

## **Data Mining Pengelompokan Nilai untuk Menentukan Minat Mata Pelajaran pada Siswa SMA Negeri 1 Kota Gajah**

**Irfan Setiawan<sup>1\*)</sup>, Heri Nurdiyanto<sup>2)</sup>**

<sup>1)2)</sup> Teknik Informatika, STMIK Dharma Wacana

**Correspondence author:** [irfansetiawan005@gmail.com](mailto:irfansetiawan005@gmail.com), Kota Metro, Indonesia

**DOI:** <https://doi.org/10.37012/jtik.v9i1.1352>

### **Abstrak**

Pilihan minat mata pelajaran Sekolah Menengah Atas merupakan usaha peningkatan kualitas pendidikan yang langsung berkenaan dengan siswa sebagai salah satu bagian dari peningkatan sumber daya manusia (SDM). Pemilihan minat mata pelajaran yang dilakukan selama ini mempunyai banyak kelemahan, antara lain berdasarkan keinginan siswa tanpa melihat latar belakang nilai akademisnya dan tidak ada acuan yang jelas untuk melakukan pengelompokan data nilai siswa. Untuk menyelesaikan permasalahan tersebut, maka dibuat sebuah sistem pengelompokan nilai siswa pada pilihan minat mata pelajaran dengan menerapkan salah satu metode *Clustering*. Algoritma yang digunakan adalah Algoritma K-Means untuk mengidentifikasi objek yang serupa dengan memperhatikan beberapa kriteria. Dengan demikian hasil dari pengelompokan yang telah dirancang akan membantu pihak SMAN 1 Kota Gajah dalam proses pengelompokan nilai siswa untuk pilihan minat mata pelajaran yang sesuai dengan kriteria, sehingga pengelompokan dapat dilakukan lebih cepat, tepat dan akurat serta terhindar dari kesalahan.

**Kata Kunci:** *Data Mining, K-means, Clustering*

### **Abstract**

*The choice of interest in senior high school subjects is an effort to improve the quality of education that is directly related to students as a part of increasing human resources (HR). The selection of subject interests that has been carried out so far has many weaknesses, including based on the wishes of students without looking at the background of their academic scores and there is no clear reference for grouping student value data. To solve this problem, a system of grouping student scores was created in the choice of subject interests by applying one of the clustering methods. The algorithm used is the K-Means Algorithm to identify similar objects by taking into account several criteria. Thus the results of the grouping that has been designed will help SMAN 1 Kota Gajah in the process of grouping student scores for the choice of subject interests that match the criteria, so that grouping can be done more quickly, precisely and accurately and avoid mistakes.*

**Keywords:** *Data Mining, K-means, Clustering*

## **PENDAHULUAN**

Penentuan minat mata pelajaran Sekolah Menengah Atas merupakan usaha peningkatan kualitas pendidikan yang langsung berkenaan dengan siswa sebagai salah satu bagian dari Kurikulum Merdeka Belajar dan untuk meningkatkan SDM. Sesuai kurikulum yang berlaku di seluruh Indonesia, maka siswa kelas X SMA yang naik ke kelas XI akan memilih minat mata pelajaran sesuai bakat siswa. Penentuan minat mata pelajaran siswa ini

bertujuan untuk mengarahkan peserta didik agar lebih fokus mengembangkan kemampuan diri dan minat yang dimiliki (Syahra et al., 2019). Keputusan memilih mata pelajaran yang tidak tepat bisa sangat merugikan siswa dan karirnya di masa mendatang. Dengan penentuan tersebut diharapkan dapat memaksimalkan potensi, bakat atau talenta individu, sehingga juga akan memaksimalkan nilai akademisnya. Penentuan minat mata pelajaran yang dilakukan selama ini mempunyai banyak kelemahan, antara lain berdasarkan keinginan siswa tanpa melihat latar belakang nilai akademisnya dan tidak ada acuan yang jelas untuk melakukan pengelompokan data siswa. Sehingga Mata Pelajaran yang dipilih terkadang menjadi masalah bagi siswa di kemudian hari (Putra & Wibowo, 2020).

*Data Mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis (Nas, 2021). Dalam *Data Mining* terdapat beberapa teknik salah satunya adalah klustering. Klustering digunakan untuk menganalisis pengelompokan data nilai siswa yang memiliki karakteristik tertentu. Algoritma yang akan diterapkan dari klustering ini adalah K-Means.

## METODE

Menurut (Hermawati, 2013:3) "*Data Mining* adalah proses yang mempekerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis." *Data Mining* berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang (Pasaribu, 2021).

Algoritma *K-Means* merupakan metode analisis kelompok yang mengarah pada partisi N objek pengamatan ke dalam kelompok (*cluster*) dimana setiap objek pengamatan dimiliki oleh sebuah kelompok dengan *mean* (rata-rata) terdekat, mirip dengan algoritma *Expectation-Maximization* untuk *Gaussian Mixture* di mana keduanya mencoba untuk menemukan pusat dari kelompok dalam data sebanyak iterasi perbaikan yang dilakukan oleh kedua algoritma (Yulianti et al., 2019).

Berikut ini adalah langkah-langkah algoritma K-Means:

1. Tentukan K sebagai jumlah *cluster* yang dibentuk.
2. Tentukan pusat (*centroid*) *cluster* awal. digunakan rumus sebagai berikut:

<http://journal.thamrin.ac.id/index.php/jtik/article/view/1352/pdf>

$$C_i = \frac{1}{M} \sum_{j=1}^M X_j$$

Keterangan:

$C_i$  : *Centroid* pada *cluster*

$x_j$  : Objek ke- $j$

$M$  : Banyaknya objek / jumlah objek yang menjadi anggota *cluster*

3. Perhitungan Untuk mengukur jarak antar data dengan pusat *cluster* digunakan *Euclidian Distance*. Rumusnya adalah:

$$D(x_2, x_1) = ||x_2 - x_1||_2 = \sqrt{\sum_{j=1}^p |x_2 - x_{1j}|^2}$$

Keterangan:

$x_2$  : Data

$x_1$  : Titik pusat (*centroid*)

$D(x_2, x_1)$  : *Euclidian Distance* yaitu jarak antara data  $x_2$  dan  $x_1$

4. Pengelompokkan data

Setelah sejumlah populasi data tersebut menemukan kedekatan dengan salah satu *centroid* yang ada maka secara otomatis populasi data tersebut masuk kedalam kelas yang memiliki *centroid* yang bersangkutan.

5. Lakukan *iterasi*, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan pertama.
6. Ulangi langkah 3 apabila masih ada data yang berpindah kelompok, atau ada perubahan nilai *centroid* diatas nilai ambang yang ditentukan, atau juga apabila perubahan nilai pada fungsi objektif yang digunakan masih diatas nilai ambang yang ditentukan.
7. Tentukan  $K$  sebagai jumlah *cluster* yang dibentuk.
8. Tentukan pusat (*centroid*) *cluster* awal.

Dalam menentukan  $n$  buah pusat *cluster* awal dilakukan pembangkitan bilangan *random* yang merepresentasikan urutan data input. Pusat awal cluster didapatkan dari data sendiri bukan dengan menentukan titik baru, yaitu dengan mengacak (*random*)

pusat awal dari data. Kemudian untuk menghitung *centroid cluster* ke-I berikutnya digunakan rumus sebagai berikut:

$$C_i = \frac{1}{M} \sum_{j=1}^M X_j$$

Keterangan:

$C_i$  : *Centroid* pada *cluster*

$x_j$  : Objek ke-j

M : Banyaknya objek / jumlah objek yang menjadi anggota *cluster*

#### 9. Perhitungan

Untuk mengukur jarak antar data dengan pusat *cluster* digunakan *Euclidian Distance*. Jarak hasil perhitungan akan dilakukan perbandingan dan dipilih jarak terdekat antar data dengan pusat *cluster*, jarak ini menggunakan asumsi bahwa data tersebut berada dalam satu kelompok dengan pusat *cluster* terdekat. Rumusnya adalah:

$$D(x_2, x_1) = ||x_2 - x_1||_2 = \sqrt{\sum_{j=1}^p |x_2 - x_{1j}|^2}$$

Keterangan :

$x_2$  : Data

$x_1$  : titik pusat (*centroid*)

$D(x_2, x_1)$ : *Euclidian Distance* yaitu jarak antara data  $x_2$  dan  $x_1$

#### 10. Pengelompokkan data

Setelah sejumlah populasi data tersebut menemukan kedekatan dengan salah satu *centroid* yang ada maka secara otomatis populasi data tersebut masuk kedalam kelas yang memiliki *centroid* yang bersangkutan.

11. Lakukan *iterasi*, kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan pertama.

12. Ulangi langkah 3 apabila masih ada data yang berpindah kelompok, atau ada perubahan nilai *centroid* diatas nilai ambang yang ditentukan, atau juga apabila perubahan nilai pada fungsi objektif yang digunakan masih diatas nilai ambang yang ditentukan.

## HASIL DAN PEMBAHASAN

Analisa permasalahan bertujuan untuk mengetahui dan mengamati permasalahan yang sedang terjadi, kemudian merumuskan sebuah solusi untuk menyelesaikan permasalahan yang ada. Untuk jurusan yang sesuai dengan minat, maka diperlukan informasi dan pengetahuan tentang informasi nilai rata-rata bidang studi siswa.

**Tabel 1.** Data Nilai Rata-rata Siswa

No	Nama Siswa	Nilai Rata-Rata IPA	Nilai Rata-Rata IPS	Nilai Tes Ujian Umum
1	Gusti Ayu Putu Widyastuti	80	80	88
2	Fresa Dwi Safitri	75	75	70
3	Rosiana Maharani	81	88	91
4	Ibnu Nurzaha	85	80	85
5	Shafa Delaila Azzahra	72	78	86
6	Aqila Weninggalih	80	88	77
7	Salwa Putri Riri Kinasih	81	91	92
8	Aditya Apdillah	90	87	91
9	Putri Ayu Pratiwi	79	81	80
10	Chlsea Raditya Putri	78	92	80
11	Medina Lingga Cahyani	91	80	95
12	Ernita Adzkia Afida	80	88	89
13	Kevin Fitrah Choirizaldi	70	78	85
14	Refal Rahmatan Gustian	92	93	90
15	Noval Syahputra	92	82	93
16	Putri Aprilia	77	88	80
17	Arum Dw Ana	80	78	80
18	Selli Setiawati	76	70	91
19	Abian Faiha Wibisono	85	73	78
20	Inaya Syafa Shabilla	93	80	91
21	Sabrina Putri	81	82	80
22	Reisha Alfira Kusuma	78	87	80
23	Ivan Adika Ramadan	86	90	87
24	Hafidz Al Juri	91	88	93
25	Nova Fardina Putri	75	77	80
..				
..				
425	Dimas Nabil Al-Hafizh	80	78	93
426	Tina Septiani	78	84	80

Berikut langkah-langkah *clustering* menggunakan algoritma K-means.

1. Tentukan jumlah *cluster* yang diinginkan (misalkan = 4)
2. Pilih *centroid* awal secara acak: pada langkah ini secara acak akan dipilih 4 buah data sebagai *centroid*,misalnya:

data {40,20,14,10}

M1 = (94,86,85), M2 = (93,80,91), M3 = (92,93,90), M4 = (78,92,80)

3. Hitung jarak dengan *centroid* .....(iterasi 1)

Pada langkah ini setiap data akan ditentukan *centroid* terdekatnya, dan data tersebut akan ditetapkan sebagai anggota kelompok yang terdekat dengan *centroid*. Untuk menghitung jarak ke *centroid* masing-masing *cluster* pada siswa sbb:

Data: (1,3), centroid M1=(94,86,85), M2=(93,80,91), M3=(92,93,90), M4=(78,92,80)

$$DM1 = \sqrt{(80 - 94)^2 + (80 - 86)^2 + (88 - 85)^2} = 15,52$$

$$DM2 = \sqrt{(80 - 93)^2 + (80 - 80)^2 + (88 - 91)^2} = 26,59$$

$$DM3 = \sqrt{(80 - 92)^2 + (80 - 93)^2 + (88 - 90)^2} = 14,46$$

$$DM4 = \sqrt{(80 - 78)^2 + (80 - 92)^2 + (88 - 80)^2} = 10,82$$

Pada berikutnya dihitung pula rasio antara besaran BCV (*Between Cluster Variation*) dengan WCV (*Within Cluster Variation*) :

$$d(m1.m2) = \sqrt{(94 - 93)^2 + (86 - 80)^2 + (85 - 91)^2} = 8,54$$

$$d(m1.m3) = \sqrt{(94 - 92)^2 + (86 - 93)^2 + (85 - 90)^2} = 8,83$$

$$d(m1.m4) = \sqrt{(94 - 78)^2 + (86 - 92)^2 + (85 - 80)^2} = 17,80$$

$$d(m2.m3) = \sqrt{(93 - 92)^2 + (80 - 93)^2 + (91 - 90)^2} = 13,08$$

$$d(m2.m4) = \sqrt{(93 - 78)^2 + (80 - 92)^2 + (91 - 80)^2} = 22,14$$

$$d(m3.m4) = \sqrt{(92 - 78)^2 + (93 - 92)^2 + (90 - 80)^2} = 17,23$$

$$BCV = d(m1.m2) + d(m1.m3) + d(m1.m4) + d(m2.m3) + d(m2.m4) + d(m3.m4) \\ = 8,54 + 8,83 + 17,80 + 13,08 + 22,14 + 17,23 = 87,63$$

Hitung nilai WCV (*Within Cluster Variation*) dengan cara memangkatkan jarak terdekat *cluster* dan menjumlahkan setiap nilai WCV.

$$WCV = 13,34^2 + 19,95^2 + 12,08^2 \dots + 8,00^2$$

$$WCV = 14908$$

$$\text{Sehingga besar rasio} = BCV/WCV = 87,63/14908 = 0.005$$

Tabel hasil perhitungan jarak selengkapnya antara masing-masing data dengan *centroid*:

**Tabel 2.** Hasil Perhitungan Jarak Data dengan *Centroid*

No	Nama Siswa	Pusat M1	Pusat M2	Pusat M3	Pusat M4	Jarak Terdekat	Klaster
1	Gusti Ayu Putu Widyastuti	15.52	13.34	17.80	14.56	13.34	2
2	Fresa Dwi Safitri	26.59	28.11	31.83	19.95	19.95	4
3	Rosiana Maharani	14.46	14.42	12.12	12.08	12.08	4
4	Ibnu Nurzeha	10.82	10.00	15.59	14.76	10.00	2
5	Shafa Delaila Azzahra	23.43	21.68	25.32	16.37	16.37	4
6	Aqila Weninggalih	16.25	20.71	18.38	5.39	5.39	4
7	Salwa Putri Riri Kinasih	15.59	16.31	11.36	12.41	11.36	3
8	Aditya Apdillah	7.28	7.62	6.40	17.03	6.40	3
9	Putri Ayu Pratiwi	16.58	17.83	20.32	11.05	11.05	4
10	Chlsea Raditya Putri	17.80	22.14	17.23	0.00	0.00	4
11	Medina Lingga Cahyani	12.04	4.47	13.96	23.19	4.47	2
12	Ernita Adzkaa Afida	14.70	15.39	13.04	10.05	10.05	4
13	Kevin Fitrah Choirizaldi	25.30	23.85	27.09	16.88	16.88	4
14	Refal Rahmatan Gustian	8.83	13.08	0.00	17.23	0.00	3
15	Noval Syahputra	9.17	3.00	11.40	21.56	3.00	2
16	Putri Aprilia	17.83	21.00	18.71	4.12	4.12	4
17	Arum Dw Ana	16.88	17.15	21.66	14.14	14.14	4
18	Selli Setiawati	24.82	19.72	28.04	24.68	19.72	2
19	Abian Faiha Wibisono	17.29	16.79	24.35	20.35	16.79	2
20	Inaya Syafa Shabilla	8.54	0.00	13.08	22.14	0.00	2
21	Sabrina Putri	14.49	16.40	18.49	10.44	10.44	4
22	Reisha Alfira Kusuma	16.79	19.87	18.22	5.00	5.00	4
23	Ivan Adika Ramadan	9.17	12.85	7.35	10.82	7.35	3
24	Hafidz Al Juri	8.77	8.49	5.92	18.81	5.92	3
-							
-							
425	Dimas Nabil Al-Hafizh	14.70	15.39	13.04	10.05	10.05	1
426	Tina Septiani	25.30	23.85	27.09	16.88	16.88	2

Karena langkah ini merupakan iterasi 1 maka lanjutkan kelangkah berikutnya. Pusat *cluster* baru ditentukan berdasarkan pengelompokkan anggota dari masing-masing *cluster*. Ulangi iterasi yang dimulai dari langkah 1, sampai rasio tetap atau *cluster* tidak mengalami

perubahan. Disini sampai dengan iterasi 4, karena bila dibandingkan maka rasio sekarang (0,014) sudah tidak lagi lebih besar dari rasio sebelumnya (0,014) oleh karena itu iterasi dihentikan.

Dari hasil klastering diatas dapat diambil kesimpulan bahwa:

1. Kelompok siswa pada klaster 1 merupakan kelompok siswa yang memilih minat mata pelajaran ilmu alam dan memilih lintas minat pelajaran ekonomi dengan jumlah 195 siswa
2. Kelompok siswa pada klaster 2 merupakan kelompok siswa yang memilih minat mata pelajaran ilmu alam dan memilih lintas minat pelajaran sejarah dengan jumlah 130 siswa
3. Kelompok siswa pada klaster 3 merupakan kelompok siswa yang memilih minat mata pelajaran ilmu sosial dan memilih lintas minat pelajaran fisika dengan jumlah 60 siswa
4. Kelompok siswa pada klaster 4 merupakan kelompok siswa yang memilih minat mata pelajaran ilmu sosial dan memilih lintas minat pelajaran Biologi dengan jumlah 42 siswa

## **KESIMPULAN DAN REKOMENDASI**

Berdasarkan hasil dari perancangan aplikasi penerapan *Data Mining* dalam pengelompokkan data nilai siswa untuk penentuan minat mata pelajaran menggunakan algoritma *K-Means*, maka dapat diperoleh beberapa kesimpulan. Adapun kesimpulan tersebut adalah sebagai berikut :

1. *Data mining* dengan algoritma *K-Means* dapat diterapkan pada SMAN 1 Kota Gajah untuk menganalisis permasalahan yang ada yang berkenaan dengan pengelompokkan data nilai siswa untuk penentuan minat mata pelajaran.
2. Aplikasi dirancang dengan mengadopsi algoritma *K-Means* yang dapat digunakan dalam pengelompokkan data nilai siswa untuk penentuan minat mata pelajaran pada SMAN 1 Kota Gajah.
3. Metode yang dirancang dapat mengimplementasikan dan dapat dijadikan sebagai solusi pemecahan masalah dalam hal pengelompokan data nilai siswa.



---

## REFERENSI

- Nas, C. (2021). Data Mining Prediksi Minat Calon Mahasiswa Memilih Perguruan Tinggi Menggunakan Algoritma C4.5. *Jurnal Manajemen Informatika (JAMIKA)*, 11(2), 131–145.
- Pasaribu, A. F. O. (2021). Analisis Pola Menggunakan Metode C4.5 Untuk Peminatan Jurusan Siswa Berdasarkan Kurikulum (Studi Kasus: Sman 1 Natar). *Jurnal Teknologi Dan Sistem Informasi*, 2(1), 80–85.
- Putra, D., & Wibowo, A. (2020). Prediksi Keputusan Minat Penjurusan Siswa SMA Yadika 5 Menggunakan Algoritma Naïve Bayes. *Prosiding Seminar Nasional Riset Information Science (SENARIS)*, 2, 84–92.
- Syakra, Y., Syahril, M., & Yusnidah, Y. (2019). Implementasi data mining dengan menggunakan algoritma fuzzy subtractive clustering dalam pengelompokan nilai untuk menentukan minat belajar siswa SMP primbana Medan. *Jurnal SAINTIKOM (Jurnal Sains Manajemen Informatika Dan Komputer)*, 17(1), 54–63.
- Yulianti, Y., Utami, D. Y., Hikmah, N., & Hasan, F. N. (2019). Penerapan Data Mining Menggunakan Algoritma K-Means Untuk Mengetahui Minat Customer di Toko Hijab. *Jurnal Pilar Nusa Mandiri*, 15(2), 241–246.