

BAB 2 TINJAUAN PUSTAKA

2.1. DATA MINING

Data *mining* adalah suatu metode pengolahan data untuk menemukan pola yang tersembunyi dari suatu data. Hasil dari pengolahan data dengan metode data *mining* ini dapat digunakan untuk mengambil keputusan di masa depan (Prayoga, Tambunan & Parlina 2019). Data *Mining* juga merupakan suatu proses untuk menggali pola-pola dari data. Pola-pola itu didapatkan dari berbagai jenis basis data seperti basis data relasional, data warehouse, data transaksi, dan data berorientasi objek (Normah, Nurajizah & Salbinda 2021).

Data *mining* merupakan metode, teknik, *artificial intelligence* dan mesin pembelajaran yang diekstraksi sehingga menghasilkan suatu pengetahuan dan informasi yang berguna dan tersimpan dalam suatu database besar. Pada prinsipnya data mining mewarisi banyak aspek dan teknik bidang-bidang ilmu. Data mining bukanlah sesuatu hal yang baru, karena data *mining* merupakan akar dari berbagai bidang ilmu tersebut (Nasir 2021).

Didalam *data mining* terdapat beberapa teknik yang dapat digunakan yaitu sebagai berikut (Tanty, Ginting & Simanjuntak 2021) :

1. Deskripsi (*Description*)

Penelitian analisis secara sederhana ingin mencoba mencari cara deskripsi (*Description*) untuk menggambarkan pola dan kecenderungan yang terdapat dalam data. Deskripsi dari pola dan kecenderungan sering memberikan kemungkinan penjelasan untuk suatu pola atau kecenderungan.

2. Estimasi (*Estimation*)

Estimasi hampir sama dengan klasifikasi, kecuali variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan *record* lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi.

3. Prediksi (*Prediction*)

Prediksi hampir sama dengan klasifikasi dan estimasi, kecuali bahwa dalam prediksi nilai dari hasil yang akan ada di masa datang.

4. Klasifikasi (*Classification*)

Dalam klasifikasi, terdapat target variabel kategori. Sebagai contoh, penggolongan pendapatan dapat dipisahkan dalam tiga kategori, yaitu pendapatan tinggi, pendapatan sedang dan pendapatan rendah. Berdasarkan atas nilai – nilai suatu variabel yang ada untuk pengklasifikasian suatu data.

5. Pengclusteran (*Clustering*)

Pengclusteran merupakan pengelompok *record*, pengamatan atau memperhatikan dan membentuk kelas objek-objek yang memiliki kemiripan. *Cluster* adalah kumpulan record yang memiliki kemiripan satu dengan yang lainnya dan memiliki ketidak miripan dengan record - record dalam *cluster* lain. Pengclusteran berbeda dengan klasifikasi yaitu tidak adanya variabel target dalam pengclusteran. Pengclusteran tidak mencoba untuk melakukan klasifikasi, mengestimasi, atau memprediksi nilai dari variabel target.

6. Asosiasi (*Assosiation*)

Tugas asosiasi dalam *data mining* adalah menemukan atribut yang muncul dalam satu waktu. Dalam dunia bisnis lebih umum disebut analisis keranjang belanja.

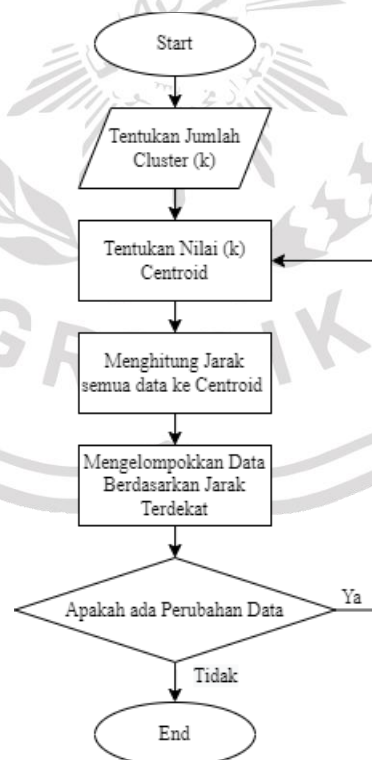
2.2. CLUSTERING

Clustering adalah salah satu teknik dari *data mining*, algoritme *cluster* yang digunakan dalam pengelompokan sejumlah data ke dalam kelompok-kelompok data tertentu (*cluster*) (Suhartini, Kerta Wijaya & Arini Pratiwi 2020). Sehingga kumpulan data yang berada di kelompok atau *cluster* yang tertentu tersebut memiliki sifat atau karakteristik yang sama (Syahputra et al. 2022). *Clustering* dapat dibedakan menjadi dua tujuan, yaitu *Clustering* untuk pemahaman dan *Clustering* untuk penggunaan. Jika tujuan untuk pemahaman maka *cluster* yang terbentuk harus menangkap struktur alami data. Biasanya proses Clustering dalam

tujuan ini hanya sebagai proses awal untuk kemudian dilanjutkan dengan pekerjaan ini seperti *summarization* (rata-rata, standar deviasi), pelabelan kelas pada setiap kelompok untuk kemudian digunakan sebagai data latih klasifikasi, dan sebagainya (Tanty, Ginting & Simanjuntak 2021).

2.3. ALGORITME *K-MEANS*

Algoritme *K-Means* merupakan salah satu algoritme pengelompokan (*clustering*) berbasis metode *non-hierarchy* yang mempartisi data dan membentuk satu atau lebih kelompok yang memiliki kesamaan dan bersifat *unsupervised* (tanpa arahan), yaitu tidak mempunyai anggota *cluster* yang pasti (Haviluddin et al. 2021). Namun, terkadang peneliti menentukan jumlah *cluster* berdasarkan pengalaman atau melalui metode tertentu, sehingga *K-Means Clustering* disebut juga sebagai *semi-supervised classification* (Qomariasih 2021). Algoritme *K-Means* bertujuan untuk menjadikan nilai rata-rata dari suatu klaster sebagai *centroid* dari *cluster* (*k*) tersebut (Wahyu & Rushendra 2022).



Gambar 2.1. Flowchart Algoritme *K-Means*

Gambar 2.1 menjelaskan tahapan algoritme *K-Means* dimana tahap pertama adalah dengan menentukan banyaknya *cluster* (k) yang akan digunakan. Setelah itu menentukan *centroid* awal secara acak yang diambil dari *dataset*. Langkah selanjutnya adalah menghitung jarak setiap data ke masing-masing titik *centroid* dengan menggunakan *Euclidean Distance*. Langkah selanjutnya adalah mengelompokkan data tersebut berdasarkan jarak terdekat antara data dengan *centroid*. Kemudian menentukan nilai *centroid* baru dengan menghitung rata-rata dari *cluster*. Setelah itu kita ulangi tahapan menghitung jarak menggunakan teori *Euclidean Distance* hingga tahapan menentukan nilai *cluster* baru. Tahap pengulangan ini akan dihentikan apabila sudah tidak ada perubahan data dalam suatu *cluster*. Untuk menentukan nilai *centroid* baru dapat dilakukan dengan menggunakan persamaan 2.1.

$$C_k = \frac{1}{n_k} \sum_{i=1}^n d_i \quad (2.1)$$

Keterangan :

n_k = jumlah semua data yang ada didalam cluster (k).

d_i = jumlah dari setiap *cluster*

2.4. ALGORITME K-MEANS++

Perbedaan antara *K-Means* dan *K-Means++* terletak pada pemilihan centroid awal. Pada *K-Means* pemilihan centroid dilakukan secara acak sebanyak nilai K yang ditentukan, namun pada *K-Means++* dilakukan pemilihan 1 centroid secara acak kemudian dilakukan perhitungan jarak *centroid* dengan data lainnya. Nilai titik terbaik adalah titik yang memiliki jarak terjauh dan probabilitas tertinggi untuk menjadi centroid baru, proses akan dilakukan sebanyak nilai *cluster*(k) yang ditentukan (Chaerudin, Murdiansyah & Imrona 2021).

Metode *K-Means++* memastikan bahwa penentuan titik pusat awal dilakukan lebih cerdas untuk meningkatkan kualitas pengelompokan (Sri Fastaf & Yamasari 2022). Langkah-langkah metode *K-Means++* sebagai berikut.

1. Menentukan titik centroid secara acak

2. Menghitung jarak semua titik dalam *dataset* dari *centroid* yang dipilih. Jarak titik x_i dari *centroid* terjauh dapat dihitung menggunakan persamaan 2.2.

$$d_i = \max_{(j:1 \rightarrow m)} \|x_i - C_j\|^2 \quad (2.2)$$

3. Menjadikan titik x_i sebagai *centroid* baru yang memiliki probabilitas maksimum.
4. Ulangi langkah 2-3 sampai titik *centroid* pada setiap *cluster*(k) terpenuhi.
5. Setelah mendapatkan *centroid*, langkah selanjutnya sama dengan metode *K-Means*

2.5. EUCLIDEAN DISTANCE

Euclidean distance adalah perhitungan jarak dari 2 buah titik dalam *Euclidean space*. *Euclidean space* diperkenalkan oleh Euclid, seorang matematikawan dari Yunani sekitar tahun 300 B.C.E. untuk mempelajari hubungan antara sudut dan jarak. *Euclidean* ini berkaitan dengan *Teorema Phytagoras* dan biasanya diterapkan pada 1, 2 dan 3 dimensi. Tapi juga sederhana jika diterapkan pada dimensi yang lebih tinggi (Abd Fauzan 2019). Teori *Euclidean distance* dapat dilihat pada persamaan 2.3.

$$d(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.3)$$

Keterangan :

- $D(x_i, y_i)$ = jarak antara clustering x dengan titik centroid y pada data ke- i
 x_i = bobot ke- i pada cluster yang ingin dicari jaraknya
 y_i = bobot data ke- i pada titik centroid
 n = merupakan jumlah data

2.6. DAVIES BOULDIN INDEX (DBI)

Davies bouldin index (DBI) adalah *metric* untuk mengevaluasi atau mempertimbangkan hasil algoritma *clustering*. Pertama kali diperkenalkan oleh David L. Davies dan Donald W. Bouldin pada tahun 1979. Dengan menggunakan DBI suatu *cluster* akan dianggap memiliki skema *clustering* yang optimal adalah yang memiliki DBI minimal (Butsianto & Saepudin 2020).

Langkah-langkah perhitungan Davies Bouldin Index adalah sebagai berikut.

1. Sum of Square Within-Cluster (SSW)

Untuk mengetahui kohesi dalam sebuah cluster ke- i salah satunya adalah dengan menghitung nilai dari *Sum of Square Within-Cluster (SSW)* seperti pada persamaan 2.4.

$$SSW_i = \frac{1}{m_i} \sum_{j=i}^{c_i} d(X_j, C_j) \quad (2.4)$$

Dimana :

m_i = jumlah data dalam *cluster* ke- i

c_i = *centroid cluster* ke- i

$d(X_j, C_j)$ = jarak setiap data ke *centroid* i yang dihitung menggunakan *Euclidean Distance*.

2. Sum of Square Between-Cluster (SSB)

Perhitungan *Sum Of Square Between-Cluster (SSB)* bertujuan untuk mengetahui separasi atau jarak antar *cluster*. dengan rumus perhitungan seperti pada persamaan 2.5.

$$SSB_{ij} = d(X_i, X_j) \quad (2.5)$$

Dimana :

$d(X_i, X_j)$ = jarak antara data ke- i dengan data ke- j di cluster lain.

3. *Ratio* (Rasio)

Perhitungan rasio ($R_{i,j}$) ini bertujuan untuk mengetahui nilai perbandingan antara *cluster* ke- i dan *cluster* ke- j untuk menghitung nilai rasio yang dimiliki oleh masing-masing *cluster*. indeks i dan j merupakan merepresentasikan jumlah *cluster*, dimana jika terdapat 4 *cluster* maka terdapat indeks sebanyak 4 yaitu i, j, k dan l . untuk menentukan nilai rasio seperti pada persamaan 2.6.

$$R_{ij,\dots,n} = \frac{SSW_i + SSW_j + \dots + SSW_n}{SSB_i + \dots + SSB_{ni,nj}} \quad (2.6)$$

Dimana :

SSW_i = *Sum of Square Within-Cluster* pada *centroid* i

SSB_i = *Sum of Square Between Cluster* data ke- i dengan j pada *cluster* yang berbeda

4. *Davies Bouldin Index* (DBI)

Nilai rasio yang diperoleh dari persamaan 2.6 digunakan untuk mencari nilai *DBI* dengan menggunakan persamaan 2.7.

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j,\dots,k}) \quad (2.7)$$

Dimana, $(R_{i,j,\dots,k})$ merupakan *ratio* dari nilai *SSW* dan *SSB* melalui persamaan 2.6 dari persamaan 2.7 maka dapat diketahui k adalah jumlah *cluster*. Dari perhitungan *Davies Bouldin Index* (*DBI*) dapat disimpulkan bahwa jika semakin kecil nilai *Davies Bouldin Index* (*DBI*) yang diperoleh (*non negatif* ≥ 0) maka *cluster* tersebut semakin baik.

2.7. *SILHOUETTE COEFFICIENT*

Metode pengujian ini menggabungkan antara metode *cohesion* dan *separation*. Range nilai yang dihasilkan oleh metode pengujian ini yakni antara -1

sampai 1. Jika nilai yang dihasilkan metode tersebut mendekati angka 1 maka dapat disimpulkan bahwa pengelompokan cluster tersebut semakin baik (Sri Fastaf & Yamasari 2022). Adapun langkah-langkah untuk menghitung nilai *Silhouette Coefficient* sebagai berikut.

1. Menghitung rata-rata jarak suatu data dengan data lain dalam suatu *cluster* yang sama menggunakan persamaan 2.8.

$$a(i) = \frac{1}{|A| - 1} \sum_{j \in A, j \neq i} d(i, j) \quad (2.8)$$

Keterangan :

$a(i)$ = Perbedaan rata-rata pada data (i) ke semua data lain di *cluster A*

$d(i, j)$ = Jarak antara data i dan data j

2. Menghitung rata-rata jarak data tersebut dengan semua data di *cluster* lain menggunakan persamaan 2.9.

$$d(i, C) = \frac{1}{|C|} \sum_{j \in C} d(i, j) \quad (2.9)$$

Keterangan :

$d(i, C)$ = Perbedaan rata-rata pada data (i) ke seluruh data C

C = Cluster lain selain A

3. Memilih nilai jarak yang paling minimum menggunakan persamaan 2.10.

$$b(i) = \min_{c \neq A} d(i, j) \quad (2.10)$$

Keterangan :

$b(i)$ = Nilai minimum jarak rata-rata data ke- i dengan semua data di *cluster* berbeda.

4. Menghitung nilai *Silhouette Coefficient* dengan menggunakan persamaan 2.11.

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.11)$$

Keterangan :

$a(i)$ = Perbedaan rata-rata pada data (i) ke semua data lain di *cluster A*

$b(i)$ = Perbedaan rata-rata pada data (i) ke semua data pada *cluster* berbeda

$s(i)$ = Nilai *Silhouette Coefficient*

Range nilai $s(i)$ yakni antara -1 dan 1, interpretasi nilai tersebut yaitu:

$s(i) = -1$: data ke- i digolongkan lemah (dekat pada *cluster B* daripada *A*)

$s(i) = 0$: data ke- i berada di tengah dua kluster (*A* dan *B*)

$s(i) = 1$: data ke- i digolongkan baik

Intepretasi nilai *Silhouette Coefficient* juga ditunjukkan dalam tabel 2.1.

Tabel 2.1. Interpretasi Nilai *Silhouette Coefficient*

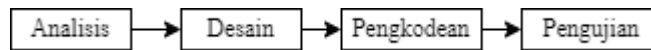
<i>Silhouette Coefficient</i>	Interpretasi
≤ 0.25	Tidak terstruktur
0.26-0.50	Hasil struktur lemah
0.51-0.70	Hasil struktur baik
0.71-1.00	Hasil struktur kuat

2.8. METODE WATERFALL

Metode *waterfall* adalah salah satu model atau pendekatan pengembangan perangkat lunak yang bersifat *linier* atau berurutan, di mana setiap tahapannya harus diselesaikan secara berurutan dan tahapan selanjutnya baru dimulai setelah tahapan sebelumnya selesai (Nur 2019).

Setiap tahap memiliki *input* dan *output* yang terdefinisi dengan jelas dan tahapannya bersifat sekuensial dan tidak berulang. Model ini sangat cocok digunakan untuk pengembangan perangkat lunak yang bersifat stabil, tidak banyak

perubahan kebutuhan, dan jelas dalam spesifikasinya. Berikut adalah gambar dari metode waterfall sendiri.



Gambar 2.2. Ilustrasi *Model Waterfall*

2.9. **BLACK BOX TESTING**

Black box testing adalah salah satu teknik pengujian perangkat lunak yang dilakukan tanpa mengetahui secara detail bagaimana kode program bekerja di dalamnya. Teknik ini melihat sistem atau program sebagai sebuah kotak hitam (*black box*) di mana input diberikan dan output yang dihasilkan dievaluasi, tanpa memperhatikan bagaimana alur program di dalamnya (Hidayat & Muttaqin 2020).

Tujuan dari *black box* testing adalah untuk memastikan bahwa sistem atau program berfungsi dengan benar sesuai dengan persyaratan fungsional dan non-fungsional yang telah ditetapkan, serta mengidentifikasi kesalahan atau kegagalan dalam program (Wijaya & Astuti 2021).

2.10. **PENYAKIT KRONIS**

Penyakit kronis adalah suatu penyakit yang diderita dalam kurun waktu lama, yaitu sekitar lebih dari enam bulan atau bahkan bertahun-tahun. Biasanya penyakit kronis tidak menimbulkan gejala pada tahap awal, tetapi gejala akan muncul ketika penyakit tersebut mulai memburuk atau semakin parah (AXA 2021).

2.11. **PENELITIAN TERKAIT**

Berikut beberapa referensi pembelajaran dari beberapa contoh kasus yang hampir sama dengan permasalahan yang dihadapi dan dijadikan penulis sebagai referensi dalam penelitian kali ini, referensi yang digunakan dapat dilihat pada tabel 2.2.

Tabel 2.2. Penelitian Terkait

Peneliti	Judul	Metode	Objek Penelitian	Hasil
(Tanty, Ginting & Simanjuntak 2021)	<i>Pengelompokan Penyakit Pada Pasien Berdasarkan Usia Dengan Metode K-Means Clustering (Studi Kasus: Puskesmas Bahorok)</i>	<i>K-Means Clustering</i>	Data pasien puskesmas Bahorok	hasil analisis dari 20 data diperoleh 3 cluster
(Adiputra 2022)	<i>Clustering Penyakit DBD Pada Rumah Sakit Dharma Kerti Menggunakan Algoritma K-Means</i>	<i>K-Means</i>	Data DBD di RS Dharma Kerti	Rumah Sakit memiliki pasien dengan diagnosa DBD yang terbagi menjadi 3 kelompok usia
(Ariyanto 2022)	<i>Data Mining Menggunakan Algoritma K-Means untuk Klasifikasi Penyakit Infeksi Saluran Pernafasan Akut</i>	<i>K-Means</i>	Data pasien gejala ISPA Klinik Rahmatan Lil Alamin	menghasilkan 3 cluster yaitu cluster C1 (ISPA Biasa), cluster C2 (ISPA Sedang), cluster C3 (ISPA Berat)

(Al-Rizki, Widaningrum & Buntoro 2020)	<i>Prediksi Penyebaran Penyakit TBC dengan Metode K-Means Clustering Menggunakan Aplikasi Rapidminer</i>	<i>K-Means Clustering</i>	Data pasien di RSUD 'Aisyiyah Ponorogo	Tingkat (pengkategorian) penyebaran penyakit TBC yang terdiri dari tinggi, dipertimbangkan, dan siaga
(Haryadi & Atmaja 2021)	<i>Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Risiko Penyakit Jantung</i>	<i>K-Means Clustering</i>	Data penyakit jantung dari website kaggle	pengelompokan tingkat resiko penyakit jantung berdasarkan 3 cluster
(Ordila et al. 2020)	<i>Penerapan Data Mining Untuk Pengelompokan Data Rekam Medis Pasien Berdasarkan Jenis Penyakit Dengan Algoritma Clustering (Studi Kasus :</i>	<i>Algoritma K-Means Clustering</i>	Data rekam medis pasien Poli Klinik PT.Inecda	Hasilnya jumlah pasien berdasarkan umur yang pertama adalah dewasa dengan jumlah pasien (4912 pasien), yang kedua adalah anak-anak (1262 pasien), dan yang

	<i>Poli Klinik Pt.Inecda)</i>			ketiga adalah balita (144) pasien
(Sari, Primajaya & Irawan 2020)	<i>Implementasi Algoritma K-Means untuk Clustering Penyebaran Tuberkulosis di Kabupaten Karawang</i>	<i>Algoritme K-Means</i>	Dataset Dinas Kesehatan Karawang tahun 2018	Hasil <i>clustering</i> terdiri dari <i>Cluster</i> 0 terdiri dari (7 kecamatan), <i>cluster</i> 1 (9 kecamatan), dan <i>cluster</i> 2 (14 kecamatan)
(Wandana, Defit & Sumijan 2020)	<i>Klasterisasi Data Rekam Medis Pasien Pengguna Layanan BPJS Kesehatan Menggunakan Metode K-Means</i>	<i>K-Means</i>	Data rekam medis pasien tahun 2019	Hasil dari perhitungan diperoleh 3 cluster. Terdiri dari cluster 0 (H0) terdapat 3 pasien yang didominasi penyakit A09.9 (Diare/Disentri), untuk cluster 1 (H1) terdapat 5 pasien dengan jenis penyakit yang lebih beragam, sedangkan untuk cluster 2 (H2) terdapat 5 pasien yang didominasi

				penyakit K30 (Dyspepsia).
(Ikhwan & Aslami 2020)	<i>Implementasi Data Mining untuk Manajemen Bantuan Sosial Menggunakan Algoritma K-Means</i>	<i>Algoritme K-Means</i>	Data kecamatan Medan Tembung dari Januari 2017	Teknik <i>K-Means</i> digunakan untuk membuat aplikasi ini, yang dapat digunakan untuk mengidentifikasi kelompok prioritas penerima hibah PKH untuk keluarga berpenghasilan rendah di Kecamatan Medan Tembung.
(Saputro & Sucihermayanti 2021)	<i>Penerapan Klasterisasi Menggunakan K-Means untuk Menentukan Tingkat Kesehatan Bayi dan Balita di Kabupaten Bengkulu Utara</i>	<i>K-Means</i>	Data diperoleh dari Dinas Kesehatan dan merupakan data pertahun sejak tahun 2019 dan 2020 dari 22 Puskesmas di Kabupaten Bengkulu Utara	Proses klasterisasi menggunakan algoritme <i>K-Means</i> menghasilkan tiga klaster tingkat kesehatan bayi dan balita di Kabupaten Bengkulu Utara yaitu rendah, sedang dan tinggi.

(Saputra Sy 2022)	<i>Klasterisasi Pasien Rawat Inap Peserta BPJS Berdasarkan Jenis Penyakit Menggunakan Algoritma K-Means</i>	<i>K-Means</i>	Data yang digunakan merupakan rekam medis pasien rawat inap rumah sakit RSUD Arifin Achmad Provinsi Riau tahun 2021	Penelitian menghasilkan pada kluster K1 didominasi dengan penyakit <i>Malignant neoplasm, breast, unspecified</i> (C50.9) dan <i>Non-hodgkin's lymphoma, unspecified type</i> (C85.9). Kluster K2 didominasi penyakit <i>Fracture of neck of femur, closed</i> (S71.00) dan <i>Dengue haemorrhagic fever</i> (A91).
(Isag & Irawan 2021)	<i>Implementasi K-means Dalam Prioritas Perbaikan Jembatan Dan Saluran Air Di Kecamatan Ngronggot</i>	<i>K-Means</i>	Data jembatan dan saluran air di Kecamatan Ngronggot	hasil perhitungan dibagi menjadi 3 (tiga) kelompok menjadi prioritas pertama sebagai perbaikan yang harus didahulukan dikarenakan hasil perhitungan berdasarkan kriteria

				menunjukkan bahwa tingkat kerusakan yang berat, prioritas kedua memiliki nilai prioritas perbaikan sedang dan prioritas ketiga memiliki nilai prioritas perbaikan yang ringan
(Rosadi et al. 2020)	<i>Penerapan K-Means Clustering Pada Pemilihan Program Prioritas Dana Desa</i>	<i>K-Means Clustering</i>	Data rancangan program kerja desa yang diajukan oleh setiap kepala dusun dan pemerintah desa tahun 2020	hasil akhir yang berupa pengelompokan program kerja desa menjadi 3 tingkat proritas, yaitu tingkat prioritas tinggi pada cluster 1, tingkat prioritas sedang pada cluster 2 dan tingkat prioritas rendah pada cluster 3
(Nugroho &	<i>Penggunaan Metode K-Means dan K-</i>	<i>K-Means dan K-Means++</i>	Data dari laman web	<i>K-Means++</i> lebih unggul dalam memberikan

Adhinata 2022)	<i>Means++ Sebagai Clustering Data Covid-19 di Pulau Jawa</i>		<i>https://m.andrafarm.com/</i>	informasi sejauh mana penyebaran virus Covid-19
(al Rivan, Gandi & Lukman 2020)	<i>Perbandingan Hybrid Genetic K-Means++ dan Hybrid Genetic K-Medoid untuk Klasterisasi Dataset EEG Eyestate</i>	<i>K- Means++, K- Medoids dan GA</i>	<i>Data EEG Eyestate yang didapat pada laman resmi UCI Machine Learning Repository (2014)</i>	Berdasarkan hasil perbandingan nilai rata-rata iterasi GA K-Means++ dan GA K-Medoids maka dapat disimpulkan bahwa GA - KMeans++ lebih baik