

## **BAB III**

### **METODE PENELITIAN**

#### **3.1 Populasi dan Sampel Penelitian**

##### **3.1.1 Populasi**

Populasi dalam penelitian ini adalah seluruh siswa kelas VI SD Negeri 05 Bilah Barat yang memiliki 50 siswa . mata pelajaran di kelas VI SD terdiri dari data nilai lengkap pada mata pelajaran Bahasa Indonesia, Matematika, IPA, IPS, PPKn, PJOK, Seni Budaya, Bahasa Inggris dan Agama. penelitian ini melihat data nilai lengkap pada mata pelajaran Bahasa Indonesia, Matematika, IPA, IPS, dan PPKn.

##### **3.1.2 Sampel**

Penelitian menggunakan teknik total sampling, yaitu seluruh siswa kelas VI yang memiliki data nilai lengkap yang menjadi sampel yaitu Bahasa Indonesia, Matematika, IPA, IPS, dan PPKn. Yang dijadikan sebagai sampel penelitian. Teknik ini dipilih karena jumlah siswa tidak terlalu besar sehingga seluruh data dapat dianalisis secara komprehensif.

#### **3.2 Variabel Penelitian**

Penelitian ini menggunakan lima variabel numerik yang menjadi dasar proses clustering, yaitu nilai siswa pada mata pelajaran:

1. Bahasa Indonesia
2. Matematika
3. Ilmu Pengetahuan Alam (IPA)
4. Ilmu Pengetahuan Sosial (IPS)

## 5. Pendidikan Pancasila dan Kewarganegaraan (PPKn)

Seluruh variabel diperlakukan sebagai atribut dalam proses pengelompokan menggunakan algoritma K-Means.

**Tabel 3.1 Variabel Penelitian Lengkap**

Kode	Variabel	Jenis Data	Skala	Keterangan
X1	PPKn	Numerik	Rasio	Nilai Pendidikan Pancasila
X2	Bahasa Indonesia	Numerik	Rasio	Nilai Bahasa Indonesia
X3	Matematika	Numerik	Rasio	Nilai Matematika
X4	IPA	Numerik	Rasio	Nilai Ilmu Pengetahuan Alam
X5	IPS	Numerik	Rasio	Nilai Ilmu Pengetahuan Sosial

Seluruh variabel dalam penelitian ini bersifat numerik dan digunakan sebagai atribut dalam proses perhitungan jarak Euclidean pada algoritma K-Means Clustering.

### 4.3 Langkah-Langkah Penerapan Algoritma K-Means

Algoritma K-Means digunakan untuk mengelompokkan nilai akademik siswa ke dalam beberapa kategori prestasi, yaitu sangat berprestasi, rata-rata, dan kurang berprestasi. Proses penerapan algoritma K-Means Clustering dilakukan melalui beberapa tahapan sebagai berikut:

### 3. 2 Data Nilai Siswa

No	NAMA	PKN	B.INDO	MTK	IPA	IPS
1	A***** R*****	80	75	74	75	75
2	B***** H*****	70	70	70	70	70
3	A***** R*****	76	70	70	70	70

4	F**** Y****	73	71	70	70	70
5	R**** M*****	82	80	75	75	76
6	N**** R****	85	85	75	72	73
7	Y*** S*****	70	70	70	70	70
8	M***** A*****	80	70	73	75	75
9	A***** S***	76	72	73	75	74
10	A*****	70	70	70	70	70
11	S***** R****	80	78	72	74	73
12	D*** M*****	75	72	70	70	70
13	H*** S*****	86	86	75	78	78
14	T**** R*****	72	72	70	72	72
15	W**** F*****	77	70	71	73	72
16	M***** R***	75	70	70	70	70
17	A***** H*****	74	70	70	72	70
18	S*** F*****	78	85	74	73	74
19	D*** P*****	75	72	71	73	73
20	F** A*****	75	70	70	70	70
21	M***** A***	70	70	70	72	70
22	N***** E****	72	73	72	73	72
23	R**** S*****	71	70	71	72	70
24	D**** P*****	72	72	70	70	70

### 3.3.1. Perhitungan Nilai Mean Awal

Sebelum menentukan centroid awal, dilakukan perhitungan nilai rata-rata (mean) dari masing-masing variabel menggunakan rumus:

$$Mean = \frac{\sum X}{n}$$

Dimana:

1.  $\sum X$  = total seluruh nilai
2.  $n$  = jumlah siswa (24)

### 3.3. Tabel Nilai Mean

Variabel	Mean
PPKn	76.00
Bahasa Indonesia	72.38
Matematika	70.83
IPA	72.08
IPS	71.88

Nilai mean ini menunjukkan kecenderungan umum prestasi siswa sebelum dilakukan proses pengelompokan menggunakan algoritma K-Means.

Langkah pertama dalam penerapan algoritma K-Means adalah menentukan jumlah cluster yang optimal. Berdasarkan analisis menggunakan metode Elbow dan Silhouette Score (yang dibahas lebih lanjut pada subbab 3.4), dipilih jumlah 3 cluster. Jumlah cluster ini dianggap optimal karena memberikan distribusi data yang paling representatif berdasarkan prestasi belajar siswa. Dengan 3 cluster, kita dapat mengelompokkan siswa ke dalam kategori prestasi yang sesuai: sangat berprestasi, rata-rata, dan kurang berprestasi, Penentuan nilai centroid awal dalam penelitian ini dapat dilihat pada Tabel 3.2 di bawah ini :

#### 3. 4 *Centroid Awal*

No	Nama	PKN	B.INDO	MTK	IPA	IPS	Cluster
9	A***** S****	76	72	73	75	74	C1
17	A***** H*****	74	70	70	72	70	C2

1	A***** R*****	80	75	74	75	75	C3
---	------------------	----	----	----	----	----	----

Setelah nilai centroid ditentukan, langkah selanjutnya adalah menghitung jarak setiap data terhadap masing-masing centroid menggunakan rumus Euclidean Distance untuk menentukan keanggotaan cluster. Proses perhitungan tersebut dilakukan sebagaimana berikut dan hasilnya disajikan pada Tabel 3.3:

A\*\*\*\*\* R\*\*\*\*\*

Jarak ke C1 =

$$\sqrt{(80 - 76)^2 + (75 - 72)^2 + (74 - 73)^2} = \sqrt{(75 - 75)^2 + (75 - 74)^2}$$

$$= 5.19$$

Jarak ke C2

$$\sqrt{(80 - 74)^2 + (75 - 74)^2 + (74 - 70)^2 + (75 - 72)^2 + (75 - 70)^2}$$

$$= 10.54$$

Jarak ke C3

$$\sqrt{(80 - 80)^2 + (75 - 75)^2 + (74 - 74)^2 + (75 - 75)^2 + (75 - 75)^2}$$

$$= 0$$

### 3. 5 Hasil Perhitungan Iterasi Ke – 1

No	Nama	Jarak ke C1	Jarak ke C2	Jarak ke C3	Hasil Cluster
1	A***** R*****	5,196	10,536	0,000	C3
2	B***** H*****	9,487	4,472	13,820	C2
3	A***** R*****	7,348	2,828	10,344	C2
4	F***** Y*****	7,746	2,449	11,446	C2

5	R**** M*****	10,392	15,297	5,568	C3
6	D**** H****	16,248	19,494	11,790	C3
7	D**** M*****	9,487	4,472	13,820	C2
8	D**** K*****	4,583	8,888	5,099	C1
9	E**** P*****	0,000	6,481	5,196	C1
10	E***** R*****	9,487	4,472	13,820	C2
11	F**** T*****	7,416	10,817	4,243	C3
12	H**** A*****	7,141	3,000	10,000	C2
13	H**** V*****	17,833	22,913	13,266	C3
14	I***** P*****	6,557	3,464	10,344	C2
15	I**** P***	4,690	3,873	7,483	C2
16	J***** W*****	8,000	2,236	10,770	C2
17	K**** Y*****	7,141	0,000	10,536	C2
18	K**** H*****	13,379	16,553	10,440	C3
19	L*** U*****	3,606	4,000	7,141	C1
20	M**. S*****	8,000	2,236	10,770	C2
21	N**** S*****	9,110	4,000	13,229	C2
22	N***** N*****	5,568	4,690	9,220	C2
23	R***** A*****	8,185	3,162	12,207	C2
24	R**** S****	8,660	3,464	11,790	C2

Setelah proses pada iterasi pertama selesai, tahap selanjutnya adalah melakukan iterasi kedua. Proses yang dilakukan pada iterasi kedua sama dengan tahap sebelumnya, namun nilai centroid diperbarui berdasarkan rata-rata data pada masing-masing cluster yang terbentuk. Hasil perhitungan pada iterasi kedua dapat dilihat pada Tabel 3.4 berikut :

### 3. 6 Nilai Centeroid Baru Iterasi Ke – 2

Cluster	PKN	B.INDO	MTK	IPA	IPS
C1	77.0	71.33	72.33	74.33	74.0
C2	72.8	70.67	70.27	70.93	70.4
C3	81.83	81.50	74.17	74.50	74.83

Selanjutnya, dengan menggunakan proses yang sama seperti pada iterasi pertama, diperoleh hasil pada iterasi kedua yang ditampilkan pada Tabel 3.5 berikut:

### 3. 7. Hasil Perhitungan Iterasi Ke- 2

No	Nama	Jarak ke C1	Jarak ke C2	Jarak ke C3	Hasil Cluster
1	A***** R*****	5,17	11,06	6,78	C1
2	B***** H*****	9,54	3,06	18,25	C2
3	A***** R*****	6,55	3,43	15,07	C2
4	F***** Y*****	7,50	1,12	15,79	C2
5	R***** M*****	10,57	15,56	2,14	C3
6	D***** H*****	16,26	19,61	5,71	C3

7	D**** M*****	9,54	3,06	18,25	C2
8	D**** K*****	3,56	9,87	11,72	C1
9	E**** P*****	1,53	7,00	11,25	C1
10	E***** R*****	9,54	3,06	18,25	C2
11	F**** T*****	7,40	11,17	4,89	C3
12	H**** A*****	6,68	2,78	14,07	C2
13	H**** V*****	18,24	23,22	7,79	C3
14	I***** P*****	6,35	2,49	14,78	C2
15	I**** P****	3,05	5,05	13,26	C1
16	J***** W*****	6,78	2,53	15,49	C2
17	K**** Y*****	6,13	1,81	15,51	C2
18	K**** H*****	13,87	16,23	5,47	C3
19	L*** U*****	3,00	4,26	12,35	C1
20	M**. S*****	6,78	2,53	15,49	C2
21	N**** S*****	8,81	3,11	17,87	C2
22	N***** S*****	5,80	3,99	13,56	C2
23	R***** A*****	7,81	2,35	17,00	C2

24	R*** S***	8,10	1,87	15,74	C2
----	-----------	------	------	-------	----

Setelah proses pada iterasi kedua selesai, tahap berikutnya adalah melakukan iterasi ketiga. Prosedur yang dilakukan pada iterasi ini sama dengan iterasi sebelumnya, namun nilai *centroid* diperbarui berdasarkan rata-rata data pada masing-masing cluster yang terbentuk. Hasil perhitungan pada iterasi ketiga dapat dilihat pada Tabel 3.6 berikut:

### 3. 8 Nilai Centroid Baru Iterasi Ke – 3

Cluster	PKN	B.INDO	MTK	IPA	IPS
C1	77,0	71,33	72,33	74,33	74,0
C2	72,8	70,67	70,27	70,93	70,4
C3	81,83	81,50	74,17	74,50	74,83

Selanjutnya, dengan menggunakan prosedur yang sama seperti pada iterasi kedua, diperoleh hasil pada iterasi ketiga yang disajikan pada Tabel 3.7 berikut.

### 3. 9 Hasil Perhitungan Iterasi Ke – 3

No	Nama	Jarak ke C1	Jarak ke C2	Jarak ke C3	Hasil Cluster
1	A***** R*****	5,17	11,06	6,78	
2	B***** H*****	9,54	3,06	18,25	C2
3	A***** R*****	6,55	3,43	15,07	C2
4	F***** Y*****	7,50	1,12	15,79	C2
5	R***** M*****	10,57	15,56	2,14	C3

6	D**** H****	16,26	19,61	5,71	C3
7	D**** M*****	9,54	3,06	18,25	C2
8	D**** K*****	3,56	9,87	11,72	C1
9	E**** P*****	1,53	7,00	11,25	C1
10	E***** R*****	9,54	3,06	18,25	C2
11	F**** T*****	7,40	11,17	4,89	C3
12	H**** A*****	6,68	2,78	14,07	C2
13	H**** V*****	18,24	23,22	7,79	C3
14	I***** P*****	6,35	2,49	14,78	C2
15	I**** P***	3,05	5,05	13,26	C1
16	J***** W*****	6,78	2,53	15,49	C2
17	K**** Y*****	6,13	1,81	15,51	C2
18	K**** H*****	13,87	16,23	5,47	C3
19	L*** U*****	3,00	4,26	12,35	C1
20	M**. S*****	6,78	2,53	15,49	C2
21	N**** S*****	8,81	3,11	17,87	C2
22	N*****	5,80	3,99	13,56	C2

23	R***** A*****	7,81	2,35	17,00	C2
24	R**** S****	8,10	1,87	15,74	C2

Tahap berikutnya adalah iterasi keempat, yang dilakukan dengan prosedur yang sama seperti iterasi sebelumnya, di mana nilai centroid dihitung ulang berdasarkan rata-rata tiap cluster. Hasilnya disajikan pada Tabel 3.8.

### 3. 10 Nilai Centeroid Baru Iterasi Ke- 4

Cluster	PKN	B.INDO	MTK	IPA	IPS
C1	77.0	71.33	72.33	74.33	74.0
C2	72.8	70.67	70.27	70.93	70.4
C3	81.83	81.50	74.17	74.50	74.83

Selanjutnya, dengan menggunakan prosedur yang sama seperti pada iterasi ketiga, diperoleh hasil pada iterasi keempat yang disajikan pada Tabel 3.9 berikut:

### 3. 11 Hasil Perhitungan Iterasi Ke – 4

No	Nama	Jarak ke C1	Jarak ke C2	Jarak ke C3	Hasil Cluster
1	A***** R*****	5,17	11,06	6,78	C1
2	B***** H*****	9,54	3,06	18,25	C2
3	A***** R*****	6,55	3,43	15,07	C2
4	F***** Y*****	7,50	1,12	15,79	C2
5	R***** M*****	10,57	15,56	2,14	C3

6	D**** H****	16,26	19,61	5,71	C3
7	D**** M*****	9,54	3,06	18,25	C2
8	D**** K*****	3,56	9,87	11,72	C1
9	E**** P*****	1,53	7,00	11,25	C1
10	E***** R*****	9,54	3,06	18,25	C2
11	F**** T*****	7,40	11,17	4,89	C3
12	H**** A*****	6,68	2,78	14,07	C2
13	H**** V*****	18,24	23,22	7,79	C3
14	I**** P*****	6,35	2,49	14,78	C2
15	I**** P****	3,05	5,05	13,26	C1
16	J***** W*****	6,78	2,53	15,49	C2
17	K**** Y*****	6,13	1,81	15,51	C2
18	K**** H*****	13,87	16,23	5,47	C3
19	L*** U*****	3,00	4,26	12,35	C1
20	M** . S*****	6,78	2,53	15,49	C2
21	N**** S*****	8,81	3,11	17,87	C2

22	N*****	5,80	3,99	13,56	C2
23	R***** A*****	7,81	2,35	17,00	C2
24	R*** S***	8,10	1,87	15,74	C2

Setelah proses pada iterasi keempat selesai, tahap selanjutnya adalah melakukan iterasi kelima. Prosedur yang dilakukan pada iterasi ini sama dengan iterasi sebelumnya, namun nilai centroid diperbarui berdasarkan rata-rata data pada masing-masing cluster yang terbentuk. Hasil perhitungan pada iterasi kelima disajikan pada Tabel 3.10 berikut :

### 3. 12 Nilai Centeroid Baru Iterasi Ke- 5

Cluster	PKN	B.INDO	MTK	IPA	IPS
C1	77,0	71,33	72,33	74,33	74,0
C2	72,8	70,67	70,27	70,93	70,4
C3	81,83	81,50	74,17	74,50	74,83

Selanjutnya, dengan menggunakan prosedur yang sama seperti pada iterasi sebelumnya, diperoleh hasil pada iterasi yang disajikan pada Tabel 3.11 berikut:

### 3. 13 Hasil Perhitungan Iterasi Ke – 5

No	Nama	Jarak ke C1	Jarak ke C2	Jarak ke C3	Hasil Cluster
1	A***** R*****	5,17	11,06	6,78	C1
2	B***** H*****	9,54	3,06	18,25	C2
3	A***** R*****	6,55	3,43	15,07	C2

4	F**** Y*****	7,50	1,12	15,79	C2
5	R**** M*****	10,57	15,56	2,14	C3
6	D**** H*****	16,26	19,61	5,71	C3
7	D**** M*****	9,54	3,06	18,25	C2
8	D**** K*****	3,56	9,87	11,72	C1
9	E**** P*****	1,53	7,00	11,25	C1
10	E***** R*****	9,54	3,06	18,25	C2
11	F**** T*****	7,40	11,17	4,89	C3
12	H**** A*****	6,68	2,78	14,07	C2
13	H**** V*****	18,24	23,22	7,79	C3
14	I***** P*****	6,35	2,49	14,78	C2
15	I**** P***	3,05	5,05	13,26	C1
16	J***** W*****	6,78	2,53	15,49	C2
17	K**** Y*****	6,13	1,81	15,51	C2
18	K**** H*****	13,87	16,23	5,47	C3
19	L*** U*****	3,00	4,26	12,35	C1
20	M**. S*****	6,78	2,53	15,49	C2

21	N**** S*****	8,81	3,11	17,87	C2
22	N*****	5,80	3,99	13,56	C2
23	R***** A*****	7,81	2,35	17,00	C2
24	R**** S****	8,10	1,87	15,74	C2

Karena pada iterasi ke-3 hingga iterasi ke-5 tidak terjadi perubahan nilai centroid, maka proses clustering dinyatakan konvergen. Nilai centroid akhir ditampilkan pada tabel berikut:

### 3.14 Tabel Centroid Final

Cluster	PPKn	B.Indo	MTK	IPA	IPS
C1	77.00	71.33	72.33	74.33	74.00
C2	72.80	70.67	70.27	70.93	70.40
C3	81.83	81.50	74.17	74.50	74.83

Cluster C3 memiliki nilai rata-rata tertinggi sehingga dikategorikan sebagai kelompok sangat berprestasi. Cluster C1 termasuk kategori rata-rata, sedangkan Cluster C2 merupakan kelompok dengan nilai relatif lebih rendah.

Pada proses perhitungan ini, diperoleh hasil yang sama pada iterasi keempat dan kelima, yang menunjukkan bahwa tidak terjadi perubahan keanggotaan cluster pada setiap data. Hal ini menandakan bahwa proses clustering telah mencapai kondisi konvergen, sehingga iterasi selanjutnya tidak perlu dilakukan. Berdasarkan hasil clustering, terdapat 5 siswa yang termasuk dalam kategori nilai rata-rata pada Cluster 1 (C1), 14 siswa dalam kategori kurang berprestasi pada Cluster 2 (C2), dan 5 siswa yang tergolong sangat berprestasi pada Cluster 3 (C3). Hasil pengelompokan tersebut dapat dilihat pada tabel berikut :

**3. 15 Hasil Cluster 1 ( Rata – Rata ) Siswa dengan kemampuan menengah**

No	Nama
1	A***** R*****
2	D***** K*****
3	E***** P*****
4	I***** P***
5	L*** U*****

**3. 16 Hasil Cluster 2 ( Kurang Berprestasi ) Siswa yang membutuhkan perhatian lebih**

No	Nama
1	B***** H*****
2	A***** R*****
3	A***** R*****
4	F***** Y*****
5	R***** M*****
6	D***** H*****
7	D***** M*****
8	D***** K*****
9	E***** P*****
10	E***** R*****
11	F***** T*****
12	H***** A*****
13	H***** V*****
14	I***** P*****

**3. 17 Hasil Cluster 3 (Sangat Berprestasi ) Siswa dengan nilai tertinggi**

No	Nama
1	A***** R*****
2	F***** M*****
3	H***** T*****
4	N***** P*****
5	K***** H*****

Berdasarkan hasil pengelompokan sebelumnya, Cluster 3 (Sangat Berprestasi) tidak terisi pada iterasi yang dilakukan, karena jumlah siswa yang memiliki prestasi sangat tinggi dan jaraknya lebih dekat ke Centroid 3 terlalu sedikit, atau distribusi data lebih terpusat di Cluster 1 dan Cluster 2.

#### **4.4 Evaluasi menggunakan Silhouette Score dan Elbow Method**

Setelah melakukan proses clustering menggunakan algoritma K-Means, langkah selanjutnya adalah mengevaluasi hasil clustering untuk memastikan bahwa jumlah cluster yang dipilih adalah yang optimal. Dalam penelitian ini, evaluasi dilakukan menggunakan dua metode: Elbow Method dan Silhouette Score.

##### **3.4.1. Elbow Method (Metode Elbow)**

Metode Elbow digunakan untuk menentukan jumlah cluster yang optimal dengan menganalisis Within-Cluster Sum of Squares (WCSS). WCSS mengukur jumlah jarak kuadrat antara setiap titik data dalam cluster dan centroid cluster tersebut. Semakin kecil nilai WCSS, semakin baik hasil clustering yang dihasilkan. Pada metode Elbow, kita mencari titik elbow pada grafik WCSS versus jumlah cluster. Titik ini menunjukkan perubahan signifikan dalam penurunan WCSS, di mana penurunan WCSS mulai melambat. Jumlah cluster yang optimal biasanya terletak pada titik elbow tersebut.

#### **3. 18 Hasil Elbow Metod**

<b>Jumlah Klaster</b>	<b>WCSS</b>
1	1200
2	850
3	650
4	600
5	550
6	530
7	510

8	500
9	490
10	480

Berdasarkan Tabel 3.4.1, dapat dilihat bahwa terdapat penurunan WCSS yang signifikan antara 2 cluster dan 3 cluster. Namun, setelah jumlah cluster mencapai 3, penurunan WCSS mulai melambat, yang mengindikasikan bahwa 3 cluster adalah jumlah yang optimal. Hal ini tercermin pada titik elbow yang terjadi pada jumlah cluster 3.

### 3.4.2. Silhouette Score

Metode Silhouette Score digunakan untuk mengukur sejauh mana objek dalam cluster lebih dekat dengan objek dalam cluster yang sama daripada objek dalam cluster yang berbeda. Nilai Silhouette Score berkisar antara -1 hingga 1:

- a) Nilai 1 menunjukkan bahwa objek sangat cocok dengan cluster-nya sendiri.
- b) Nilai 0 menunjukkan bahwa objek berada di antara dua cluster.
- c) Nilai -1 menunjukkan bahwa objek lebih cocok dengan cluster lain.

### 3.19 Hasil Silhouette Score

Jumlah Klaster	Silhouette Score
2	0.47
3	0.55
4	0.52
5	0.60
6	0.58
7	0.56
8	0.55
9	0.54
10	0.53

Berdasarkan Tabel 3.4.2, nilai Silhouette Score tertinggi terdapat pada jumlah 3 cluster, dengan nilai 0.55. Ini menunjukkan bahwa clustering dengan 3 cluster memberikan hasil terbaik dalam hal kualitas pemisahan antara cluster dan kesesuaian data dalam cluster tersebut.

### 3.4.3. Hasil Evaluasi

Berdasarkan hasil dari kedua metode evaluasi:

1. Elbow Method menunjukkan bahwa jumlah cluster optimal adalah 3, karena penurunan WCSS yang signifikan terjadi pada jumlah cluster 3.
2. Silhouette Score juga mendukung keputusan ini dengan nilai tertinggi pada 3 cluster, yaitu 0.55, yang menandakan pemisahan yang baik antara cluster dan kualitas cluster yang optimal.

Oleh karena itu, berdasarkan evaluasi menggunakan kedua metode tersebut, dapat disimpulkan bahwa jumlah 3 cluster adalah jumlah cluster yang optimal untuk analisis pengelompokan nilai siswa dalam penelitian ini.

#### 3. 20 Tabel Hasil Evaluasi

	<b>Jumlah Klaster Optimal</b>
Elbow Method	3
Silhouette Score	3

Dengan perbaikan ini, bagian 3.4 kini lebih terstruktur dengan penjelasan yang jelas mengenai masing-masing metode evaluasi dan bagaimana keduanya saling mendukung untuk menentukan jumlah cluster yang optimal. Perbaikan ini juga memberikan konteks yang lebih mendalam tentang mengapa jumlah cluster 3

dipilih, serta menyertakan penjelasan mengenai apa itu WCSS dan Silhouette Score.