

# Perbandingan Algoritma K-Means dan K-Median Clustering Terhadap Nilai Ujian Nasional SMP di Jawa Tengah

Nur Aini Fadhila<sup>1</sup>, Sulastri\*<sup>2</sup>

<sup>1,2</sup>Program Studi Sistem Informasi Universitas Stikubank Semarang  
E-mail: <sup>1</sup>[nurainifadhila26@gmail.com](mailto:nurainifadhila26@gmail.com), \*<sup>2</sup>[sulastri@edu.unisbank.ac.id](mailto:sulastri@edu.unisbank.ac.id)

## Abstrak

*Ujian Nasional (UN) merupakan kegiatan yang diselenggarakan oleh Pemerintah untuk mengukur kemampuan siswa terhadap pencapaian kompetensi lulusan secara nasional sekaligus untuk mengukur mutu dan kualitas pendidikan sekolah. Tujuan dari penelitian ini untuk mengelompokkan mutu dan kualitas pendidikan di wilayah Jawa Tengah untuk memudahkan mengetahui perbedaan. Penelitian ini menggunakan metode clustering dengan algoritma k-means dan algoritma k-median. Proses kalkulasi menggunakan software RStudio. Data yang digunakan dalam penelitian ini yaitu data nilai Ujian Nasional di Jawa Tengah yang diperoleh dari website resmi Pusat Penilaian Pendidikan dan Kebudayaan Kementerian Pendidikan dan Kebudayaan Republik Indonesia sebanyak 1769 sekolah. Hasil dari penelitian kedua algoritma menggunakan metode silhouette diperoleh jumlah kluster sebanyak 5 cluster dengan kategori sangat baik, baik, cukup baik, buruk, dan sangat buruk. Perbedaan hasil dari kedua algoritma terletak pada jumlah kluster, pusat kluster, dan nilai rata – rata pada masing – masing kluster.*

**Kata Kunci**—clustering, k-means, k-median, rstudio, ujian nasional

## 1. PENDAHULUAN

Ujian Nasional (UN) merupakan kegiatan yang diselenggarakan oleh Pemerintah untuk mengukur kemampuan siswa terhadap pencapaian kompetensi lulusan secara nasional pada mata pelajaran tertentu dan hasilnya dapat digunakan sebagai seleksi masuk jenjang pendidikan selanjutnya [1]. Mata pelajaran yang diujikan yaitu terdiri dari Bahasa Indonesia, Bahasa Inggris, Matematika dan IPA. Selain mengukur kemampuan siswa, Ujian Nasional juga dapat memperbaiki mutu dan kualitas pendidikan agar semua sekolah di Jawa Tengah dapat memiliki kualitas pendidikan yang baik. Data mining merupakan suatu proses pengumpulan data yang memiliki jumlah banyak yang kemudian diolah dengan cara mengidentifikasi data yang bersifat penting yang berisi informasi dengan menggunakan beberapa teknik yaitu teknik statistika, teknik matematika, teknik kecerdasan buatan, dan yang terakhir menggunakan teknik machine learning [2]. Oleh karena itu, penelitian ini dilakukan untuk mengelompokkan mutu dan kualitas pendidikan berdasarkan hasil nilai ujian nasional. dalam penelitian ini, metode yang digunakan yaitu metode clustering untuk mengelompokkan data.

Metode clustering merupakan pengelompokan suatu data tertentu yang memiliki kesamaan karakteristik kedalam beberapa kelompok yang disebut cluster [3]. Dalam metode cluster terdapat berbagai macam algoritma, dalam penelitian ini algoritma yang digunakan yaitu algoritma k-means dan algoritma k-median. Algoritma K-Means merupakan sebuah metode dalam clusterisasi yang dilakukan secara penyekatan (partisi) yang memisahkan data yang memiliki nilai berbeda untuk dikelompokkan dengan data yang memiliki nilai sama ke dalam kelompok cluster yang sudah ditentukan. Sedangkan algoritma k-median yaitu pengelompokan data dengan cara menentukan centroid berdasarkan nilai median atau nilai tengah [4]. Penelitian ini di proses menggunakan software rstudio. Rstudio merupakan sebuah sistem pengembang terintegrasi yang

dibuka secara gratis dengan menggunakan bahasa R pemrograman untuk komputasi statistik dan grafik [5].

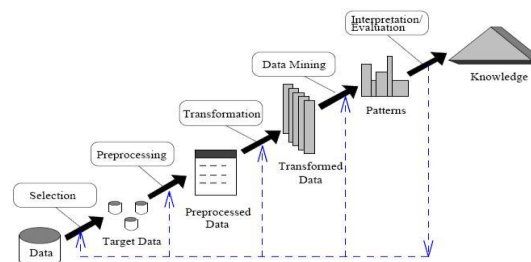
Beberapa penelitian telah membuktikan bahwa algoritma k-means sangat efektif apabila diterapkan dalam pengelompokan data. Penelitian tersebut dilakukan oleh Oktania Purwaningrum, Yudha Yunanto Putra, Amalia Anjani Arifiyan (2021) diperoleh bahwa algoritma k-means sangat efektif apabila diterapkan dalam pengelompokan data, dan dapat diterapkan pada data posyandu untuk mengelompokkan status gizi balita. Penelitian ini memperoleh hasil 3 kluster dari data gizi balita di Posyandu Tanjung 1 RW 06 desa Pepelegi Sidoarjo dengan menggunakan algoritma k-means [6]. Penelitian lainnya juga memberikan hasil yang sama dengan penelitian – penelitian sebelumnya bahwa algoritma k-means sangat efektif digunakan untuk pengelompokan data hasil nilai ujian nasional. dalam penelitian ini menghasilkan 3 kluster yang di kelompokkan menurut nilai ujian nasional kategori baik jika jumlah nilai 233.45 – 263.84, sedangkan kluster dengan kategori cukup baik memiliki nilai 218.38 – 214.12, yang terakhir memiliki nilai 197.98 – 205.47 termasuk dalam kluster kurang baik [7]. Penelitian yang dilakukan menggunakan metode clustering dan algoritma k-means yang diterapkan pada data titik api di Provinsi Riau tahun 2016, memperoleh hasil 3 cluster dengan jumlah 133 daerah sangat rawan titik api, 101 daerah rawan titik api, dan terakhir ada 77 daerah tidak rawan titik api [8]. Adapun penelitian clustering menggunakan algoritma k-means pada data penjualan toko perlengkapan outdoor yang dilakukan oleh Fintri Indriyani, Eni Irfiani menggunakan software Rapid Miner menghasilkan 3 cluster yaitu jenis barang yang paling laris, cukup laris dan jenis barang yang kurang laris [9].

Berdasarkan penelitian – penelitian yang sudah dilakukan membuktikan bahwa algoritma k-means sangat efektif untuk digunakan dalam pengelompokan data dan dapat digunakan untuk berbagai jenis data. Dalam penelitian ini ditambahkan algoritma k-median sebagai alat perbandingan dengan algoritma k-means untuk mengetahui hasil yang lebih bagus.

## 2. METODE PENELITIAN

### 2.1. Tahapan Data Mining

Langkah pertama yang dilakukan dalam tahap data mining yaitu mengumpulkan data (data collection) yang bertujuan untuk memperoleh data yang bersifat mentah, selanjutnya dilakukan preprocessing atau pembersihan data pada data yang masih mentah. Kemudian melakukan transformasi data sampai data siap untuk di proses dalam data mining menggunakan metode dan algoritma yang sudah di tentukan. setelah hasil diperoleh langkah terakhir yang harus dilakukan yaitu melakukan evaluasi untuk mendapatkan kesimpulan dari data yang sudah di olah [10].



Gambar 1. Tahapan data mining

### 2.2. Pengumpulan Data

Dalam penelitian ini data yang digunakan yaitu data nilai ujian nasional tingkat Sekolah Menengah Pertama di Jawa Tengah pada tahun 2019 yang diperoleh dari website resmi Pusat

Penilaian Pendidikan dan Kebudayaan Kementerian Pendidikan dan Kebudayaan Republik Indonesia. Data ini berjumlah 1769 sekolah dari berbagai Kabupaten/Kota yang berada di Jawa Tengah. Dapat dilihat pada Gambar 1.

NO	KODE	NAMA SATUAN PENDIDIKAN	NPSN	STATUS	JUMLAH PESERTA	RERATA NILAI PADA MATA UJI				RERATA NILAI
						BAHASA INDONESIA	BAHASA INGGRIS	MATEMATIKA	IPA	
1	03010001	SMP NEGERI 1 SEMARANG	20328861	N	324	87,62	84,63	83,29	81,69	84,31
2	03010002	SMP NEGERI 2 SEMARANG	20328860	N	285	90,24	91,96	93,63	87,96	90,95
3	03010003	SMP NEGERI 3 SEMARANG	20328859	N	285	89,11	83,76	83,43	76,24	83,14
4	03010004	SMP NEGERI 4 SEMARANG	20337597	N	275	80,15	61,38	60,88	63,74	66,54
5	03010005	SMP NEGERI 5 SEMARANG	20328858	N	284	87,31	84,68	83,92	79,57	83,87
6	03010006	SMP NEGERI 6 SEMARANG	20328857	N	255	83,42	70,27	73,98	71,53	74,80
7	03010007	SMP NEGERI 7 SEMARANG	20328856	N	237	83,7	73,81	73,24	69,81	75,14
8	03010008	SMP NEGERI 8 SEMARANG	20328855	N	287	83,73	72,03	71,12	70,16	74,26
9	03010009	SMP NEGERI 9 SEMARANG	20328854	N	286	87,01	82,76	83,6	78,9	83,07
10	03010010	SMP NEGERI 10 SEMARANG	20328840	N	277	75,99	57,79	51,53	56,44	60,44
11	03010011	SMP NEGERI 11 SEMARANG	20328839	N	295	76,57	60,3	57,03	55,31	62,30
12	03010012	SMP NEGERI 12 SEMARANG	20328825	N	289	84,39	76	72,29	66,65	74,83
13	03010013	SMP NEGERI 13 SEMARANG	20328824	N	282	82,02	70,17	73,4	67,6	73,30
14	03010014	SMP NEGERI 14 SEMARANG	20328823	N	283	83,23	68,47	75,21	71,78	74,67
15	03010015	SMP NEGERI 15 SEMARANG	20328822	N	321	82,87	72,27	73,81	69,96	74,73

Gambar 1. Potongan dataset mentah

### 2.3. Pembersihan Data

Pada tahap ini dilakukan pembersihan data atau variable pada data. Sebelum di proses dalam data mining, data yang masih bersifat mentah harus melewati tahap pembersihan data (data cleaning). Variable awal pada data nilai ujian nasional berjumlah 11, setelah di bersihkan hanya tersisa 5 variabel. Dapat dilihat pada Tabel 1.

Tabel 1. Pembersihan data

No	Variable
1	Nama Satuan Sekolah
2	Nilai Bahasa Indonesia
3	Nilai Bahasa Inggris
4	Nilai Matematika
5	Nilai IPA

### 2.4. Transformasi Data

Tahap ini perlu dilakukan karena bertujuan untuk memudahkan proses data mining. Transformasi data yang dilakukan dalam tahap ini yaitu mengubah variable nama sekolah, nilai Bahasa Indonesia, Bahasa Inggris, Matematika, IPA. Dapat dilihat pada Tabel 2.

Tabel 2. Transformasi data

Variable lama	Variable baru
Nama Satuan Sekolah	Sekolah
Nilai Bahasa Indonesia	BI
Nilai Bahasa Inggris	BING

Nilai Matematika	MTK
Nilai IPA	IPA

## 2.5. Analisis Data

### a. K-Means

Algoritma k-means merupakan sebuah metode dalam clusterisasi yang bekerja dengan cara melakukan pengelompokan yang memisahkan suatu data yang memiliki nilai berbeda kemudian dikumpulkan dengan data yang bernilai sama ke dalam kelompok kluster yang sudah di tentukan[11]. Langkah – langkah dalam algoritma k-means yaitu :

1. Menentukan jumlah cluster yang akan digunakan.
2. Menentukan pusat cluster.
3. Menghitung jarak objek ke pusat dengan menggunakan persamaan *Euclidean Distance* (1).

$$d(i, j) = \sqrt{|X_{i1} - X_{j1}|^2 + |X_{i2} - X_{j2}|^2 + \dots + |X_{ip} - X_{jp}|^2} \quad (1)$$

Keterangan :

$d(i,j)$  = jarak *Euclidean*

$X_i$  = nilai titik 1

$X_j$  = nilai titik 2

4. Mengelompokkan objek berdasarkan jarak minimum.
5. Melakukan pengecekan objek apakah ada yang harus di pindah atau tidak. Jika iya maka harus melakukan penentuan cluster lagi, jika tidak maka proses berakhir.

### b. K-Median

Algoritma k-median merupakan pengelompokan data dengan cara menentukan centroid berdasarkan nilai median atau nilai tengah. Berikut merupakan langkah – langkah algoritma k-median :

1. Menentukan jumlah cluster.
2. Menentukan titik pusat cluster terbaik menggunakan persamaan atau rumus *centroid* (2).

$$c = \sum_{i=1}^n \frac{x_i}{n} \quad (2)$$

Keterangan :

$c$  = *centroid* kluster

$x_i$  = objek ke-n

3. Menghitung jarak data terhadap pusat cluster menggunakan persamaan (1).
4. Mengelompokkan data sesuai nilai minimum jarak ke pusat cluster.
5. Melakukan pengecekan objek apakah ada yang harus di pindah atau tidak. Jika iya maka harus kembali mengulang dari langkah 2, jika tidak maka proses berakhir.

## 2.6. Evaluasi Cluster

Untuk mengatasi kesulitan dalam menentukan jumlah kluster yang sesuai berdasarkan data yang sudah digunakan, maka perlu dilakukan evaluasi pada cluster yang bertujuan untuk

mengevaluasi hasil dari algoritma agar mendapat hasil kluster terbaik. Dalam penelitian ini menggunakan metode *Silhouette* yang digunakan untuk proses evaluasi data. pada metode *Silhouette* dilakukan dengan cara menghitung nilai rata – rata masing – masing cluster [12].

### 3. HASIL DAN PEMBAHASAN

#### 3.1. Implementasi Rstudio

##### 3.3.1. Persiapan Data

##### 1. Input Data

Langkah awal yang perlu dilakukan sebelum memulai proses clustering pada rstudio yaitu mengupload data yang siap di proses ke dalam software rstudio. Setelah berhasil di upload untuk dapat melihat data menggunakan *source code* “*readxl*” kemudian “*view*”. Berikut potongan dataset yang siap di proses dapat dilihat pada Gambar 2.

	Sekolah	BI	BING	MTK	IPA
1	SMP NEGERI 1 SEMARANG	87.62	84.63	83.29	81.69
2	SMP NEGERI 2 SEMARANG	90.24	91.96	93.63	87.96
3	SMP NEGERI 3 SEMARANG	89.11	83.76	83.43	76.24
4	SMP NEGERI 4 SEMARANG	80.15	61.38	60.88	63.74
5	SMP NEGERI 5 SEMARANG	87.31	84.68	83.92	79.57
6	SMP NEGERI 6 SEMARANG	83.42	70.27	73.98	71.53
7	SMP NEGERI 7 SEMARANG	83.70	73.81	73.24	69.81
8	SMP NEGERI 8 SEMARANG	83.73	72.03	71.12	70.16
9	SMP NEGERI 9 SEMARANG	87.01	82.76	83.60	78.90
10	SMP NEGERI 10 SEMARANG	75.99	57.79	51.53	56.44
11	SMP NEGERI 11 SEMARANG	76.57	60.30	57.03	55.31
12	SMP NEGERI 12 SEMARANG	84.39	76.00	72.29	66.65
13	SMP NEGERI 13 SEMARANG	82.02	70.17	73.40	67.60
14	SMP NEGERI 14 SEMARANG	83.23	68.47	75.21	71.78
15	SMP NEGERI 15 SEMARANG	82.87	72.27	73.81	69.96

Gambar 2. Potongan dataset nilaiun

##### 2. Data Selection

Setelah melakukan input data selanjutnya dilakukan pemilihan data yaitu memilih data yang digunakan dalam proses cluster. Data yang digunakan dalam proses clustering yaitu jenis data yang bersifat *numerik* atau angka yang sudah di proses melalui software rstudio. Dapat dilihat pada Gambar 3.

BI	BING	MTK	IPA
87.62	84.63	83.29	81.69
90.24	91.96	93.63	87.96
89.11	83.76	83.43	76.24
80.15	61.38	60.88	63.74
87.31	84.68	83.92	79.57
83.42	70.27	73.98	71.53
83.70	73.81	73.24	69.81
83.73	72.03	71.12	70.16
87.01	82.76	83.60	78.90
75.99	57.79	51.53	56.44
76.57	60.30	57.03	55.31
84.39	76.00	72.29	66.65
82.02	70.17	73.40	67.60
83.23	68.47	75.21	71.78
82.87	72.27	73.81	69.96

Gambar 3. Dataset numerik

3. normalisasi Data

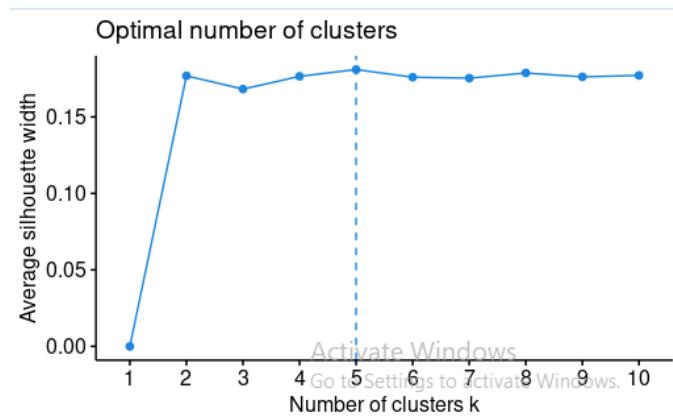
Normalisasi data dilakukan untuk menstandarisasikan angka yang bertujuan untuk menyamakan rentang nilai pada data agar dalam satu dataset memiliki satuan nilai yang seragam. Normalisasi data dapat dilakukan dengan berbagai cara salah satunya menggunakan rstudio dengan perintah “Scale”. potongan hasil normlisasi dapat dilihat pada Gambar 4.

BI	BING	MTK	IPA
1.94000653	3.58242999	2.807441551	2.719042567
2.25039745	4.36060688	3.683324985	3.327299513
2.11652656	3.49006793	2.819300708	2.190334377
1.05503700	1.11413358	0.909129391	0.977700914
1.90328089	3.58773815	2.860807757	2.513379932
1.44243330	2.05792520	2.018807629	1.733414088
1.47560485	2.43374323	1.956123514	1.566555724
1.47915895	2.24477258	1.776541998	1.600509461
1.86773995	3.38390464	2.833701113	2.448382778
0.56220257	0.73300738	0.117107138	0.269522971
0.63091506	0.99947723	0.583002581	0.159900906
1.55734903	2.66624083	1.875650665	1.260001984
1.27657556	2.04730887	1.969676836	1.352162127
1.41992404	1.86683128	2.122998791	1.757666757
1.37727490	2.27025177	2.004407224	1.581107325

Gambar 4. Tabel normalisasi data

3.3.2. Hasil Clustering

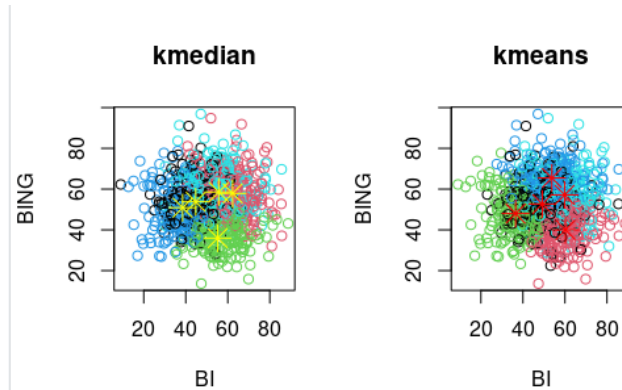
Dari proses perhitungan komputasi yang dilakukan menggunakan algoritma k-means dan algoritma k-median pada software rstudio memperoleh hasil cluster sebanyak 5 cluster. Gambar penentuan titik optimal cluster dapat dilihat pada gambar 5.



Gambar 5. Titik pusat cluster terbaik

Dari gambar diatas membuktikan bahwa cluster terbaik berada pada titik 5 dan dapat di kategorikan sebagai sekolah yang memiliki mutu dan kualitas pendidikan sangat baik, baik, cukup

baik, buruk, dan sangat buruk. Setelah mengetahui titik kluster langkah selanjutnya yaitu menampilkan visualisasi pengelompokan dataset yang sudah dilakukan menggunakan software rstudio. dapat dilihat pada gambar 6.



Gambar 6. Visualisasi pengelompokan kluster

Dari visualisasi pengelompokan di atas dapat dilihat perbedaan hasil antara algoritma k-median dan algoritma k-means. Pada algoritma k-median letak titik kluster di berikan tanda bintang warna kuning dan algoritma k-means warna merah. Selain dari hasil visualisasi ada juga perbedaan dari jumlah kluster, nilai pusat, dan nilai rata – rata pada masing – masing kluster.

#### 1. Perbedaan Hasil Jumlah Kluster

```
$size
  [,1]
[1,] 204
[2,] 189
[3,] 173
[4,] 242
[5,] 192

> #melakukan clustering kmeans sebagai pembanding
> cl.kmeans=kmeans(data,5)
> print(cl.kmeans)
K-means clustering with 5 clusters of sizes 193, 209, 196, 179, 223
```

Gambar 7. hasil jumlah kluster

Dari hasil yang diperoleh pada Gambar 7 dapat dilihat hasil perbedaan jumlah kluster dari algoritma k-median dan algoritma k-means. Jumlah kluster dari kedua algoritma tersebut memiliki nilai yang berbeda pada masing – masing kluster. Algoritma k-median jumlah anggota tertinggi berada pada kluster 4 yang berjumlah 242 sekolah, dan jumlah anggota paling sedikit berada pada kluster 3 yang berjumlah 173 sekolah. Algoritma k-means memiliki jumlah anggota tertinggi berada pada kluster 5 yang berjumlah 223 sekolah, dan jumlah kluster paling sedikit berada pada kluster 4 dengan jumlah 179 sekolah.

#### 2. Perbedaan Nilai Centroid Kluster

```
$centers
  BI      BING      MTK      IPA
1 56.82765 58.45519 40.25281 48.22402
2 38.48780 50.66877 47.33096 59.39409
3 54.94882 36.19127 49.76364 51.37935
4 62.30236 57.54301 62.65194 59.62031
5 44.55278 55.90567 62.92719 39.79081
```

```
Cluster means:
  BI      BING      MTK      IPA
1 46.44293 63.56488 54.10630 40.43094
2 60.90106 62.22763 54.23402 63.55899
3 37.21056 47.83920 46.88726 58.27074
4 55.03747 47.56851 68.99071 52.62142
5 59.65242 42.90576 43.50346 47.07526
```

Gambar 8. Hasil nilai centroid

Dari hasil yang diperoleh pada Gambar 8 dapat dilihat hasil perbedaan nilai centroid dari algoritma k-median dan algoritma k-means. Nilai centroid dari kedua algoritma tersebut memiliki jumlah yang berbeda dari masing – masing nilai mata uji pada tiap kluster.

### 3. Perbedaan Nilai Rata – Rata pada Masing – Masing Kluster

```
> dd1[1:5]%>%
+ mutate(cluster=cl.kmedian$cluster)%>%
+ group_by(cluster)%>%
+ summarise_all("mean")
# A tibble: 5 x 6
  cluster[,1]    BI    BING    MTK    IPA cl.kmedian.cluster
  <dbl> <dbl> <dbl> <dbl> <dbl> <dbl>
1           1  57.2  59.7  39.2  48.1         4.92
2           2   37.4  50.2  47.8  60.5         3.97
3           3   56.5  35.9  50.6  51.0         2.95
4           4   61.9  57.9  62.8  60.4         2.05
5           5   45.0  56.4  62.6  40.3         1.43
> dd2[1:5]%>%
+ mutate(cluster=cl.kmeans$cluster)%>%
+ group_by(cluster)%>%
+ summarise_all("mean")
# A tibble: 5 x 6
  cluster    BI    BING    MTK    IPA cl.kmeans.cluster
  <int> <dbl> <dbl> <dbl> <dbl> <dbl>
1       1  46.4  63.6  54.1  40.4         2.61
2       2  60.9  62.2  54.2  63.6         4.59
3       3  37.2  47.8  46.9  58.3         3.06
4       4  55.0  47.6  69.0  52.6         2.52
5       5  59.7  42.9  43.5  47.1         2.24
```

Gambar 9. Hasil nilai rata-rata masing – masing kluster

Dari hasil yang diperoleh pada Gambar 9 dapat dilihat hasil perbedaan nilai rata – rata dari masing – masing kluster dari algoritma k-median dan algoritma k-means. Algoritma k-median diperoleh hasil nilai rata – rata sangat baik berada pada kluster 4, kluster 5 memiliki nilai baik, kluster 1 memiliki nilai cukup baik, kluster 2 memiliki nilai buruk, dan kluster 3 memiliki nilai sangat buruk. Algoritma k-means diperoleh hasil nilai rata – rata sangat baik berada pada kluster 2, kluster 4 memiliki nilai baik, kluster 1 memiliki nilai cukup baik, kluster 5 memiliki nilai buruk, dan kluster 3 memiliki nilai sangat buruk.

## 4. KESIMPULAN

Berdasarkan hasil penelitian yang sudah dilakukan pada data nilai ujian nasional tingkat SMP di Jawa Tengah dapat di simpulkan bahwa :

1. Titik optimal kluster algoritma k-means dan algoritma k-median menghasilkan 5 kluster yaitu sangat baik, baik, cukup baik, buruk, dan sangat buruk.
2. Hasil algoritma k-means menunjukkan bahwa kluster 1 berjumlah 198 sekolah dengan kategori cukup baik, kluster 2 berjumlah 209 sekolah dengan kategori sangat baik, kluster 3 berjumlah 196 sekolah dengan kategori sangat buruk, kluster 4 berjumlah 179 sekolah dengan kategori baik, dan kluster 5 berjumlah 223 sekolah dengan kategori buruk.

3. Hasil algoritma k-median menunjukkan bahwa kluster 1 berjumlah 204 sekolah dengan kategori cukup baik, kluster 2 berjumlah 189 sekolah dengan kategori buruk, kluster 3 berjumlah 173 sekolah dengan kategori sangat buruk, kluster 4 berjumlah 242 sekolah dengan kategori sangat baik, dan kluster 5 berjumlah 192 sekolah dengan kategori baik.
4. Dari hasil klustering menggunakan dua algoritma memperoleh hasil persamaan pada titik optimal kluster yaitu sama berjumlah 5. Adapun perbedaan yang di hasilkan yaitu jumlah anggota kluster, nilai centroid dari masing – masing kluster, dan nilai rata – rata dari masing – masing kluster.

## DAFTAR PUSTAKA

- [1] C. S. Triwiniastuti and W. Sabatini, “Implementasi Program Ujian Nasional Di Sma Negeri,” *Kelola J. Manaj. Pendidik.*, vol. 6, no. 1, pp. 54–63, 2019, doi: 10.24246/j.jk.2019.v6.i1.p54-63.
- [2] Yuli Mardi, “Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database ( KDD ) . Jurnal Edik Informatika,” *J. Edik Inform.*, vol. 2, no. 2, pp. 213–219, 2019.
- [3] S. Arwansyah, “Clusterisasi Varietas Benih Tanaman Padi Menggunakan,” *Pros. Semin. Ilm. Sist. Inf. Dan Teknol. Inf.*, vol. IX, no. 2, pp. 1–10, 2020.
- [4] F. N. Adlin and T. Purwaningsih, “1 Perbandingan Cluster K-Medoids Dan K-Median Tingkat Produksi Mangga Dan Pisang Kabupaten Indramayu,” pp. 1–11, 2021.
- [5] A. Wahit *et al.*, “FUZZY C-MEANS ALGORITHM FOR PROVINCE GROUPING IN Kata kunci : Prasarana Kesehatan , Clustering , Fuzzy C-Means , Davies Boulding Index .,” vol. 2, no. 14, pp. 1–9, 2021.
- [6] O. Purwaningrum, Y. Y. Putra, and A. A. Arifiyanti, “Penentuan Kelompok Status Gizi Balita dengan Menggunakan Metode K-Means,” vol. 15, no. 2, pp. 129–136, 2021.
- [7] T. P. Sutriyani, A. M. Siregar, and D. S. Kusumaningrum, “Implementasi Algoritma K-Means terhadap Pengelompokan Nilai Ujian Nasional Tingkat SMP di Provinsi Jawa Barat,” *Techno Xplore J. Ilmu Komput. dan Teknol. Inf.*, vol. 3, no. 1, pp. 30–36, 2018, doi: 10.36805/technoexplore.v3i1.797.
- [8] S. Sukamto, I. D. Id, and T. R. Angraini, “Penentuan Daerah Rawan Titik Api di Provinsi Riau Menggunakan Clustering Algoritma K-Means,” *JUITA J. Inform.*, vol. 6, no. 2, p. 137, 2018, doi: 10.30595/juita.v6i2.3172.
- [9] F. Indriyani and E. Irfiani, “Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means,” *JUITA J. Inform.*, vol. 7, no. 2, p. 109, 2019, doi: 10.30595/juita.v7i2.5529.
- [10] D. Firdaus, “Penggunaan Data Mining dalam Kegiatan Sistem Pembelajaran Berbantuan Komputer,” *J. Format*, vol. 6, no. 2, pp. 91–97, 2017.
- [11] M. W. Talakua, Z. A. Leleury, and A. W. Taluta, “Analisis Cluster Dengan Menggunakan Metode K-Means Untuk Pengelompokan Kabupaten/Kota Di Provinsi Maluku Berdasarkan Indikator Indeks Pembangunan Manusia Tahun 2014,” *BAREKENG J. Ilmu Mat. dan Terap.*, vol. 11, no. 2, pp. 119–128, 2017, doi: 10.30598/barekengvol11iss2pp119-128.

- [12] A. Aditya, I. Jovian, and B. N. Sari, "Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019," *J. Media Inform. Budidarma*, vol. 4, no. 1, p. 51, 2020, doi: 10.30865/mib.v4i1.1784.