dengan *Elbow*

Algoritma *Elbow* digunakan untuk mendapatkan jumlah kluster yang optimal pada metode *K-Means* [20]. Hasil pengujian didapatkan hasil kluster yang terbaik adalah 3 kluster seperti terlihat



Gambar 2. Grafik *Elbow*

Hasil pengujian dapat dilihat pada gambar 2 bahwa jumlah kluster yang optimal yaitu pada titik 3 ($k=3$) dimana penurunan setelahnya mulai melambat. Hal ini menunjukkan bahwa menambah lebih banyak cluster tidak memberikan keuntungan signifikan dalam mengurangi jarak dalam cluster.

3.3 *K-Means Clustering* tanpa optimasi *centroid* awal

Pengujian kali bertujuan untuk melihat hasil evaluasi *K-Means Clustering* tanpa optimasi *centroid* awal dimana akan menggunakan jumlah kluster yang diperoleh dari hasil pengujian *elbow* yaitu $k=3$.

```

Hasil Evaluasi Clustering menggunakan K-means:
DBI (Davies-Bouldin Index): 1.905494021128303
Silhouette Score: 0.2803158077887847
SSE (Sum of Squared Errors): 35.925176913954715
Quantization Error: 3.430727404282892
    
```

Gambar 3. Evaluasi *K-Means* Tanpa Optimasi

Hasil evaluasi *K-Means* tanpa optimasi *centroid* awal dapat dilihat pada gambar 3 yaitu nilai DBI yaitu 1.905494021128303, *Silhouette Score* yaitu 0.2803158077887847, SSE yaitu 35.925176913954715, dan *Quantization Error* yaitu 3.430727404282892.

Inisiasi Centroids awal:											
	Dimensi 1	Dimensi 2	Dimensi 3	Dimensi 4	Dimensi 5	Dimensi 6	Dimensi 7	Dimensi 8	Dimensi 9	Dimensi 10	Dimensi 11
Centroid											
0	0.140845	0.181818	0.000000	0.0	0.000000	0.0	0.229508	0.095238	0.117647	0.00	0.2
1	0.112676	0.113636	0.090909	0.0	0.142857	0.0	0.147541	0.238095	0.235294	0.25	0.0
2	0.042254	0.272727	0.000000	0.0	0.000000	0.0	0.491803	0.238095	0.176471	0.00	0.0

Gambar 4. *Centroid* Awal *K-Means*

Gambar 4 merupakan *centroid* awal yang digunakan dalam proses klusterisasi dimana pemilihan *centroid* tersebut masih dilakukan secara acak. Dimensi pada gambar 4 mewakili atribut dari data jumlah sekolah yaitu TK, KB, TPA dan lainnya. *Centroid* awal yang digunakan pada gambar 4 sebelumnya telah dilakukannya proses normalisasi *MinMax* pada data jumlah sekolah.

3.4 Optimasi *Centroid* awal dengan algoritma PSO

Pada pengujian sebelumnya telah melakukan klusterisasi data jumlah sekolah tanpa optimasi *centroid* awal menggunakan *K-means*. Namun, pada proses tersebut dalam menentukan pusat *cluster* awal masih menggunakan data acak sehingga hasil klusterisasi kurang optimal. Oleh karna itu, dapat mengujinya menggunakan metode optimasi PSO agar mendapatkan solusi hasil klusterisasi yang optimal. Dalam pengujianya langkah awal yaitu menginisialisasi jumlah *particles* yang akan digunakan. Pada penelitian ini pengujiannya menggunakan jumlah *particles* sebanyak 50. Setiap partikel dalam PSO mewakili sekumpulan *centroid* awal *K-Means* yang dimana tiap partikel ini akan melakukan proses klusterisasi *K-Means*. Hasil optimasi ini akan digunakan sebagai *centroid* awal pada proses *K-Means*. Pengujian kali bertujuan untuk melihat hasil evaluasi *K-Means Clustering* dengan optimasi *centroid* awal menggunakan PSO.

Hasil Evaluasi Clustering menggunakan Hybrid Particle Swarm Optimization *K-Means* dengan jumlah 50 Particle:
DBI (Davies-Bouldin Index): 0.9796658196970688
Silhouette Score: 0.5133949992062032
SSE (Sum of Squared Errors): 32.228553489562074
Quantization Error: 3.040925788784388

Gambar 5. Evaluasi *K-Means* Dengan Optimasi PSO

Hasil evaluasi *K-Means* tanpa optimasi *centroid* awal dapat dilihat pada gambar 5 yaitu nilai DBI yaitu 0.9796658196970688, *Silhouette Score* yaitu 0.5133949992062032, SSE yaitu 32.228553489562074, dan *Quantization Error* yaitu 3.040925788784388.

Tabel 2. *Centroid* awal *K-Means* Optimasi PSO

Cluster	TK	KB	TPA	SPS	PKBM	SKB	SD	SMP	SMA	SMK	SLB
C1	0,12 7	0,068	0	0	0	1	0,197	0,095	0,059	0,167	0
C2	0,04 2	0,227	0	0	0	0	0,295	0,286	0,118	0	0,2
C3	0,25 4	0,341	0,546	0,143	0,143	0	0,377	0,238	0,294	0,167	0,4

Tabel 2 merupakan solusi dari *centroid* awal yang optimal digunakan dalam klusterisasi. *Centroid* awal yang digunakan pada gambar 5 sebelumnya telah dilakukannya proses normalisasi *MinMax* pada data jumlah sekolah.

3.5 Evaluasi Akhir

Hasil proses optimasi K-means menggunakan algoritma PSO akan di evaluasi menggunakan *Davies–Bouldin index* (DBI), *Silhouette Score*, *Sum of Square Error* (SSE) dan *Quantization Error* (QE). Hasil evaluasi didapatkan sebagai berikut ini:

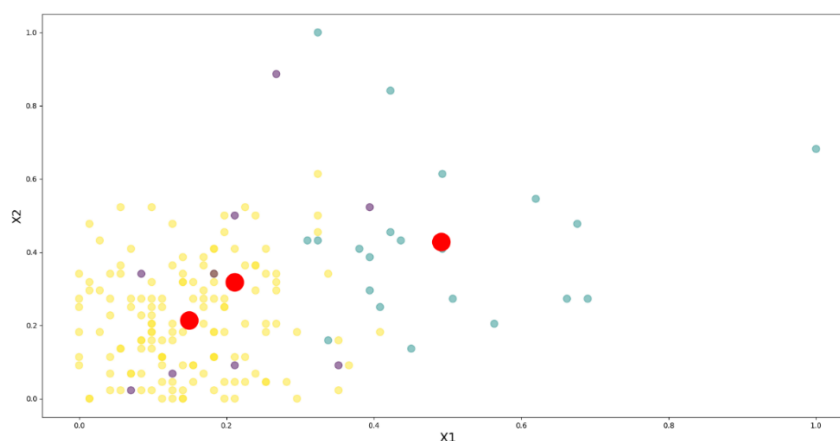
Tabel 3. Evaluasi Akhir

Algoritma	DBI	<i>Silhouette</i>	SSE	<i>Quantization</i>
K-means	1,90549402	0,28031581	35,9251769	3,430727404
Hybrid PSO K-Means (50)	0,97966582	0,513395	32,2285535	3,040925789

Berdasarkan hasil pengujian tersebut dapat disimpulkan bahwa pengujian *K-Means* dengan optimasi *centroid* awal menggunakan PSO mendapatkan hasil yang lebih baik daripada *K-Means* tanpa optimasi *centroid* awal dilihat dari:

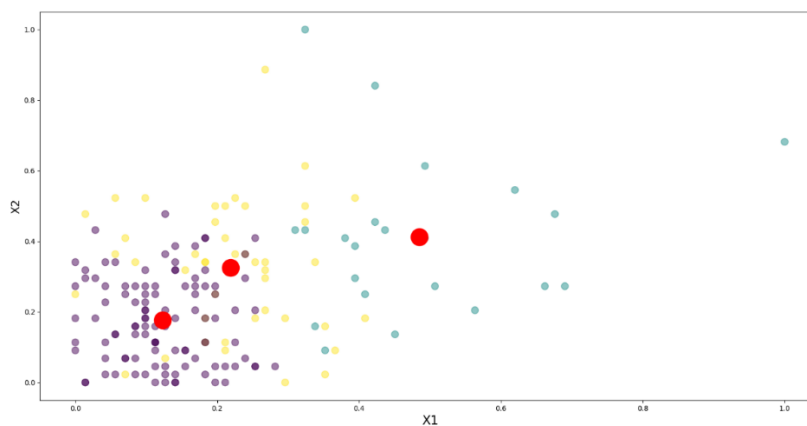
- Nilai *Davies–Bouldin index* (DBI) dari optimasi K-means menggunakan PSO dengan jumlah 50 *particle* menunjukan bahwa nilai DBI lebih daripada *K-Means* tanpa optimasi *centroid* awal yang berarti memiliki hasil yang lebih optimal atau lebih baik.
- Nilai *Silhouette Score* optimasi K-means menggunakan PSO dengan jumlah 50 *particle* lebih baik dari yang tidak di optimasi dikarenakan Nilai yang lebih tinggi menunjukan bahwa objek dalam satu kelompok lebih mirip satu sama lain dan lebih berbeda dari objek kelompok lain.
- Nilai *Sum of Square Error* (SSE) optimasi PSO K-means dengan jumlah 50 *particle* lebih baik dari yang tidak di optimasi dikarenakan nilai lebih kecil adalah lebih baik.
- Nilai *Quantization Error* optimasi PSO K-means dengan jumlah 50 *particle* lebih baik dari yang tidak di optimasi dikarenakan nilai lebih rendah adalah lebih baik.

Hasil visualisasi dari pengujian *K-Means* dengan optimasi *centroid* awal menggunakan PSO dengan jumlah 50 *particle* dapat dilihat pada gambar 6 dibawah ini :



Gambar 6. Hasil Visualisasi *K-Means* Dengan Optimasi PSO

Hasil visualisasi dari pengujian *K-Means* tanpa optimasi *centroid* awal dapat dilihat pada gambar 7 dibawah ini:

Gambar 7. Hasil Visualisasi *K-Means* Tanpa Optimasi

4. KESIMPULAN

K-means merupakan sebuah algoritma yang clustering menawarkan kombinasi yang baik antara kesederhanaan, efisiensi, dan fleksibilitas membuatnya menjadi salah satu metode yang banyak digunakan pada bidang analisa klasterisasi data, walaupun terdapat kelemahan dalam pada algoritma ini yaitu pemilihan *centroid* secara acak. Penelitian ini bertujuan untuk melakukan optimasi *centroid* awal pada *k-means* menggunakan algoritma *particle swarm optimization*, Dimana dari hasil uji coba dapat dilihat bahwa melakukan optimasi pada nilai *centroid* menghasilkan klaster yang lebih baik dibandingkan dengan klasterisasi *k-means* tanpa optimasi *centroid*. Hal ini dapat dilihat dari nilai evaluasi *Davies-Bouldin index* (DBI), *Silhouette Score*, *Sum of Square Error* dan *Quantization error* yang lebih baik jika dibandingkan dengan pemilihan *centroid* awal *K-Means* secara acak. Adapun saran untuk penelitian selanjutnya yaitu diharapkan dapat menggunakan metode clustering atau metode optimasi yang berbeda, sehingga dapat dijadikan sebagai perbandingan hasil algoritma mana yang jauh lebih baik.

UCAPAN TERIMA KASIH

Terimakasih kepada pihak pemberi dana penelitian, kepada lembaga/ orang yang membantu penelitian, kepada orang yang membantu dalam diskusi, dll.

DAFTAR PUSTAKA

- [1] E. Y. T. P. Dewi And I. Kamila, "Pengelompokan Wilayah Berdasarkan Faktor Pendukung Pendidikan Dengan Jumlah Sekolah Dan Jumlah Guru Menggunakan Algoritma K-Means," *Interval: Jurnal Ilmiah Matematika*, Vol. 2, No. 1, Pp. 1–12, 2022.
- [2] C. I. S. Puspa, D. N. O. Rahayu, And M. Parhan, "Transformasi Pendidikan Abad 21 Dalam Merealisasikan Sumber Daya Manusia Unggul Menuju Indonesia Emas 2045," *Jurnal Basicedu*, Vol. 7, No. 5, Pp. 3309–3321, 2023.
- [3] Z. I. Alfianti, "Pengelompokan Tingkat Pendidikan Berdasarkan Jumlah Sekolah Di Provinsi Jawa Barat Menggunakan Algoritma K-Means," 2022.
- [4] Budiman And T. P. Yoga, "Optimalisasi K-Means Berbasis Particle Swarm Optimization Untuk Hasil Produksi Tanaman Sayuran Di Indonesia," *Jurnal Nuansa Informatika*, Vol. 17, Pp. 2614–5405, 2023, Doi: 10.25134/Nuansa.
- [5] D. Toresa, "Implementasi K-Means Terhadap Penyebaran Penyakit Tbc Di Riau Menggunakan Rapid Miner," *Jutim (Jurnal Teknik Informatika Musirawas)*, Vol. 5, No. 1, Pp. 35–42, May 2020, Doi: 10.32767/Jutim.V5i1.809.
- [6] Taslim, D. Toresa, D. Jollyta, D. Suryani, And E. Sabna, "Optimasi K-Means Dengan Algoritma Genetika Untuk Target Pemanfaat Air Bersih Provinsi Riau," *Indonesian Journal Of Computer Science*, Vol. 10, No. 1, 2021.

- [7] A. R. Lashiyanti, I. R. Munthe, And F. A. Nasution, "Optimisasi Klasterisasi Nilai Ujian Nasional Dengan Pendekatan Algoritma K-Means, Elbow, Dan Silhouette," *Jurnal Ilmu Komputer Dan Sistem Informasi (Jikoms)*, Vol. 6, No. 1, Pp. 14–20, 2023.
- [8] R. G. Prasasti Alam And Y. Everhard, "Optimasi K-Means Dengan Particle Swarm Optimization Dalam Penentuan Titik Awal Pusat Klaster Data Telekomunikasi," *Techno. Com*, Vol. 23, No. 1, 2024.
- [9] A. Y. Hendrawan, "Peningkatan Kinerja Algoritma K Means Dengan Menggunakan Particle Swarm Optimization Dalam Pengelompokan Data Penyediaan Akses," *Electro Luceat*, Vol. 6, No. 2, Pp. 213–227, 2020.
- [10] I. Setiaji And A. Zainul Fanani, "Optimasi K-Means Clustering Dengan Menggunakan Particle Swarm Optimization Untuk Menentukan Jumlah Cluster Pada Kanker Serviks," Vol. 7, No. 3, Pp. 1463–1473, 2023, Doi: 10.30865/Mib.V7i3.6292.
- [11] F. Yusuf Bisilisin, Y. Herdiyeni, And B. Paruhum Silalahi, "Optimasi K-Means Clustering Menggunakan Particle Swarm Optimization Pada Sistem Identifikasi Tumbuhan Obat Berbasis Citra K-Means Clustering Optimization Using Particle Swarm Optimization On Image Based Medicinal Plant Identification System", [Online]. Available: [Http://Journal.lpb.ac.id/Index.php/Jika](http://Journal.lpb.ac.id/Index.php/Jika)
- [12] D. N. P. Sari And Y. L. Sukestiyarno, "Analisis Cluster Dengan Metode K-Means Pada Persebaran Kasus Covid-19 Berdasarkan Provinsi Di Indonesia," *Prisma, Prosiding Seminar Nasional Matematika*, Vol. 4, Pp. 602–610, 2021, [Online]. Available: [Https://Journal.unnes.ac.id/Sju/Index.php/Prisma/](https://Journal.unnes.ac.id/Sju/Index.php/Prisma/)
- [13] Y. S. Sari, H. Oktavianto, And H. Wahyu Sulisty, "Algoritma K-Means Dengan Metode Elbow Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah Berdasarkan Komponen Pembentuk Indeks Pembangunan Manusia K-Means Algorithm With Elbow Method To Grouping District/City In Central Java Based On Components Of Human Development Index," 2022. [Online]. Available: [Http://Jurnal.unmuhjember.ac.id/Index.php/Jst](http://Jurnal.unmuhjember.ac.id/Index.php/Jst)
- [14] Kuncoro Renaldy Surya, "Optimasi Penjadwalan Jam Kerja Dimasa Pandemi Covid-19 Menggunakan Metode Particle Swarm Optimization," *Jurnal Industri Kreatif Dan Informatika Series (Jikis)*, Vol. 01, May 2021.
- [15] A. S. Rahmadhnai, A. A. Supianto, And C. Dewi, "Penerapan Particle Swarm Optimization Pada Algoritme K-Means Untuk Pengelompokan Proses Berpikir Siswa Dalam Belajar," 2020. [Online]. Available: [Http://J-ptiik.ub.ac.id](http://J-ptiik.ub.ac.id)
- [16] I. T. Umagapi, B. Umaternate, H. Hazriani, And Y. Yuyun, "Uji Kinerja K-Means Clustering Menggunakan Davies-Bouldin Index Pada Pengelompokan Data Prestasi Siswa," *Prosiding Sisfotek*, Vol. 7, No. 1, Pp. 303–308, 2023.
- [17] L. Petra Refialy, H. Maitimu, And M. Soyano Pesulima, "Perbaikan Kinerja Clustering K-Means Pada Data Ekonomi Nelayan Dengan Perhitungan Sum Of Square Error (Sse) Dan Optimasi Nilai K Cluster," 2021.
- [18] D. Jollyta, S. Efendi, M. Zarlis, And H. Mawengkang, "Optimasi Cluster Pada Data Stunting: Teknik Evaluasi Cluster Sum Of Square Error Dan Davies Bouldin Index," 2019.
- [19] G. K. Patel, V. K. Dabhi, And H. B. Prajapati, "Clustering Using A Combination Of Particle Swarm Optimization And K-Means," *Journal Of Intelligent Systems*, Vol. 26, No. 3, Pp. 457–469, 2019.
- [20] R. B. Trianto, A. S. Nugroho, And E. Supriyadi, "Klasterisasi Menggunakan Algoritma K-Means Dan Elbow Pada Opini Masyarakat Tentang Kebijakan Sekolah Luring Tahun 2022," *Jurnal Inovtek Polbeng Seri Informatika*, Vol. 8, No. 1, Pp. 1–13, 2023.

