

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Pustaka

Dalam penelitian ini, akan dilakukan analisis beberapa literatur yang relevan untuk memberikan dukungan kepada penelitian. Tinjauan studi yang akan digunakan sebagai berikut:

Terdapat penelitian yang dengan menggunakan metode K-Means dan Algoritma Genetika untuk menyelesaikan masalah optimasi rute pendistribusian produk di PT Indomarco Adi Prima (Stock Point Nganjuk), yang melibatkan sejumlah titik distribusi yang perlu dikunjungi. Tujuan dalam penelitian ini adalah mencari rute teroptimal dengan total jarak tempuh sejauh 259,722337 km dan nilai keoptimalan *fitness* sebesar 0,385026. Hasil optimal diperoleh ketika menggunakan parameter tertentu, yaitu ukuran populasi sebanyak 1600, jumlah generasi 800, serta kombinasi nilai *crossover rate* (cr) sebesar 0,4 dan *mutation rate* (mr) sebesar 0,6 [18].

Selanjutnya terdapat penelitian lain menggunakan Algoritma Genetika dan K-Means dalam masalah didalam mengenai berkurangnya sumber air bersih di Provinsi di Provinsi Riau akibat epenbangan hutan dan transformasinya menjadi areal pertanian monokultur. Untuk mengukur validitas klaster, penelitian ini menggunakan *Davies Bouldin Index* (DBI). Klasterisasi tanpa optimasi menghasilkan DBI sekitar 2.164763, sedangkan dengan menerapkan optimasi genetika pada nilai centroid, DBI meningkat menjadi 2.06894 [19].

Masalah yang muncul dalam penelitian berikutnya adalah kekurangan variasi pada nilai awal *centroid* dalam algoritma K-Means. Kekurangan ini mengakibatkan hasil pengklasteran yang sangat responsif terhadap nilai awal tersebut. Oleh karena itu, metode DAC GA + K-Means dianggap lebih unggul dibandingkan dengan K-Means dan GA + K-Means dengan 3 klaster. Hasil dari metode ini mencapai nilai DBI masing-masing sekitar 0.138, 0.279, dan 0.382,

dengan *Sum Square Error* berturut-turut sekitar 92.56, 332.39, dan 1280.68. Selain itu, nilai *fitness* dari metode tersebut juga menunjukkan keunggulan, yaitu sekitar 7.12, 3.57, dan 2.13 [20].

Penelitian berikutnya menghadapi kendala dalam mengoptimalkan nilai *centroid* menggunakan K-Means, sehingga diperlukan penambahan algoritma genetika secara paralel untuk meningkatkan optimasi nilai *centroid* tersebut. Melalui kombinasi tersebut, hasil menunjukkan bahwa Algoritma Genetika berhasil mencapai tingkat keberhasilan sebesar 72,145%, sedangkan pengklasteran dengan menggunakan K-Means hanya mencapai 68,03% [21].

Penelitian selanjutnya sulit untuk menentukan jumlah kluster dan menginisialisasi kluster pusat secara sensitif, sehingga ditambahkan algoritma K-Means untuk mengatasi masalah tersebut. Kemudian dihasilkan percobaan menunjukkan bahwa algoritma kebisingan dapat menentukan jumlah kluster dan menginisialisasi kluster. Algoritma pengelompokan noise K-Means diusulkan menunjukkan kinerja yang lebih baik dan menghasilkan pengelompokan yang lebih akurat [22].

Penelitian berikutnya kurangnya pengelompokan daerah yang berpotensi menghasilkan jagung di Lampung Selatan. Setelah diterapkannya metode K-Means klustering dari data tanaman jagung, dihasilkan bahwa Kecamatan Penegahan dengan *centroid* 2 sebesar 79448.30257 memiliki jumlah panen terbanyak. Sedangkan Candipuro adalah wilayah dengan jumlah panen jagung terendah dengan *centroid* 2 sebesar 1424.036868 [23].

Penelitian selanjutnya kurangnya pengelompokan provinsi COVID-19. Maka diperlukannya algoritma K-Means yang menghasilkan 3 kluster provinsi, seperti kluster 1 (Jawa Barat, Jawa Timur, Sulawesi Selatan, dan Jawa Tengah), kluster 2 (Bali, Papua, Kalimantan Selatan, Sumatera Selatan, Sumatera Utara, Kepulauan Riau, Sumatera Barat, Kalimantan Utara, Daerah Istimewa Yogyakarta, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Tengah, Sulawesi Tenggara, Riau, Sulawesi Tengah, Lampung, Kalimantan Barat, Sulawesi Utara, Maluku, Jambi, Aceh, Kepulauan Bangka Belitung, Sulawesi Barat, Papua Barat, Bengkulu, Gorontalo, Maluku Utara, Nusa Tenggara Timur), dan kluster 3 (DKI Jakarta) [24].

Penelitian berikutnya dilakukan penerapan algoritma genetika, karena dalam penjadwalan mata pelajaran yang masih dilakukan secara manual. Sehingga menghasilkan penjadwalan mengajar guru menjadi solusi yang sangat optimal di SMK Negeri 1 Tambolaka, yang memungkinkan kesamaan waktu dan jam mengajar guru yang sama sangat kecil [25].

Penelitian selanjutnya dikarenakan pendekatan analisis yang berusaha memadukan material yang akan memungkinkan akan membebani para insinyur desain. Oleh karena itu, Algoritma Genetika ini dapat merekomendasikan untuk membantu meningkatkan dalam memecahkan masalah optimasi komposit di masa depan [26].

Penelitian berikutnya dikarenakan kurangnya identifikasi pola dalam data hasil pembelajaran, membagi siswa ke dalam kelompok berdasarkan tingkat pencapaian. Hasilnya pengujian *Silhouette score* menunjukkan untuk klaster 2 yang paling tinggi sebesar 0,9168 dan menunjukkan bahwa pengelompokan data menjadi kelompok lebih baik, penggunaan *silhouette score* sebagai metrik evaluasi memberikan panduan yang berguna dalam menentukan jumlah klaster yang optimal dalam analisis *clustering* dan interpretasi data[27].

Tabel 2. 1 Kajian Pustaka

No.	Peneliti, Tahun	Judul	Metode	Masalah	Hasil
1.	Taslim, Dafwen Toresa, Deny Jollyta, Des Suryani, dan Eka Sabna, 2021	Optimasi K-Means dengan Algoritma Genetika untuk Target Pemanfaatan Air Bersih	K-Means dan Algoritma Genetika	Kurang optimalnya hasil klasterisasi dari K-Means tanpa algoritma genetika apabila dilihat dari akurasi	Penelitian ini mencapai hasil dengan mengukur validitas klaster menggunakan Davies Bouldin Index (DBI). Tanpa melakukan optimasi, klasterisasi menghasilkan DBI sekitar 2.164763. Namun, setelah menerapkan optimasi genetika pada nilai <i>centroid</i> , nilai DBI berhasil diperbaiki menjadi 2.06894.
2.	Mursalim, Purwanto, M Arief Soeleman, 2021	Penentuan <i>Centroid</i> Awal pada Algoritma K-Means dengan <i>Dynamic Artificial Chromosomes Genetic Algorithm</i> untuk <i>Tuberculosis Dataset</i>	K-Means dan <i>Dynamic Artificial Chromosomes Genetic Algorithm</i>	Ketidakoptimalan inisiasi nilai awal <i>centroid</i> yang bersifat acak dalam algoritma K-Means dapat mengakibatkan sensitivitas	Penelitian ini menunjukkan bahwa metode DAC GA + K-Means secara signifikan lebih baik dibandingkan dengan K-Means dan GA + K- Means dengan 3 klaster. Metode ini mencapai nilai DBI sebesar 0.138, 0.279, dan 0.382, dengan nilai <i>Sum Square Error</i> masing-masing adalah 92.56, 332.39, dan 1280.68. Selain itu, nilai <i>fitness</i>

No.	Peneliti, Tahun	Judul	Metode	Masalah	Hasil
				terhadap hasil akhir pengklasteran.	metode tersebut juga mencapai 7.12, 3.57, dan 2.13.
3.	Yusuf Afandi, Imam Cholissodin, dan Bayu Rahayudi, 2021	Optimasi Multiple Travelling Salesmen Problem Distribusi Produk PT Indomarco Adi Prima (Stock Point Nganjuk) dengan menggunakan Algoritma K-Means dan Algoritma Genetika (GKA)	K-Means dan Algoritma Genetika	Peningkatan efisiensi rute distribusi produk yang melibatkan sejumlah titik distribusi yang perlu dikunjungi.	Penelitian ini berhasil mengidentifikasi rute paling optimal dengan total jarak tempuh sejauh 259,722337 km, dan mencapai nilai <i>fitness</i> sebesar 0,385026. Pengaturan parameter yang optimal terjadi ketika ukuran populasi adalah 1600, jumlah generasi mencapai 800, dan menggunakan kombinasi nilai <i>crossover rate</i> (cr) sebesar 0,4, serta <i>mutation rate</i> (mr) sebesar 0,6.
4.	Endah Utik Wahyuningtyas, Rekyan Regasari Mardi Putri, Sutrisno, 2018	Optimasi K-Means Untuk Klastering Dosen Berdasarkan Kinerja Akademik Menggunakan	Algoritma Genetika Paralel dan K-Means	Ketidakoptimalan dalam menentukan <i>centroid</i> pada metode K-Means dalam	Klastering dalam penelitian tersebut menunjukkan bahwa penggunaan algoritma genetika menghasilkan kualitas yang lebih unggul dibandingkan dengan penggunaan K-Means secara mandiri. Algoritma genetika mencapai tingkat

No.	Peneliti, Tahun	Judul	Metode	Masalah	Hasil
		Algoritma Genetika Paralel		pengelompokan kinerja akademik dosen.	keberhasilan sebesar 72,145%, sementara pengklasteran dengan menggunakan K-Means hanya mencapai 68,03%.
5.	Xiaojuan Ran, Xiangbing Zhou, Mu Lei, Worawit Tepsan, and Wu Deng, 2021	<i>A Novel K-Means Clustering Algorithm with a Noise Algorithm for Capturing Urban Hotspots</i>	Fuzzy C-means, K-Means, dan K-Means plus	Sulit untuk menentukan jumlah kluster dan menginisialisasi kluster pusat secara sensitive.	Hasil percobaan menunjukkan bahwa algoritma kebisingan dapat menentukan jumlah kluster dan menginisialisasi kluster. Algoritma pengelompokan noise K-Means diusulkan menunjukkan kinerja yang lebih baik dan menghasilkan pengelompokan yang lebih akurat.
6.	A A Aldino, D Darwis, A T Prastowo, dan C Sujana, 2021	Implementation of K-Means Algorithm for Clustering Corn Planting Feasibility Area in South Lampung	K-Means	Kurangnya pengelompokan daerah yang berpotensi menghasilkan jagung di Lampung Selatan	Setelah diterapkannya metode K-Means klustering dari data tanaman jagung, dihasilkan bahwa Kecamatan Penegahan dengan <i>centroid 2</i> sebesar 79448.30257 memiliki jumlah panen terbanyak. Sedangkan Candipuro adalah wilayah dengan jumlah panen jagung terendah dengan <i>centroid 2</i> sebesar 1424.036868.

No.	Peneliti, Tahun	Judul	Metode	Masalah	Hasil
7.	Dahlan Abdullah, S. Susilo, Ansari Saleh Ahmar, R. Rusli, dan Rahmat Hidayat, 2021	The application of K-Means klastering for province klastering in Indonesia of the risk of the COVID-19 pandemic based on COVID-19 data	K-Means	Kurangnya pengelompokkan provinsi COVID-19	Berdasarkan hasil penelitian terdapat 3 klaster provinsi, seperti klaster 1 (Jawa Barat, Jawa Timur, Sulawesi Selatan, dan Jawa Tengah), klaster 2 (Bali, Papua, Kalimantan Selatan, Sumatera Selatan, Sumatera Utara, Kepulauan Riau, Sumatera Barat, Kalimantan Utara, Daerah Istimewa Yogyakarta, Nusa Tenggara Barat, Kalimantan Timur, Kalimantan Tengah, Sulawesi Tenggara, Riau, Sulawesi Tengah, Lampung, Kalimantan Barat, Sulawesi Utara, Maluku, Jambi, Aceh, Kepulauan Bangka Belitung, Sulawesi Barat, Papua Barat, Bengkulu, Gorontalo, Maluku Utara, Nusa Tenggara Timur), dan klaster 3 (DKI Jakarta).
8.	Emirensiana Padaka, Yulius Nahak Tetik, dan	Penerapan Algoritma Genetika untuk Penjadwalan Mata	Algoritma Genetika	Penjadwalan mata pelajaran yang	Penerapan algoritma genetika dalam penjadwalan mengajar menjadi solusi yang sangat optimal di SMK Negeri 1 Tambolaka, yang memungkinkan

No.	Peneliti, Tahun	Judul	Metode	Masalah	Hasil
	Dian Fransiska Ledi, 2023	Pelajaran di SMK Negeri 1 Kota Tambolaka		dilakukan secara manual	kesamaan waktu dan jam mengajar guru yang sama sangat kecil.
9.	ZhenZhou Wang, dan Adam Sobey 2020	A comparative review between Genetic Algorithm use in composite optimisation and state-of-the-art in evolutionary computation	Genetic Algorithm	Kompleksitas tambahan yang membedakan para insinyur desain	Algoritma Genetika ini membantu meningkatkan dalam memecahkan masalah optimasi komposit di masa depan.
10.	Nirwana Hendrastuty, 2024	Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Dalam Evaluasi Hasil Pembelajaran Siswa	K-Means	Kurangnya identifikasi pola dalam data hasil pembelajaran, membagi siswa ke dalam kelompok berdasarkan tingkat pencapaian	Hasilnya pengujian <i>Silhouette score</i> menunjukkan untuk kluster 2 yang paling tinggi sebesar 0,9168 dan menunjukkan bahwa pengelompokan data menjadi kelompok lebih baik, penggunaan <i>silhouette score</i> sebagai metrik evaluasi memberikan panduan yang berguna dalam menentukan jumlah kluster yang optimal dalam analisis <i>clustering</i> dan interpretasi data.

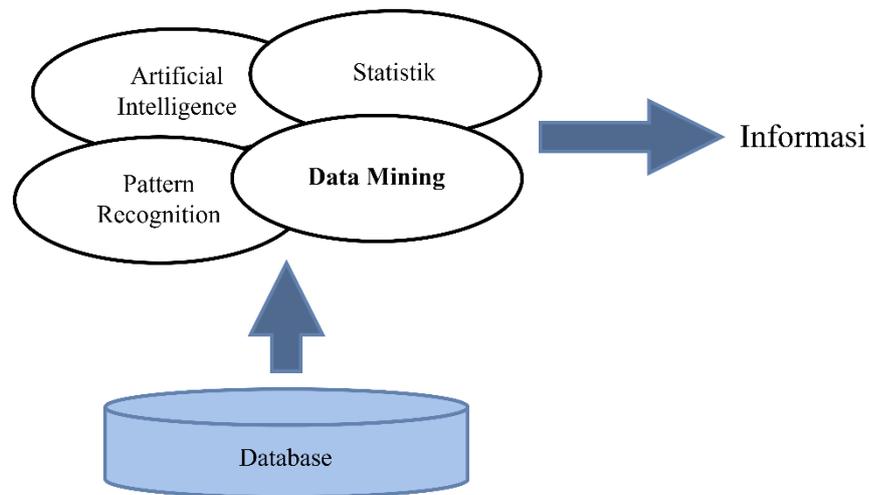
2.2 Dasar Teori

2.2.1 LINE Webtoon

LINE Webtoon adalah suatu platform komik digital yang berasal dari Korea Selatan, diperkenalkan oleh LINE Corporation dan Naver Corporation, dua perusahaan teknologi ternama. Platform ini dirancang untuk memberikan akses luas kepada pembaca untuk menikmati berbagai konten komik secara daring, menawarkan berbagai judul dari berbagai genre yang dapat diakses dengan mudah melalui aplikasi maupun situs web. LINE Webtoon memadukan kekuatan teknologi dan seni, menciptakan pengalaman membaca komik yang modern dan terhubung dengan pengguna di seluruh dunia [1]. Webtoon adalah kategori komik online atau komik digital yang menampilkan sejumlah judul yang beragam, mencakup berbagai genre seperti drama, fantasi, komedi, horor, dan slice of life. Platform ini menyajikan pengalaman membaca yang unik dengan menggunakan format yang lebih terstruktur dan mudah diakses secara daring. Dengan ragam genre yang ditawarkan, pembaca dapat menemukan berbagai cerita yang sesuai dengan preferensi mereka, mulai dari kisah-kisah penuh emosi hingga petualangan fantastis, komedi menghibur, cerita horor yang mencekam, hingga potret kehidupan sehari-hari melalui genre *slice of life*. Webtoon telah menjadi alternatif yang populer bagi penggemar komik yang ingin menikmati konten berkualitas secara digital dengan cara yang lebih modern dan mudah dijangkau [2]. Disisi lain, Webtoon juga menjadi wadah bagi para komikus lokal untuk berkarya.

2.2.2 Data Mining

Data mining merupakan suatu proses dimana data dan informasi yang signifikan yang sebelumnya tidak diketahui dapat diekstraksi dan digali dari set data yang besar. Proses ini melibatkan penemuan pola atau hubungan yang bermanfaat, yang dapat dipahami dari database yang luas. Hasil dari data mining dapat memberikan kontribusi penting dalam pembuatan keputusan bisnis [28]. Data mining juga dapat didefinisikan sebagai proses ekstraksi dan analisis pola serta pengetahuan yang menarik dari sejumlah besar data [29].



Gambar 2. 1 Diagram Data Mining

Secara garis besar, data mining adalah proses ekstraksi informasi penting dari kumpulan besar menggunakan metode kompleks seperti kecerdasan buatan, teknik statistik, matematika, dan pembelajaran mesin. Teknik ini membantu mengidentifikasi dan mengekstraksi informasi berharga dari database, memungkinkan untuk pengambilan keputusan kritis di berbagai bidang [30].

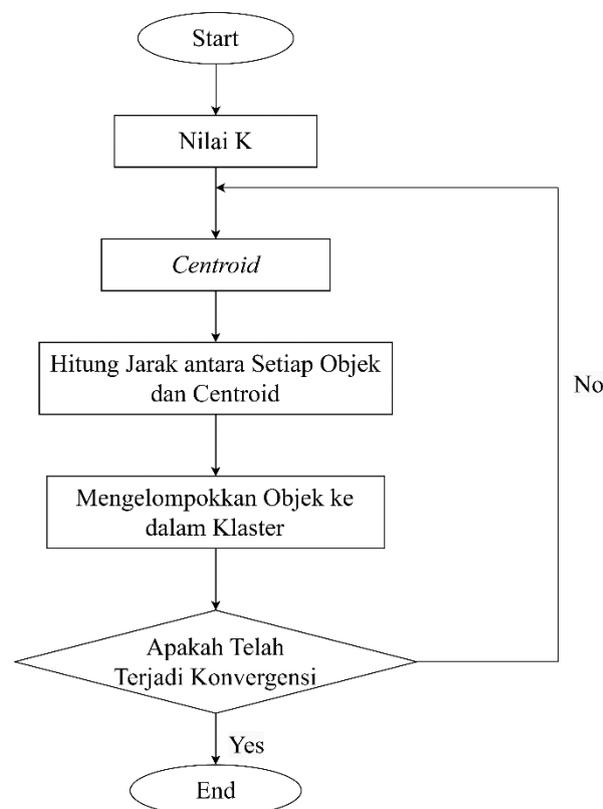
2.2.3 Clustering

Clustering adalah proses analisis data untuk mengelompokkan data ke dalam beberapa klaster berdasarkan karakteristik kesamaan. Klaster terdiri dari objek data yang mirip satu sama lain, tetapi beberapa dari objek di klaster lain. Proses ini mengidentifikasi dan membentuk kelompok berdasarkan pola atau atribut tertentu, memungkinkan pengguna memahami struktur data lebih baik dan mendapatkan wawasan bermanfaat dari informasi yang besar [31]. Metode pengelompokan, yang sering diidentifikasi sebagai teknik dalam data mining, dengan tujuan mencari dan menggabungkan data berdasarkan kesamaan karakteristik di antara satu set data dengan yang lain. Pendekatan ini bertujuan untuk mengidentifikasi pola atau hubungan yang ada di antara data, memungkinkan pembentukan kelompok yang mencerminkan kesamaan fitur [32]. Dalam konteks ini, tujuannya adalah untuk membentuk kelompok-kelompok yang menunjukkan kesamaan tertentu di antara entitas-entitas tersebut. Proses klustering membantu

mengidentifikasi pola atau struktur yang ada dalam data, memudahkan pemahaman terhadap hubungan antar objek, dan memberikan pandangan yang lebih terperinci terhadap kumpulan informasi yang diberikan [33].

2.2.4 K-Means

Metode K-Means merupakan suatu pendekatan dalam analisis data yang menerapkan sistem partisi untuk mengelompokkan data. Metode ini mendasarkan pembentukan kelompok pada pemilihan *centroid*, yaitu titik pusat representatif dalam setiap klaster. Dalam konteks ini, algoritma K-Means berfungsi dengan mendefinisikan sejumlah klaster yang diinginkan, kemudian mengelompokkan data ke dalam klaster-klaster tersebut berdasarkan kedekatan dengan *centroid* masing-masing. Pendekatan ini memberikan kerangka kerja yang efektif untuk membentuk klaster secara iteratif, memungkinkan pengguna untuk memahami struktur data dengan lebih baik dan mengidentifikasi pola-pola yang mungkin tersembunyi dalam himpunan data yang kompleks [36]. Seperti pada Gambar 2.3.



Gambar 2. 2 Diagram K-Means

1. Tentukan nilai k sebagai jumlah klaster yang dibentuk.

Menentukan jumlah klaster k , beberapa pertimbangan dapat digunakan, termasuk pertimbangan teoritis dan konseptual yang mungkin diajukan untuk menentukan berapa banyak klaster yang diperlukan.

2. Penentuan nilai centroid

Dalam menetapkan nilai awal centroid pada awal iterasi, dilakukan secara acak. Namun, pada tahap menentukan nilai *centroid* pada iterasi berikutnya, menggunakan rumus (2.1) berikut:

$$v = \frac{\sum_{i=1}^n x_i}{n}; i = 1, 2, 3, \dots, n \quad (2.1)$$

dimana :

v : centroid pada klaster

x_i : objek ke- i .

n : banyaknya objek/jumlah objek yang menjadi anggota klaster

3. Hitung jarak antara setiap objek dan centroid

K-Means juga dapat diartikan sebagai suatu algoritma untuk mengelompokkan n objek berdasarkan atributnya menjadi k partisi, dengan catatan bahwa nilai k lebih kecil dari n . maka digunakan rumus (2.2) sebagai berikut :

$$d_{(x,y)} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2}; i = 1, 2, 3, \dots, n \quad (2.2)$$

dimana :

$d_{(x,y)}$: jarak data ke i ke pusat klaster i

x_i : objek x ke- i

y_i : data y ke- i

n : banyaknya objek

4. Mengelompokkan objek-objek ke dalam klaster

Dalam tahap ini mengelompokkan sekumpulan data atau objek berdasarkan karakteristik atau fitur yang sama atau serupa. Tujuannya untuk memisahkan data ke dalam klaster, dimana objek-objek dalam satu klaster memiliki kesamaan yang lebih besar satu sama lain dibandingkan dengan objek di klaster lain.

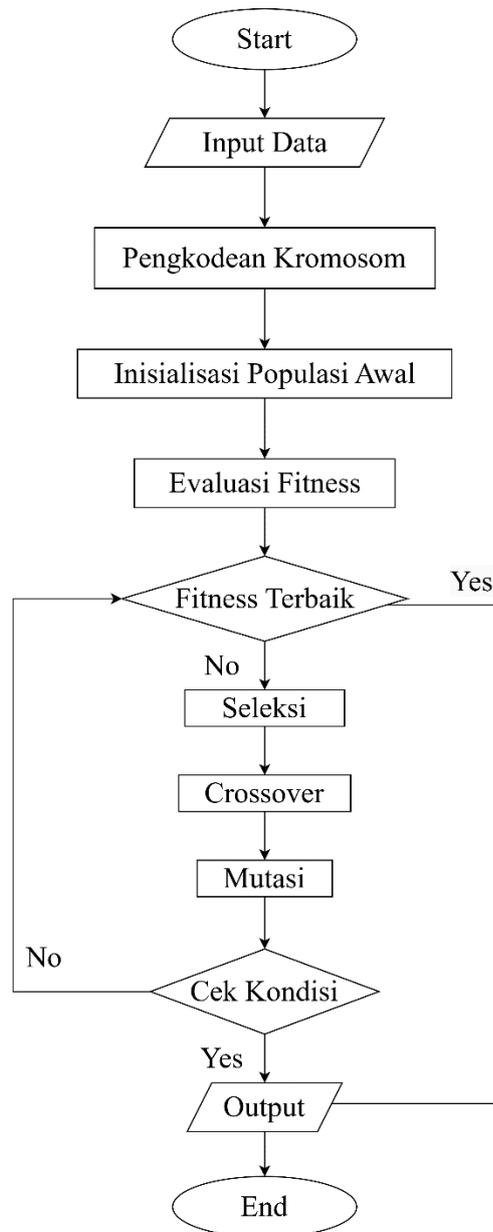
5. Apakah telah terjadi konvergensi

Dalam hal ini konvergensi mengacu pada saat ketika algoritma mencapai titik dimana tidak ada perubahan signifikan dalam solusi atau ketika mencapai kriteria berhenti yang ditetapkan sebelumnya. Ini menandakan bahwa proses iterative telah mencapai kestabilan atau solusi yang cukup optimal untuk masalah yang diberikan. Dengan kata lain, konvergensi menunjukkan bahwa algoritma telah menemukan solusi yang stabil dan cukup baik.

2.2.5 Algoritma Genetika

Algoritma genetika adalah suatu metode optimasi yang didasarkan pada prinsip seleksi alam dan genetika alami. Diperkenalkan oleh John Holland pada tahun 1970, pendekatan ini terinspirasi oleh evolusi alam untuk menciptakan teknik pencarian dan optimasi. Algoritma genetika melakukan pencarian dan optimasi melalui iterasi dan manipulasi struktur data yang meniru konsep-konsep genetika seperti seleksi alam, persilangan (*crossover*), dan mutasi. Metode ini merupakan pendekatan heuristik yang efisien untuk menemukan solusi mendekati optimal dalam berbagai masalah kompleks, terutama yang melibatkan ruang pencarian besar dan kompleksitas tinggi [37].

Algoritma genetika memberikan suatu kerangka kerja yang menyelidiki pengaruh faktor biologis seperti seleksi pasangan, reproduksi, mutase dan pertukaran informasi genetik. Algoritma ini memodelkan mekanisme evolusi alamiah melalui seleksi alam, variasi genetik, dan mutasi. Metode ini efektif untuk mengoptimalkan solusi dalam ruang solusi yang luas dan fleksibel dalam beradaptasi dengan berbagai jenis permasalahan [38]. Dalam fase perkembangannya, populasi terdiri dari sejumlah individu tertentu yang dikenal sebagai filamen atau kelompok kromosom.



Gambar 2. 3 Diagram Algoritma Genetika

Keterangan:

- a. Pengkodean Kromosom: Tahap ini mengacu pada representasi struktur solusi dalam bentuk rangkaian genetik yang bisa dioperasikan oleh algoritma. Ini melibatkan cara mengonversi elemen solusi ke dalam bentuk biner, bilangan bulat, atau simbol lainnya yang dapat dimengerti dan dimanipulasi oleh algoritma genetika. Dengan pengkodean yang tepat, algoritma dapat memanipulasi dan memodifikasi kromosom untuk mencari solusi optimal.

- b. Inisialisasi Populasi Awal: Tahap ini dimana populasi awal yang beragam dapat meningkatkan peluang menemukan solusi optimal melalui evolusi genetika. Dengan cara ini, algoritma dapat memulai proses pencarian solusi dari titik awal yang beragam dan terdapat berbagai kemungkinan solusi yang mungkin.
- c. Evaluasi Fitness : langkah ini menilai seberapa baik atau buruk kromosom dalam populasi berkinerja dalam menyelesaikan masalah yang dihadapi. Ini dilakukan dengan menghitung nilai fitness untuk setiap kromosom berdasarkan seberapa baik kromosom tersebut memenuhi kriteria atau tujuan yang ditetapkan. Dengan demikian, evaluasi fitness membantu algoritma genetika untuk memiliki kromosom yang lebih baik, yang nantinya diambil nilai fitness yang terbaik. Apabila tidak maka akan dilanjutkan pada tahap selanjutnya.
- d. Seleksi : Langkah ini ditujukan untuk memilih individu terbaik dari populasi untuk menjadi orang tua (*parents*) berdasarkan nilai kecocokan.
- e. *Crossover* / Pindah Silang : Tahap ini menggabungkan 2 orang tua untuk menghasilkan 1 atau lebih keturunan (*offspring*).
- f. Mutasi : Proses mutasi bertujuan untuk menggantikan gen yang hilang akibat proses seleksi. Mutasi gen dilakukan dengan cara menghasilkan nilai acak untuk masing-masing gen. Rentang nilai untuk proses mutasi berkisar antara 0 dan 1, di mana nilai 0 dapat mengalami mutasi menjadi 1, dan nilai 1 dapat mengalami mutasi menjadi 0. Setelah itu akan dilakukan cek kondisi apabila tidak memenuhi kembali pada tahap pemilihan fitness terbaik, jika memenuhi akan dilanjutkan pada tahap selanjutnya.
- g. Output : *Outcome* dari proses mutasi akan menjadi populasi awal untuk generasi berikutnya. Populasi yang baru ini akan menjalani proses iteratif kembali, serupa dengan langkah awal pada populasi sebelumnya. Iterasi ini berlanjut hingga mencapai kesamaan dalam fitness di antara semua individu.

2.2.6 *Silhouette Score*

Silhouette merupakan proses evaluasi yang digunakan untuk mengukur seberapa baik suatu algoritma klustering mampu memisahkan data menjadi kelompok yang berbeda. Nilai *silhouette score* yang tinggi menunjukkan bahwa

klustering baik, sementara nilai yang rendah menunjukkan bahwa ada persebaran yang tidak homogen dalam satu atau lebih kelompok [27]. Untuk titik data $i \in C_i$ (titik data i dalam kluster C_i , biarkan jarak rata-rata antara i dan semua titik data lain dalam kluster yang sama, dimana $|C_i|$ adalah jumlah titik yang termasuk dalam kluster i ,

a_i = jarak rata-rata dari i ke titik lain dalam kluster yang sama

$$a_i = \frac{1}{|C_i|-1} \sum_{j \in C_i, i \neq j} d(i, j) \quad (2.3)$$

Kemudian mendefinisikan ketidaksamaan rata-rata dari titik i ke beberapa kluster C_i sebagai rata-rata jarak dari i ke semua titik di C_i , yaitu:

b_i = jarak rata-rata ke kluster lain terdekat

$$b_i = \min_{k \neq i} \frac{1}{|C_k|} \sum_{j \in C_k} d(i, j) \quad (2.4)$$

Sekarang, definisikan *Silhouette* (nilai) dari satu titik data i :

$s(i) = b(i) - a(i) / \max\{a(i), b(i)\}$ [39]

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.5)$$

2.2.7 Davies Bouldin Index

Perhitungan DBI diperoleh dari nilai keterikatan data di dalam setiap kluster (*SSW*) dan jarak antar centroid kluster (*SSB*). Rumus digunakan untuk menentukan nilai *inter-cluster*:

$$SSW_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (2.6)$$

Keterangan:

SSW : *sum of square within cluster*

m_i : Jumlah data dalam kluster ke- i

c_i : *Centroid* kluster ke- i

$d(x_j, c_i)$: Jarak dari data ke- i ke titik kluster i

Berikutnya, dilakukan langkah selanjutnya menggunakan untuk mengetahui separasi antar kluster.

$$SSB_{ij} = d(i, j) \quad (2.7)$$

Keterangan:

SSB : *sum of square between* klaster

$d(i, j)$: Jarak *Euclidiene distance* data ke- i dan data ke- j

Selanjutnya, yaitu mencari rasio dari hasil perhitungan SSW dan SSB seperti pada rumus (2.8):

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (2.8)$$

Keterangan:

SSW : *sum of square within cluster*

SSB : *sum of square between* klaster

R_{ij} : rasio kesamaan antara klaster i dan klaster j

Untuk proses perhitungan tahap akhir rumus (2.9) digunakan untuk memperoleh nilai DBI.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{ij}) \quad (2.9)$$

Keterangan:

k : Jumlah klaster yang dihitung

R_{ij} : Rasio kesamaan antara klaster i dan klaster j

i dan j : Indeks klaster