

BAB III

METODE PENELITIAN

3.1. Bahan Penelitian

Bahan penelitian diambil melalui media internet dan mulai diakses pada tanggal 10 April 2024 sampai dengan tanggal 10 Mei 2024 dari website resmi Kementerian Pendidikan dan Kebudayaan (KEMENDIKBUD) dengan alamat <https://dapo.dikdasmen.kemdikbud.go.id>. Data yang diambil yaitu data-data pokok pendidikan yang memiliki karakteristik serupa berdasarkan fasilitas, pendidik, tenaga pendidik di kabupaten Karawang, data yang digunakan data pada tingkat SMA dan SMK dengan tahun ajaran 2023/2024 yang akan dibagi menjadi 2 *cluster* yaitu baik dan kurang.

3.2. Peralatan Penelitian

Alat yang digunakan dalam penelitian yaitu:

A. Perangkat Keras

- 1) Personal Computer (PC)
- 2) Laptop
- 3) Printer
- 4) Flasdisk 6 GB

B. Perangkat Lunak

- 1) Sistem Operasi Windows 10
- 2) Microsoft Word
- 3) Microsoft Excel
- 4) *RapidMiner*

3.3. Lokasi dan Waktu Penelitian

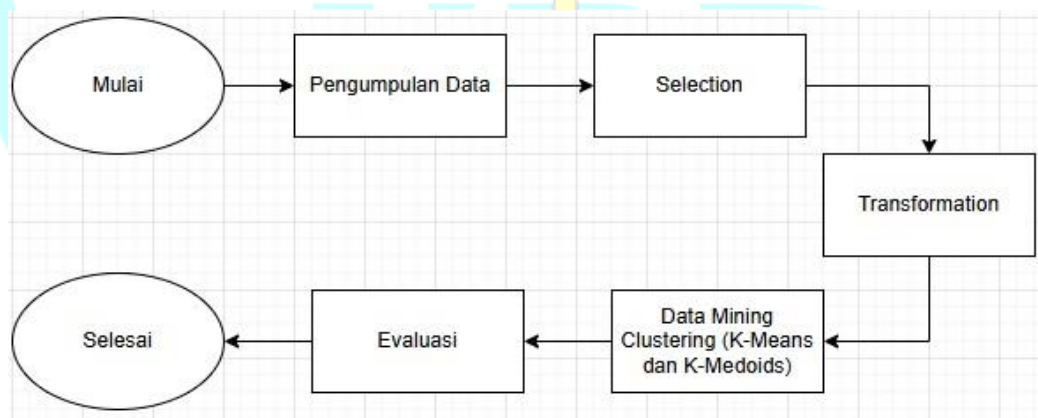
Lokasi penelitian ini tidak menunjukkan pada suatu letak lokasi tetapi merupakan studi literatur dimana data yang didapatkan berupa data-data pokok pendidikan yang diambil secara online dari website resmi KEMENDIKBUD. Adapun waktu penelitian terdapat pada table di bawah ini:

Tabel 3. 1 Waktu penelitian

Kegiatan	Bulan			
	Mei	Juni	Juli	Agustus
Pengumpulan Data				
Seleksi Data				
Data Mining				
Evaluasi				

3.4. Prosedur Penelitian

Berikut adalah prosedur percobaan dalam penelitian ini:



Gambar 3. 1 Flowchart Penelitian

3.4.1. Pengumpulan Data

Tahap pertama dari penelitian ini yaitu mengumpulkan data-data pokok terkait dengan pendidikan. Dataset yang digunakan ada beberapa atribut yaitu: data sekolah, data fasilitas, data pendidik, dan data tenaga pendidik. Data yang diambil yaitu data tahun ajaran 2023/2024 tingkat SMA/K yang berada di wilayah Karawang yang di peroleh di website <https://dapo.dikdasmen.kemdikbud.go.id>.

Nama Sekolah	SMA/K	Status	Peserta Didik	Rombel	PENDIDIK								TENAGA PENDIDIK	Fasilitas		
					Guru Matematika	Guru B. Indo	Guru B. Inggris	Guru Penjas	Guru Agama	Guru Ips	Guru Ipa	Guru Pkn		Kelas	Lab	Perpus
SMKN 1 CIKAMPEK	SMK	Negeri	1,904	54	11	11	11	1	2	25	30	1	29	52	5	3
SMAN 1 KLARI	SMA	Negeri	1,182	33	11	12	12	2	1	13	13	1	20	34	4	1
SMAN 1 CIKAMPEK	SMA	Negeri	1,182	33	10	10	10	2	1	10	12	3	11	33	5	1
SMKN PERTANIAN	SMK	Negeri	1,019	30	8	7	7	1	1	8	8	3	19	21	1	1
SMKN 1 RAWAMERTA	SMK	Negeri	1,047	30	8	7	9	1	2	8	7	2	16	16	0	1
SMAN 3 KARAWANG	SMA	Negeri	1,066	30	10	10	10	1	2	10	10	1	15	56	5	1
SMAN 2 CIKAMPEK	SMA	Negeri	1,168	33	10	10	10	2	2	9	10	1	11	28	8	1
SMAN 1 TELUKJAMBE	SMA	Negeri	1,086	31	9	10	10	2	2	8	10	1	14	31	6	1

Gambar 3. 2 Data Sekolah Kabupaten Karawang

3.4.2. Selections

Selection data dilakukan karena data yang didapatkan tidak digunakan semua dalam penelitian ini. Proses *selection* data mempunyai tujuan dalam Meng *selection* atribut data yang tidak diperlukan. Data yang diperoleh dari website resmi KEMENDIKBUD tidak langsung sempurna. Ada beberapa atribut yang tidak relevan atau tidak dibutuhkan dalam penelitian ini. Data hanya sesuai analisis yang akan digunakan seperti data fasilitas, data pendidik, dan data tenaga pendidik. Pengambilan data difokuskan hanya mengambil data sekolah tingkat SMA/K di wilayah Karawang.

Tabel 3. 1 Data Setelah di *Selection*

Nama Sekolah	Jenis Sekolah	Status	Guru Matematika	...	Perpustakaan
SMKN 1 KARAWANG	SMK	Negeri	16	...	1
SMKN 2	SMK	Negeri	15	...	1

KARAWANG		i			
SMKN 1	SMK	Neger	11	...	3
CIKAMPEK		i			
SMKN 3	SMK	Neger	10	...	1
KARAWANG		i			
⋮	⋮	⋮	⋮	⋮	⋮
SMA BAITUL	SMA	Swast	0	...	0
ULYA		a			
BOARDING					
SCHOOL					

3.4.3. Transformasi Data

Mengubah dan memperbaiki data sehingga mendapatkan data yang sesuai. Tujuan utama dari transformasi data adalah untuk meningkatkan kualitas data, memungkinkan deteksi pola atau tren yang lebih baik, serta mempersiapkan data

Nama Sekolah	Jenis Sekolah	Status	Guru Matematika	...	Perpustakaan
SMKN 1 KARAWANG	2	1	16	...	1
SMKN 2 KARAWANG	2	1	15	...	1
SMKN 1 CIKAMPEK	2	1	11	...	3
SMKN 3 KARAWANG	2	1	10	...	1
⋮	⋮	⋮	⋮	⋮	⋮
SMA BAITUL ULYA BOARDING SCHOOL	1	2	0	...	0

untuk penggunaan oleh algoritma.

3.4.4. Data Mining Klasterisasi

Penelitian ini diselesaikan dengan menggunakan dua model algoritma klasterisasi yaitu algoritma K-Means dan algoritma K-Medoids, di mana hasilnya akan dibandingkan dan akan mengetahui tingkat akurasi yang baik pada penelitian ini. Jadi semua data yang sudah melalui tahapan-tahapan sebelumnya akan

dihitung atau dianalisis dengan 2 algoritma tersebut. Sebelum pengelompokan data dalam penelitian ini ditentukan dengan 2 *Cluster* yaitu *Cluster* 1 dan *Cluster* 2. Penelitian ini menghasilkan sekelompok sekolah yang memiliki baik dan kurang dalam fasilitas, pendidik, dan tenaga pendidik. Perangkat lunak yang digunakan dalam proses data mining ini menggunakan aplikasi *RapidMiner* yang sudah di install pada computer dan berikut pseudocode algoritma.

Judul:

Pseudocode Algoritma K-Means

Input:

X, K, Max_Iterasi

Inisiasi:

Centroid = randomly select k points from x

Iterasi:

For iter from 1 to max_iters do

Labels = []

For each data point x in x do

Distance = []

For each centroid c in centroid do

Distance = calculate euclidean distance (x,c)

Append distance to distance

End for

Cluster_label = index of the minimum distance in distance

Append cluster_label to labels

End for

New_centroids = []

For each cluster_index from 0 to K-1 do

cluster_points = [x for x in X if labels[x] == cluster_index]

If cluster_points is not empty then

new_centroid = CalculateMean(cluster_points)

Append new_centroid to new_centroids

Else

End If

```

End For
If centroids == new_centroids then
Break the loop
End If
centroids = new_centroids
End For
Return (centroids, labels)
End Function

```

Output:

Centroid, Labels

Judul:

Pseudocode Algoritma K-Medoids

Input:

X, K, Max_Iterasi

Inisiasi:

Medoids = randomly select k points from x

Iterasi:

```

For iter from 1 to max_iters do
labels = []
For each data point x in X do
distances = []
For each medoid m in medoids do
distance = CalculateDistance(x, m)
Append distance to distances
End For
cluster_label = Index of the minimum distance in distances
Append cluster_label to labels
End For
new_medoids = []

```

```

For each cluster_index from 0 to K-1 do
  cluster_points = [x for x in X if labels[x] == cluster_index]
  If cluster_points is not empty then
    best_medoid = None
    best_cost = Infinity
    For each candidate_medoid in cluster_points do
      cost = Sum of distances between candidate_medoid and all points in
      cluster_points
      If cost < best_cost then
        best_cost = cost
        best_medoid = candidate_medoid
      End If
    End For
    Append best_medoid to new_medoids
  Else
    End If
End For
If medoids == new_medoids then
  Break the loop
End If
medoids = new_medoids # Perbarui medoid
End For
Return (medoids, labels)
End Function

```

Output:

Medoids, Labels

3.4.5. Evaluasi

Evaluasi klusterisasi dilakukan menggunakan metrik utama. *Davies Bouldin Index* digunakan untuk menilai rasio antara jarak intra-klaster dan antar-klaster, memberikan indikasi mengenai kualitas klusterisasi. Hasil evaluasi dari kedua algoritma dibandingkan untuk menentukan kinerja relatif mereka. Analisis

dilakukan dengan membandingkan kualitas nilai *Davies Bouldin Index* dari setiap algoritma pada dataset. Selain itu, perbandingan juga dilakukan untuk menentukan algoritma mana yang lebih baik.

