

Klasterisasi Penentuan Minat Siswa dalam Pemilihan Sekolah Menggunakan Metode Algoritma *K-Means Clustering*

Suhefi Oktarian^{1✉}, Sarjon Defit², Sumijan³

^{1,2,3}Univeristas Putra Indonesia YPTK Padang

suhefioktarian05@gmail.com

Abstract

Education is one of the main focuses of the Indragiri Hilir Regency Government work program. Based on data from the Regional Central Statistics Agency of Indragiri district in 2019, the high level of student interest in attending school is at the elementary and junior high school levels. K-means clustering is a data grouping technique by dividing existing data into one or more clusters. School grouping based on student interest is important because at the high school level students' interest in education has decreased so that information is needed which schools are in great demand, sufficient interest and less interest by students at the junior high school level when after finishing elementary school education. This study aims to assist the Education Office in the decision-making process to determine which school students are most interested in as a reference in development both in terms of quality and quantity. The data used in this study is the Dapodikdasmen data in 2019. Data processing in this study uses the K-means clustering method with a total of 3 clusters, namely cluster 0 (C0) is less attractive, Cluster 1 (C1) is quite attractive, cluster 2 (c2) is very interested in students in choosing a school. The results of the clustering process with 2 iterations state that for cluster 0 there are 6 school data, for cluster 1 there are 3 school data, cluster 2 is 1 school data.

Keywords: Data Mining, K-Means, Clustering, School, Interests.

Abstrak

Pendidikan merupakan satu diantara fokus utama program kerja Pemerintah Kabupaten Indragiri Hilir. Berdasarkan data Badan Pusat Statistik Daerah kabupaten Indragiri tahun 2019 memaparkan, tingginya tingkat minat siswa dalam mengenyam bangku sekolah adalah pada jenjang SD dan SMP. *K-means Clustering* merupakan salah satu Teknik pengelompokan data dengan cara membagi data yang ada ke dalam bentuk satu atau lebih *cluster*. Pengelompokan sekolah berdasarkan minat siswa merupakan hal penting dikarenakan pada tingkat SMA minat siswa dalam mengenyam pendidikan sudah berkurang sehingga di perlukan informasi sekolah mana yang sangat diminati, cukup diminati dan kurang diminati oleh siswa pada tingkat SMP ketika setelah selesai dari pendidikan SD. Penelitian ini bertujuan membantu pihak Dinas Pendidikan dalam proses pengambilan keputusan untuk menentukan sekolah mana yang paling banyak diminati oleh siswa guna sebagai acuan dalam pembangunan baik dari segi kualitas maupun kuantitas. Data yang digunakan dalam penelitian ini adalah data Dapodikdasmen tahun 2019. pengolahan data dalam penelitian ini menggunakan metode *K-means Clustering* dengan jumlah 3 cluster yaitu cluster 0 (C0) kurang diminati, Cluster 1 (C1) cukup diminati, cluster 2 (c2) sangat diminati siswa dalam memilih sekolah. Hasil dari proses *clustering* dengan 2 kali iterasi menyatakan bahwa untuk cluster 0 berjumlah 6 data sekolah, untuk cluster 1 berjumlah 3 data sekolah cluster 2 berjumlah 1 data sekolah.

Kata kunci: *Data Mining*, *K-Means*, *Clustering*, Sekolah, Minat.

© 2020 JIDT

1. Pendahuluan

Kabupaten Indragiri Hilir merupakan kabupaten terluas kedua di Provinsi Riau dengan luas mencapai 11.605,97 km². Kabupaten Indragiri Hilir terdiri dari 20 kecamatan, 39 kelurahan dan 197 desa dan memiliki 664 lembaga Pendidikan sekolah terdiri dari SD dan SMP dengan jumlah penduduk 731.236 jiwa sebagian besar penduduk Indragiri hilir berprofesi menjadi petani.

Dinas Pendidikan kabupaten Indragiri Hilir berfokus pada pengelolaan pada bidang SD dan SMP. Hasil data yang di peroleh total siswa dari tingkat SD berjumlah 71944 siswa dan SMP berjumlah 19550 siswa. Data Badan statistik kabupaten Indragiri hilir tahun 2019

menjabarkan tingkat partisipasi masyarakat mengakses pendidikan secara umum ada pada jenjang SD yaitu 94,92 persen dan SMP yaitu 82,39 sementara itu jenjang SMA 51,9 persen. Rendahnya minat pada SMA antara lain banyaknya anak yang putus sekolah untuk membantu orang tuanya bertani dan alasan finansial lainnya.

Klasterisasi Minat Siswa dalam pemilihan sekolah dengan mengelompokkan data sekolah yang yang paling banyak diminati khususnya tingkat SMP dalam penelitian ini menggunakan metode *clustering*, data yang diperlukan sebagai acuan untuk melakukan pengelompokan adalah data pokok pendidikan. Data yang dikelompokkan dengan menggunakan metode *k-*

means mempermudah siswa mengetahui sekolah mana yang sangat diminati, cukup diminati dan kurang diminati dalam memilih sekolah tingkat SMP.

Sejumlah Penelitian yang pernah menggunakan metode algoritma K-Means diantaranya adalah Gustientiedina dkk (2019) Penelitian ini menerapkan *Algoritma K-Means Untuk Clustering* data obat-obatan pada RSUD Pekanbaru dengan hasil klasterisasi pada data obat – obatan ini bahwa kelompok pemakaian sedikit, sedang dan tinggi [1]. Penelitian yang sudah dilakukan oleh Mar'i & Supianto (2018) menerapkan *Clustering Credit Card Holder* yaitu menghasilkan fakta dengan menerapkan *Improved K-Means* dan PSO memberikan model *cluster* yang lebih relevan [2].

Sari dan Ayu (2018) dalam penelitian penentuan kriteria dalam memilih sekolah dasar dengan menerapkan *K-Means Clustering* menghasilkan kriteria memilih sekolah berdasarkan lokasi sekolah, latarbelakang keluarga yang pernah bersekolah ditempat yang sama, pendidikan orang tua, penghasilan orang tua dan alamat rumah [3]. Penelitian Aditya dkk (2020) dalam penelitian Implementasi *K-Means Clustering* ujian nasional sekolah menengah pertama di Indonesia tahun 2018/2019 menghasilkan 3 kluster dengan nilai evaluasi Connectivity 11.916, Dunn 0.246 dan Silhouette 0.464 (struktur data lemah) [4]. Penelitian yang sudah dilakukan oleh Rosmini, dkk (2018) untuk implementasi metode *k-means* dalam pemetaan kelompok mahasiswa melalui data aktivitas kuliah dengan hasil 2 kategori yaitu memprediksi lulus tepat waktu dan tidak tepat waktu [5].

Penelitian Ridho & Kusuma (2018) deteksi intrusi jaringan dengan *k-means clustering* pada akses log dengan teknik pengolahan *big data*. Hasil Penelitian ini menerangkan satu model yang bisa digunakan untuk mendeteksi sebuah serangan dengan probabilitas deteksi sebesar 99.68%. serta 80% presisi model bentukan dari K-Means [6]. Penelitian Indriyani & Irfiani (2019) *clustering* data penjualan pada toko perlengkapan *Outdoor* Menggunakan Metode *K-Means* dan didapatkan hasil akhir berupa tiga cluster yang berguna bagi pihak manajemen [7]. Peneliti Putra dan Wadisman, (2018) implementasi data mining pemilihan pelanggan potensial menggunakan pengujian *Algoritma K-Means* memberikan dampak yang lebih akurat dalam pengelompokan pelanggan potensial [8]. Penelitian Rustam (2018) optimasi *k-means clustering* untuk identifikasi daerah endemik penyakit menular dengan *algoritma particle swarm optimization* Hasil penelitian dengan 10 kali iterasi serta menggunakan *Algoritma K-Means* dengan PSO memiliki *performance* yang lebih baik [9].

Parlina dkk (2018) memanfaatkan algoritma k-means dalam menentukan pegawai yang layak mengikuti *assessment center* untuk *clustering* program SDP. diperoleh dari penelitian ini untuk 3 kelompok yang dikategorikan dengan kelompok lolos, hampir lolos, tidak lolos. Dengan nilai yang diperoleh C1=8;66;13,

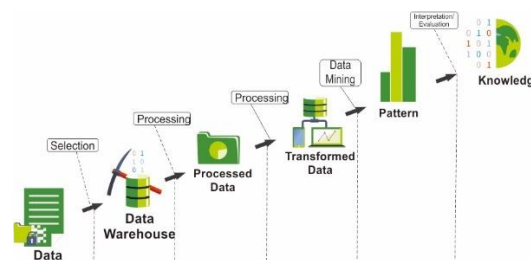
C2=10;71;14, C3=7;60;12 dengan keterangan optimal [10]. Maulida (2018) penerapan data mining dalam mengelompokkan kunjungan wisatawan ke objek wisata unggulan di prov. Dki Jakarta dengan *K-Means*. penelitian ini menghasilkan 3 cluster dengan jumlah kunjungan C1=15.438.488, C2=4.464.577, C3=342.332 [11]. Penelitian Nalendra (2020) implementasi algoritma *k-mean* dalam pengelompokan data kecelakaan hasil dari clustering terdapat 3 jenis yaitu jarang terjadi kecelakaan, rawan kecelakaan dan sangat rawan kecelakaan [12]. Priambodo & Prasetyo (2018) pemetaan penyebaran guru di provinsi banten dengan Menggunakan Metode *Spatial Clustering K-Means*. Hasil yang diperoleh adalah mampu memetakan persebaran data guru [13].

2. Metodologi Penelitian

Objek penelitian ini menggunakan data yang diperoleh dari data rekam data per satuan pendidikan (Dapodik) pada Dinas Pendidikan Kabupaten Indragiri Hilir tahun 2019.

2.1. Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah suatu ilmu dari percabangan ilmu statistik, database, AI, visualisasi dan komputer paralel yang mempengaruhi pengetahuan “interdisciplinary knowledge” [14] yang melibatkan hasil ekstraksi kecenderungan suatu pola, sehingga mengubah hasilnya secara tepat dan akurat serta menjadi informasi yang mudah dipahami [15] proses KDD seperti pada gambar berikut ini yang diadopsi dari



Gambar 1. Langkah Knowledge Discovery in Database (KDD)

Adapun penjelasan pada Gambar 2 menjabarkan langkah *Knowledge Discovery in Database* (KDD) yaitu:

1.Data Selection

Data Selection adalah tahap menetapkan variabel yang akan digunakan dalam proses data mining

2.Preprocessing

preprocessing memiliki dua tahapan yaitu tahap pertama melakukan proses *cleaning* yakni memeriksa data apakah terdapat penggadaan data atau data yang tidak konsisten, serta memperbaiki kesalahan pada data, seperti kesalahan huruf. Tahapan kedua yaitu *integration*, tahap ini melakukan terhadap atribut yang mengidentifikasi entitas yang unik.

3. Transformation

Tahap untuk merubah data sesuai dengan format yang mendukung pada pengolahan data mining.

4. Data Mining

Tahapan utama untuk formula yang di jalankan agar memperoleh *knowledge* dari data yang di olah. Untuk penelitian ini di terapkan teknik *clustering* yaitu metode *K-Means Clustering*.

5. Interpretation/Evaluation

Tahap ini dimana mengidentifikasi hubungan– hubungan yang menarik di dalam *knowdge base* yang sudah identifikasi serta menghasilkan pola-pola khas maupun model prediksi yang dievaluasi untuk menilai kajian yang ada sudah memenuhi target yang diinginkan.

6. Knowledge

Pola-pola yang dihasilkan akan dipresentasikan kepada pengguna. Pada tahapan ini pengetahuan baru yang dihasilkan bisa dipahami semua orang yang akan dijadikan acuan pengambilan keputusan.

2.2 Clustering

Clustering adalah metode untuk membagi data menjadi beberapa kelompok data sesuai kateristik kesamaan data yang khas akan ditempatkan pada cluster yang sama dan pada data yang memiliki kemiripan maksimum akan ditempatkan pada cluster yang sama pula [16] cara berikut ini untuk melihat jarak antar cluster diantaranya sebagai berikut [17] :

- *Single link* adalah jarak terdekat antara satu elemen di dalam suatu cluster dengan elemen lainnya di dalam cluster yang berbeda.
- *Complete link* merupakan jarak terbesar antara satu elemen di dalam suatu cluster dengan elemen lainnya di dalam cluster yang berbeda.
- *Average* merupakan jarak rata-rata antara di dalam suatu cluster dengan elemen lainnya di dalam cluster yang berbeda
- *Centroid* merupakan jarak antara centroid dari tiap cluster dengan *centroid* cluster lainnya.
- *Medoid* merupakan jarak antara medoid dari tiap bagian cluster dengan *medoid* cluster lainnya.

2.3. Algoritma K-Means

Algoritma *K-mean* adalah salah satu cara pengelompokkan data sekatan (*nonhierarki*) yang berusaha membagi data yang ada kedalam bentuk dua atau lebih kelompok [18]. Teknik pengelompokan ini bekerja berdasarkan partisi cluster, dan pada setiap iterasi dari hasil pengelompokan hirarki cluster [19].

Tahapan implementasi algoritma *K-Means clustering* yakni dengan menentukan jumlah cluster *K* yang telah di ditetapkan diawal, kemudian menetapkan nilai *centroid* secara random, selanjutnya mengestimasi jarak pada data ke setiap *centroid* dengan menerapkan metode korelasi antara dua buah objek yang terdapat pada rumus *Euclidean Distance* berikut ini.

$$Sh_i = \sqrt{(m_h - m_i)^2 + (n_h - n_i)^2 + \dots + (q_h - q_i)^2} \quad (1)$$

Dimana:

Sh_i = Jarak antara *h* dan *i*

m_h = Koordinat *m* objek

m_i = Koordinat *m* pusat

n_h = Koordinat *n* objek

Hasil yang di peroleh dari perhitungan jarak dari setiap data, langkah berikutnya menggabungkan setiap data yang dilihat berdasarkan jarak terdekat antara data dan cluster. Kemudian Langkah berikutnya yaitu menentukan tempat centroid baru dengan menghitung nilai rata-rata berdasarkan data pada cluster yang serupa. Untuk menentukan tempat centroid baru untuk menghitung jarak antara data dan centroid dapat menggunakan rumus berikut ini.

$$K_c = \frac{s_1 + s_2 + s_3 + \dots + s_n}{\sum s} \quad (2)$$

Dimana :

K_c = data *cluster* ke *c*

s_1 = nilai *record* data ke 1

s_2 = nilai *record* data ke 2

$\sum s$ = jumlah *record* data

Kelebihan metode algoritma *k-means clustering* sebagai berikut [20]:

- Sangat mudah dalam proses penerapan;
- Mudah untuk beradaptasi;
- Sangat umum digunakan.

3. Hasil dan Pembahasan

Data yang digunakan dalam penelitian ini adalah data Pokok Pendidikan (*Dapodik*) Kabupaten Indragiri Hilir tahun 2019 yang diperoleh dari Dinas Pendidikan Kabuapten Idragiri Hilir. Adapun sampel data sekolah yang menjadi objek penelitian seperti pada Tabel 1.

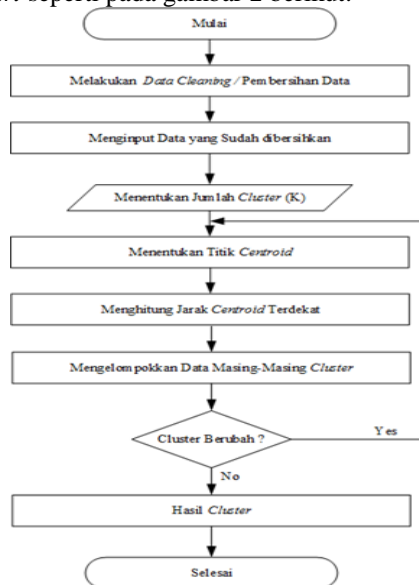
Tabel 1. Sampel Data Pokok Pendidikan

No	Nama Satuan Pendidikan	NPSN	Status Sekolah	Jumlah Rombel	Jumlah Peserta Didik	Tenaga Guru & Tendik
1	SMP Bhakti Nusa	10496624	Swasta	3	9	10
2	SMP Budi Daya Pulau Palas	10402080	Swasta	4	76	9
3	SMPN 1 Enok	10402011	Negeri	6	131	13
4	SMPN 1 Gaung	10402010	Negeri	16	332	38
5	SMPN 1 Gaung Anak Serka	10402008	Negeri	8	162	22
6	SMPN 1 Kuindra	10402005	Negeri	6	136	12
7	SMPN 1 Pulau Burung	10402023	Negeri	9	222	18
8	SMPN 1 Tembilahan Hulu	10402001	Negeri	33	1014	61
9	SMPN 3 Tembilahan Hulu	10402084	Negeri	18	369	54
10	SMPS IT Daarul Rahman	10497402	Swasta	3	9	10

Penelitian ini menerapkan analisa berdasarkan metode *k-means clustering* untuk mengklaster minat siswa dalam memilih sekolah berdasarkan jumlah Rombel (ruangan belajar), Jumlah Peserta Didik serta Jumlah data guru dan tendik dengan menjadi 3 *cluster*, yaitu *cluster C0* untuk sekolah kurang diminati, *cluster C1* untuk sekolah cukup diminati, dan *cluster C2* untuk sekolah sangat diminati. Berikut ini merupakan Langkah-langkah metode algoritma *k-means* adalah:

1. Melakukan pembersihan data atau di kenal dengan istilah *Data Cleaning*
2. Menginput data yang sudah *cleaning*
3. Menentukan jumlah *cluster* (*K*)
4. Menentukan titik *centroid* secara acak
5. Menghitung jarak *centroid* terdekat
 - a. Proses iterasi 1
 - b. Proses iterasi 2
 - c. Proses iterasi ... *n*
6. Mengelompokkan data masing-masing *cluster*
7. Mengecek perubahan data
8. Hasil *cluster*

Agar lebih jelas dan terarah, penerapan algoritma *k-means* pada penelitian ini digambarkan dalam bentuk *flowchart* seperti pada gambar 2 berikut:



Gambar 2. Flowchart Langkah-langka K-means

3.1. Melakukan Data Cleaning

Langkah ini merupakan tahapan dari proses pengoreksian, pengurangan dari data sebelumnya atau menghilangkan sebagian data yang terduplikasi, menghapus data yang salah, serta memilah data mana yang dianggap tidak konsisten serta memverifikasi kesalahan yang ada pada data. Maksud dari di lakukan data *cleaning* yaitu untuk mempermudah proses perhitungan algoritma *k-means*. Hasil dari pembersihan data dapat dilihat pada Tabel 2 berikut ini:

Tabel 2. Sampel Data Cleaning Rekap Data per Satuan Pendidikan

No	Nama Satuan Pendidikan	Jumlah Rombel	Jumlah Peserta Didik	Tenaga Guru & Tendik
1	SMP Bhakti Nusa	3	9	10
2	SMP Budi Daya Pulau Palas	4	76	9
3	SMPN 1 Enok	6	131	13
4	SMPN 1 Gaung	16	332	38
5	SMPN 1 GAS	8	162	22
6	SMPN 1 Kuindra	6	136	12
7	SMPN 1 Pulau Burung	9	222	18
8	SMPN 1 Tembilahan Hulu	33	1014	61
9	SMPN 3 Tembilahan Hulu	18	369	54
10	SMPS IT Daarul Rahman	11	427	15

3.2. Menginput Data yang Sudah dibersihkan

Tahap pemasukan data diterapkan pada waktu sistem melakukan data *cleaning*. Data tersebut diperuntukan pada tahap penginputan. Data training yang digunakan penelitian ini adalah data yang sudah dilakukan pembersihan data sampel yaitu 10 data sekolah.

3.3 Menentukan Jumlah Cluster K

Jumlah *cluster* pada penelitian ini sebanyak 3 *cluster* yang dimana, untuk ke-3 *cluster* ini akan memastikan hasil penelitian mengenai sekolah mana yang masuk dalam kategori kurang diminati, cukup diminati, dan sangat diminati terhadap pemilihan sekolah tingkat SMP di Kabupaten Indragiri Hilir.

3.4 Menentukan Titik Centroid

Penentuan titik pusat atau *centroid* pada tahap ini dilakukan dengan cara mengambil data secara random. Adapun Peneliti mengambil nilai dengan kateogri rendah, nilai menengah, dan nilai tertinggi pada jumlah rombel, jumlah peserta didik serta jumlah tenaga pendidik dan tendik pada tahun 2019 di Dinas Pendidikan Kabupaten Indragiri Hilir, dengan data *centroid* sebagai berikut:

Tabel 3. Sampel *Centroid* Acak

<i>Centroid</i>	Nama Satuan Pendidikan	Jumlah Rombel	Jumlah Peserta Didik	Tenaga Guru & Tendik
C0	SMPN 1 Enok	6	131	13
C1	SMPN 1 Gaung	16	332	38
C2	SMPN 1 Tembilahan Hulu	33	1014	61

3.5 Menghitung Jarak Centroid Terdekat

Formula yang diterapkan untuk mengkalkulasi jarak ketiap-tiap *centroid* yaitu menerapkan formula korelasi antara tiga objek (*Euclidean distance*), sehingga dengan formula ini didapatkan jarak nilai yang paling dekat dari setiap data dengan masing-masing mendapatkan kesamaan data *centroid*. perhitungan sebagai berikut :

- Perulangan Tahap pertama atau iterasi ke-1
 - Menghitung *cluster* 0 (C0)

Proses kalkulasi diambil dari data sekolah nomor urut 1 dan nilai *centroid* C0, yaitu:

Tabel 4 Sampel Data Perhitungan Iterasi 1

No	Nama Satuan Pendidikan	Jumlah Rombel	Jumlah Peserta Didik	Tenaga Guru & Tendik	C0	C1	C2	Jarak Terdekat	C0	C1	C2
1	SMP Bhakti Nusa	3	9	10	122	324	1.007	324	1		
2	SMP Budi Daya Pulau Palas	4	76	9	55	258	940	258	1		
3	SMP Negeri 1 Enok	6	131	13	0	203	885	203	1		
4	SMP Negeri 1 Gaung	16	332	38	203	0	683	0		1	
5	SMP Negeri 1 Gaung Anak Serka	8	162	22	32	171	853	171	1		
6	SMP Negeri 1 Kuindra	6	136	12	5	198	880	198	1		
7	SMP Negeri 1 Pulau Burung	9	222	18	91	112	794	112	1		
8	SMP Negeri 1 Tembilahan Hulu	33	1014	61	885	683	0	0			1
9	SMP Negeri 3 Tembilahan Hulu	18	369	54	242	40	645	40		1	
10	SMPS I T Daarul Rahman	11	427	15	296	98	589	98		1	
									6	3	1

Tabel 4 diatas merupakan perolehan klasterisasi tiap-tiap data sampel pada setiap sekolah dengan menerapkan formula *euclidean distance*. Mengenai perolehan *cluster* 0 berjumlah 6 Sekolah, untuk *cluster* 1 berjumlah 3 sekolah, dan untuk *cluster* 2 berjumlah 1 sekolah.

3.7 Menentukan Titik Centroid

Tahapan berikutnya adalah menemukan nilai *centroid baru* dengan mengkalkulasi dari nilai rata rata pada iterasi pertama yang sebelumnya sudah di kelompokkan. Adapun kalkulasi nilai untuk *centroid* baru sebagai berikut:

$$S_1 = \sqrt{(3-6)^2 + (9-131)^2 + (10-13)^2}$$

$$S_1 = 122$$

- Mengkalkulasi *cluster* 1 (C1)

Proses kalkulasi diambil dari data sekolah nomor urut 1 dan nilai *centroid* C1, yaitu:

$$S_1 = \sqrt{(3-16)^2 + (9-332)^2 + (10-38)^2}$$

$$S_1 = 324$$

- Mengkalkulasi *cluster* 2 (C2)

Proses kalkulasi diambil dari data sekolah nomor urut 1 dan nilai *centroid* C2, yaitu:

$$S_1 = \sqrt{(3-33)^2 + (9-1014)^2 + (10-61)^2}$$

$$S_1 = 1.007$$

3.6 Mengelompokkan data masing-masing *cluster*

Ketika proses kalkulasi selesai pada proses selanjutnya adalah mengkalkulasi jarak *centroid* terdekat sebelumnya, jadi data yang dikelompokkan memiliki kecocokan dengan *cluster* tiap-tiap data, seperti yang terlihat pada Tabel 4 dibawah ini.

- Cluster* 0 (C0) berjumlah 6 data

$$C0(A1) = (3+4+6+8+6+9) / 6$$

$$C0(A1) = 6$$

$$C0(A2) = (9+76+131+162+136+222) / 6$$

$$C0(A2) = 123$$

$$C0(A3) = (10+9+13+22+12+18) / 6$$

$$C0(A3) = 14$$

- Cluster* 1 (C1) berjumlah 3 data

$$C1(A1) = (16+18+11) / 3$$

$$C1(A1) = 15$$

$$C1(A2) = (332+369+427) / 3$$

$$C1(A2) = 376$$

$$C1(A3) = (38+54+15) / 3$$

$$C1(A1) = 36$$

3. Cluster 2 (C2) berjumlah 1 data

$$C2(A1) = (33) / 1$$

$$C2(A1) = 33$$

$$C2(A2) = (1.014) / 1$$

$$C2(A2) = 1.014$$

$$C2(A3) = (61) / 1$$

$$C2(A3) = 61$$

Proses kalkulasi yang sudah dilakukan, maka diperoleh nilai *centroid* baru yang selanjutnya di terapkan untuk proses iterasi kedua. Data *centroid* baru dapat dilihat pada Tabel 5 berikut ini:

Tabel 5. *Centroid* Baru

<i>Centroid</i>	Jumlah Rombel	Jumlah Peserta Didik	Tenaga Guru & Tendik
C0	6	123	14
C1	15	376	36
C2	33	1014	61

3.8 Menghitung Jarak Centroid Terdekat

Formula yang dimplementasikan untuk mengkalkulasi masing-masing *centroid* dapat dikalkulasikan pada jarak setiap data adalah menggunakan formula korelasi antara tiga objek (Euclidean distance), sehingga formula ini dapat menentukan jarak yang sangat terdekat untuk setiap data dengan masing-masing

Tabel 6 Sampel Data Perhitungan Iterasi 2

No	Nama Satuan Pendidikan	Jumlah Rombel	Jumlah Peserta Didik	Tenaga Guru & Tendik	C0	C1	C2	Jarak Terdekat	C0	C1	C2
1	SMP Bhakti Nusa	3	9	10	114	368	1.007	165	1		
2	SMP Budi Daya Pulau Palas	4	76	9	47	301	940	98	1		
3	SMPN 1 Enok	6	131	13	8	246	885	43	1		
4	SMPN 1 Gaung	16	332	38	211	44	683	44		1	
5	SMPN 1 GAS	8	162	22	40	215	853	12	1		
6	SMPN 1 Kuindra	6	136	12	13	241	880	38	1		
7	SMPN 1 Pulau Burung	9	222	18	99	155	794	49	1		
8	SMPN 1 Tembilahan Hulu	33	1014	61	893	639	0	0			1
9	SMPN 3 Tembilahan Hulu	18	369	54	250	20	645	20		1	
10	SMPS IT Daarul Rahman	11	427	15	304	55	589	55		1	
Jumlah									6	3	1

Tabel diatas menjabarkan hasil dari pengelompokan pada iterasi kedua, sehingga dapat disimpulkan untuk data iterasi kedua tidak terdapat perubahan untuk setiap *cluster* atau hasil iterasi kedua sama dengan hasil iterasi pertama. Dengan begitu proses iterasi hanya dilakukan sampai iterasi kedua dengan hasil *cluster* 0 berjumlah 6 Sekolah, *cluster* 1 berjumlah 3 sekolah, dan *cluster* 2 berjumlah 1 sekolah. Hasil untuk setiap *cluster* dapat dilihat pada masing-masing tabel di bawah ini.

meperoleh kesamaan data *centroid*. *Centroid* baru ini diperoleh dari hasil kalkulasi sebelumnya digunakan untuk Proses kalkulasi jarak masing-masing data ketitik pusat *cluster* seperti berikut ini:

2. Proses iterasi 2

a. Mengkalkulasi *cluster* 0 (C0)

Proses kalkulasi diambil dari data sekolah nomor urut 1 dan nilai *centroid* baru C0, yaitu:

$$S_1 = \sqrt{(3-6)^2 + (9-123)^2 + (10-14)^2}$$

$$S_1 = 114$$

b. Mengkalkulasi *cluster* 1 (C1)

Proses kalkulasi diambil dari data sekolah nomor urut 1 dan nilai *centroid* baru C1, yaitu:

$$S_1 = \sqrt{(3-15)^2 + (9-376)^2 + (10-36)^2}$$

$$S_1 = 368$$

c. Mengkalkulasi *cluster* 2 (C2)

Proses kalkulasi diambil dari data sekolah nomor urut 1 dan nilai *centroid* baru C2, yaitu:

$$S_1 = \sqrt{(3-33)^2 + (9-1014)^2 + (10-61)^2}$$

$$S_1 = 1007$$

3.9 Mengelompokkan data tiap-tiap *cluster*

Ketika tahapan kalkulasi telah dilakukan untuk menghitung jarak *centroid* terdekat sebelumnya, menerapkan hasil dari *centroid* baru selanjutnya, data dikelompokkan sesuai dengan *cluster* masing-masing, seperti yang dapat dilihat pada Tabel 6 di bawah ini

Tabel 7. Cluster 0 Kategori Kurang Diminati

No	Nama Satuan Pendidikan	Jumlah Rombe	Jumlah Peserta Didik	Tenaga Guru & Tendik
1	SMP Bhakti Nusa	3	9	10
2	SMP Budi Daya Pulau Palas	4	76	9
3	SMPN 1 ENOK	6	131	13
4	SMPN 1 GAS	8	162	22
5	SMPN 1 Kuindra	6	136	12
6	SMPN 1 Pulau Burung	9	222	18
Jumlah		6	123	14

Tabel 7 diatas menjabarkan hasil *cluster* untuk kategori sekolah kurang diminati oleh siswa di Kabupaten Indragiri Hilir dengan jumlah 6 sekolah.

Tabel 8. Cluster 1 Kategori Cukup Diminati

No	Nama Satuan Pendidikan	Jumlah Rombel	Jumlah Peserta Didik	Tenaga Guru & Tendik
1	SMPN 1 GAUNG	16	332	38
2	SMPN 3 TEMBILAHAN HULU	18	369	54
3	SMPS IT DAARUL RAHMAN	11	427	15
Jumlah		15	376	36

Tabel 8 diatas menjabarkan hasil *cluster* untuk kategori sekolah cukup diminati oleh siswa di Kabupaten Indragiri Hilir dengan jumlah 3 sekolah.

Tabel 9. Cluster 2 Kategori Sangat diminati

No	Nama Satuan Pendidikan	Jumlah Rombel	Jumlah Peserta Didik	Tenaga Guru & Tendik
1	SMPN 1 TEMBILAHAN HULU	33	1.014	61
Jumlah		33	1.014	61

Tabel 9 diatas menjabarkan hasil *cluster* untuk kategori sekolah sangat diminati oleh siswa di Kabupaten Indragiri Hilir dengan jumlah 1 sekolah.

3. Kesimpulan

Algoritma *K-Means Clustering* yang digunakan sebagai suatu untuk mengklasterisasikan minat Siswa terhadap pemilihan sekolah. Dari 10 data sekolah pada tingkat SMP yang telah oleh didapatkan hasil untuk kategori kurang diminati (C0) sebanyak 6 sekolah, kategori cukup diminati (C1) sebanyak 3 sekolah, dan kategori sangat diminati (C2) sebanyak 1 sekolah.

Daftar Rujukan

[1] Gustientiedina, G., Adiya, M. H., & Desnelita, Y. (2019). Penerapan Algoritma K-Means untuk Clustering Data Obat-Obatan. *Jurnal Nasional Teknologi Dan Sistem Informasi*, 5(1), 17–24. DOI: <https://doi.org/10.25077/teknosi.v5i1.2019.17-24>

[2] Mar'i, F., & Supianto, A. A. (2018). Clustering Credit Card Holder Berdasarkan Pembayaran Tagihan Menggunakan Improved K-Means dengan Particle Swarm Optimization. *Jurnal*

Teknologi Informasi Dan Ilmu Komputer, 5(6), 737. DOI: <https://doi.org/10.25126/jtiik.201856858>.

[3] Sari, D. W., & Ayu, K. G. (2019). Penentuan Kriteria Dalam Memilih Sekolah Dasar dengan Menerapkan K-Means Clustering (Studi Kasus : Wilayah Kecamatan Mampang). *Sainstech* 29(2), 24–28.

[4] Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, 4(1), 51. DOI: <https://doi.org/10.30865/mib.v4i1.1784>.

[5] Rosmini, R., Fadlil, A., & Sunardi, S. (2018). Implementasi Metode K-Means dalam Pemetaan Kelompok Mahasiswa Melalui Data Aktivitas Kuliah. *It Journal Research And Development*, 3(1), 22–31. DOI: [https://doi.org/10.25299/itjrd.2018.vol3\(1\).1773](https://doi.org/10.25299/itjrd.2018.vol3(1).1773).

[6] Ridho, F., & Kusuma, A. A. (2019). Deteksi Intrusi Jaringan dengan K-Means Clustering pada Akses Log dengan Teknik Pengolahan Big Data. *Jurnal Aplikasi Statistika & Komputasi Statistik*, 10(1), 53. DOI: <https://doi.org/10.34123/jumalasks.v10i1.202>.

[7] Indriyani, Fintri., & Irfiani, Eni. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA : Jurnal Informatika*, 7(2), 109. DOI: <https://doi.org/10.30595/juita.v7i2.5529>.

[8] Putra, R. R., & Wadisman, C. (2018). Implementasi Data Mining Pemilihan Pelanggan Potensial Menggunakan Algoritma K Means. *INTECOMS: Journal of Information Technology and Computer Science*, 1(1), 72–77. DOI: <https://doi.org/10.31539/intecom.v1i1.141>.

[9] Rustam, S., Santoso, H. A., & Supriyanto, C. (2018). Optimasi K-Means Clustering Untuk Identifikasi Daerah Endemik Penyakit Menular Dengan Algoritma Particle Swarm Optimization Di Kota Semarang. *ILKOM Jurnal Ilmiah*, 10(3), 251–259. DOI: <https://doi.org/10.33096/ilkom.v10i3.342.251-259>.

[10] Parlina, I., Windarto, A. P., Wanto, A., & Lubis, M. R. (2018). Memanfaatkan Algoritma K-Means Dalam Menentukan Pegawai Yang Layak Mengikuti Assesment Center Untuk Clustering Program SDP. *Computer Engineering, Science and System Journal*, 3(1), 87. DOI: <https://doi.org/10.24114/cess.v3i1.8192>.

[11] Maulida, L. (2018). Penerapan Datamining Dalam Mengelompokkan Kunjungan Wisatawan Ke Objek Wisata Unggulan Di Prov. Dki Jakarta Dengan K-Means. *JISKA (Jurnal Informatika Sunan Kalijaga)*, 2(3), 167. DOI: <https://doi.org/10.14421/jiska.2018.23-06>.

[12] Nalendra, A. K., Mujiono, M., Akhsani, R., & Utama, A. S. W. (2020). Implementasi Algoritma K-Mean dalam Pengelompokan Data Kecelakaan di Kabupaten Kediri. *VOCATECH: Vocational Education and Technology Journal*, 1(2), 53–60. DOI: <https://doi.org/10.38038/vocatech.v1i2.28>.

[13] Priambodo, Y. A., & Prasetyo, S. Y. J. (2018). Pemetaan Penyebaran Guru di Provinsi Banten dengan Menggunakan Metode Spatial Clustering K-Means (Studi kasus : Wilayah Provinsi Banten). *Indonesian Journal of Computing and Modeling*, 1(1), 18–27. DOI: <https://doi.org/10.24246/j.icm.2018.v1.i1.p18-27>.

[14] Ependi, U., & Putra, A. (2019). Solusi Predeksi Persediaan Barang dengan Menggunakan Algoritma Apriori (Studi Kasus: Regional Part Depo Auto 2000 Palembang). *JEPIN (Jurnal Edukasi & Penelitian Informatika)*, 5(2).

[15] Nastuti, A., & Harahap, S. Z. (2019). Teknik Data Mining untuk Penentuan Paket Hemat Sembako dan Kebutuhan Harian dengan Menggunakan Algoritma Fp-Growth (Studi Kasus di Ulfamart Lubuk Alung). *Jurnal Ilmiah Fakultas Sains dan Teknologi*, 7(3), 111-119.

- [16] Muhammad I. (2019). Klasterisasi Data Jamaah Umroh Pada Auliya Tour & Travel Menggunakan Metode K-Means Clustering. *JURTEKSI (Jurnal Teknologi dan Sistem Informasi)*, 5(2), 97-104.
- [17] Silalahi, M. (2018). Analisis Clustering Menggunakan Algoritma K-Means Terhadap Penjualan Produk Pada PT Batamas Niaga Jaya. *Computer Based Information System Journal*, 6(2). DOI: <https://doi.org/10.33884/cbis.v6i2.709>.
- [18] Indriyani, F., & Irfiani, E. (2019). Clustering Data Penjualan pada Toko Perlengkapan Outdoor Menggunakan Metode K-Means. *JUITA: Jurnal Informatika*, 7(2), 109-114. DOI: <https://doi.org/10.30595/juita.v7i2.5529>.
- [19] Fatmawati, K., & Windarto, A. P. (2018). Data Mining : Penerapan Rapidminer Dengan K-Means Cluster Pada Daerah Terjangkit Demam Berdarah Dengue (Dbd) Berdasarkan Provinsi CESS (*Journal of Computer Engineering System and Science*), 3(2), 173-178. DOI: <https://doi.org/10.24114/cess.v3i2.9661>.
- [20] Kelebihan Algoritma k-means clustering [online] (diperbaharui 21 Januari 2017) <https://id.wikipedia.org/wiki/K-means> [Diakses 11 Mei 2020].