Jurnal Interaksi Komputer
Volume 1, Nomor 1, Agustus 2025, Page 11-21

Klasifikasi data dengan kategori Siswa dan siswi BARU dengan algoritma K-MEANS CLUSTERING

[1]Leni Mawarni Halawa
[1] Fakultas Ilmu Komputer, Program Studi Teknik Informatika, Universitas Budi Darma, Medan, Indonesia
Email:lenimawarni6@gmail.com*

Abstrak Siswa adalah komponen masukan dalam sistem pendidikan yang berpengaruh besar terkait dengan identitas sekolah, siswa merupakan komponen yang diutamakan dalam proses pendidikan, sehingga menjadi manusia yang berkualitas sesuai dengan tujuan pendidikan nasional (Budi Sutedjo dkk.2010). Dalam proses belajar mengajar terdapat dari kategori siswa terkhusus pada siswa baru kategori ini di berikan kepada siswa berdasarkan aktifitas atau nilai-nilai yang diperoleh pada saat proses belajar mengajar berlangsung.

Kata Kunci: Data Mining,K-Means, clasification

Abstract− Students are an input component in the education system that has a big influence on school identity, students are a prioritized component in the education process, so that they become quality human beings in accordance with national education goals (Budi Sutedjo et al. 2010). In the teaching and learning process there are categories of students, especially new students. This category is given to students based on the activities or values obtained during the teaching and learning process.

Keywords: Data Mining, K-Means, classification

1. PENDAHULUAN

Data mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan dalam database. *Data Mining* adalah proses ekstraksi informasi dari kumpulan data melalui penggunaan algoritma dan teknik yang melibatkan bidang ilmu stastistik, mesin pembelajaran, matematika, dan kecerdasan buatan (Turban, dkk.2005). Data mining digunakan untuk ekstrasi informasi penting yang tersembunyi dari dataset yang besar. Dengan adanya *data mining* maka akan didapatkan suatu permata berupa pengetahuan di dalam kumpulan data-data yang banyak jumlahnya.

Siswa adalah komponen masukan dalam sistem pendidikan yang berpengaruh besar terkait dengan identitas sekolah, siswa merupakan komponen yang diutamakan dalam proses pendidikan, sehingga menjadi manusia yang berkualitas sesuai dengan tujuan pendidikan nasional[1]. Dalam proses belajar mengajar terdapat dari kategori siswa terkhusus pada siswa baru kategori ini di berikan kepada siswa berdasarkan aktifitas atau nilai-nilai yang diperoleh pada saat proses belajar mengajar berlangsung.

Dalam penelitian sebelumnya pada jurnal yang berjudul "Pengelompokan Surat Dalam Al-Quran menggunakan *Algoritma K-Means* " [2]menunjukan bahwa penerapan data *mining* ada pada bidang pengelompokan, pengelompokan tersebut juga bisa di terapkan pada pembagian kategori siswa baru masuk, pengkategorian dilakukan untuk mengetahui golongan-golongan siswa berdasarkan nilai yang di raih siswa dalam ujian, pembagian kategori siswa merupakan hal penting dan perlu dilakukan pada setiap sekolah untuk evaluasi peningkatan nilai dan mutu sekolah.

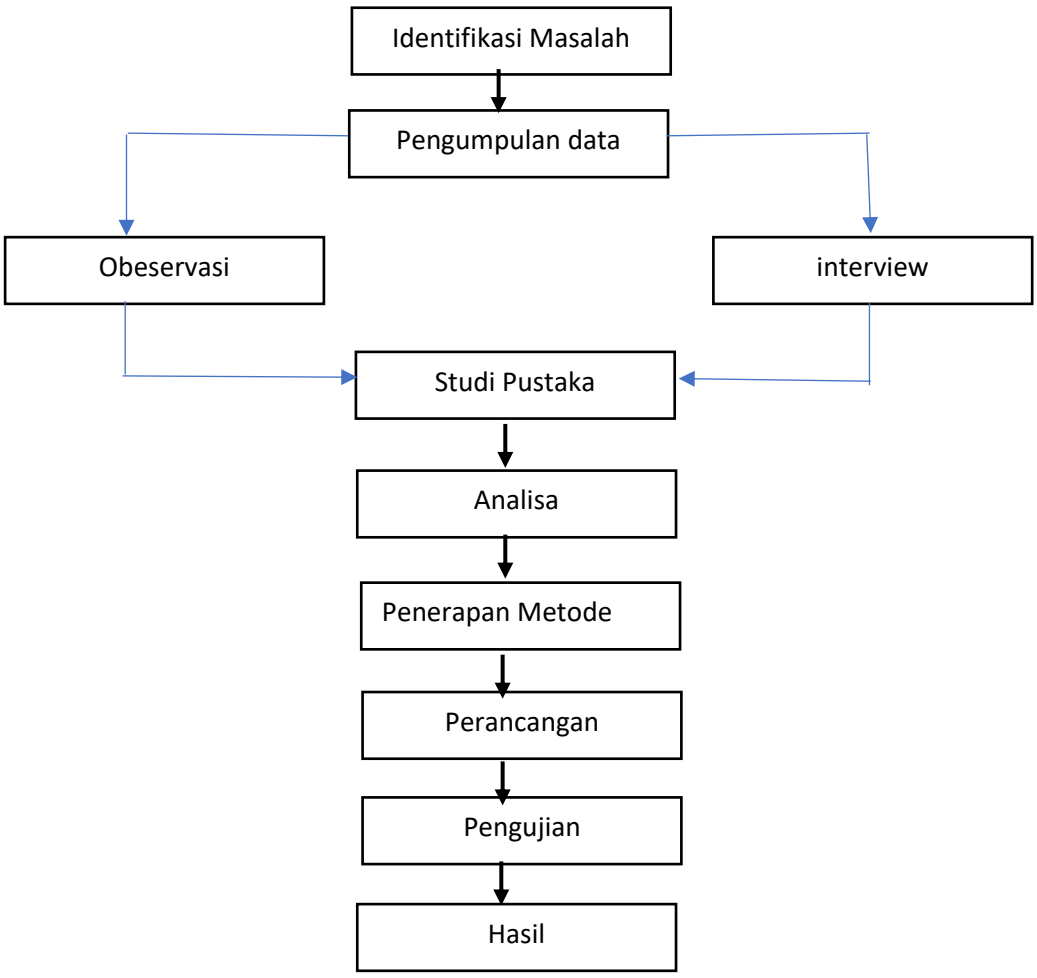
K-Means merupakan salah satu metode pengelompokan data *nonhierarki* (sekatan) berusaha mempartisi data yang ada kedalam bentuk dua atau lebih kelompok. Metode ini mempartisi data ke dalam kelompok sehingga data berkarakteristik sama di masukkan ke dalam suatu kelompok yang sama dan data yang berkarakteristik berbeda di kelompokkan kedalam kelompok yang lain [3][4].

Siswa SMA adalah merupakan yang memiliki banyak jumlah siswa baru, hal itu akan mempersulit pihak sekolah dalam menentukan mana siswa yang pintar, sedang dan biasa dalam bidang akademik masalah tersebut adalah masalah yang wajar dikarenakan banyaknya jumlah siswa dengan karakter dan kemampuan yang berbeda-beda[5]. Nilai-nilai prestasi akademik yang di raih siswa akan menjadi identitas sekolah, jika siswa-siswanya berpestasi maka pandangan masyarakat terhadap sekolah akan baik serta reputasi sekolah akan meningkat dan diminati. Dengan adanya penelitian ini diharapkan dapat mempermudah pihak sekolah dalam menentukan pembagian kategori siswa baru yang berprestasi dalam bidang akademik guna untuk mengetahui mana siswa pintar sedang dan biasa, serta di jadikan cara untuk pemilihan beasiswa pada siswa berprestasi dan juga bisa digunakan untuk penentuan pembagian jurusan.

2. METODOLOGI PENELITIAN

2.1 Tahapan Penelitian

Adapun tahapan – tahapan penelitian yang dilakukan dalam penelitian adalah sebagai berikut :



2.2 Algoritma *Clustering* Dan *K-Means*

Secara umum clustering merupakan suatu metode untuk mencari dan mengelompokkan data yang memiliki kemiripan karakteriktik *(similarity)* antara satu data dengan data yang lain[6]. *Clustering* merupakan salah satu metode *data mining* yang bersifat tanpa arahan *(unsupervised)*, maksudnya metode ini diterapkan tanpa adanya latihan *(taining)* dan tanpa ada guru *(teacher)* serta tidak memerlukan target output.

Proses pengelompokkan sekumpulan obyek ke dalam kelas – kelas obyek yang sama disebut clustering/pengelompokkan. pengklusteran adalah menyatakan kesimpulan pola ke kelompok yang sesuai yang berguna untuk menemukan kesamaan dan perbedaan sehingga dapat menghasilkan kesimpulan yang berharga.

langkah-langkah melakukan clustering dengan metode *K-Means* adalah sebagai berikut[7]:

a. Pilih jumlah *cluster k.*
b. Inisialisasi *k* pusat *cluster* ini bisa dilakukan dengan berbagai cara. Namun yang paling sering dilakukan adalah dengan cara random. Pusat-pusat cluster diberiduberi nilai awal dengan angka-angka random.
c. Alokasikan semua data/ objek ke cluster terdekat. Kedekatan dua objek ditentukan berdasarkan jarak kedua objek tersebut. Demikian juga kedekatan suatu data ke *cluster* tertentu ditentukan jarak antara data dengan pusat *cluster*. Dalam tahap ini perlu dihitung jarak tiap data ke tiap pusat *cluster*. Jarak paling antara satu data dengan satu *cluster* tertentu akan menentukan suatu data masuk dalam *cluster* mana. Untuk menghiutng jarak semua data ke setiap tiitk pusat *cluster* dapat menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut[8]:

$$d(i,j) = \left[\sum_{k=1}^{p} | x_{ik} - y_{jk} |^2 \right]^2$$

Dimana
dij = Jarak objek antara objek *i* dan *j*

P = Dimensi data
Xik = Koordinat dari objek *i* pada dimensi *k*
Xjk = Koordinat dari obyek *i*

Atau

$$D(i,j)=\sqrt{(x_{1i}-y_1j)^2+(x_{2i}-y_2j)^2+...+(x_{ki}-y_{kj})^2}$$

.

Dimana
D (*i*,*j*) = Jarak data ke *i* ke pusat cluster *j*
X*ki* = Data ke *i* pada atribut data ke *k*
X*k j* = Titik pusat ke *j* pada atribut ke *k*

d. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata – rata dari semua data/ objek dalam *cluster* tertentu. Jika dikehendaki bisa juga menggunakan median dari *cluster* tersebut. Jadi rata – rata (mean) bukan satu – satunya ukuran yang bisa di pakai.
e. Tugaskan lagi setiap objek memakai *ckuster* yang bary. Jika pusat *cluster* tidak berubah lagi maka proses *clustering* selesai. Atau, kembali ke langkah nomor 3 sampai pusat *cluster* tidak berubah lagi.

3. HASIL DAN PEMBAHASAN

Berdasarkan data diri siswa baru, namun dalam penelitian ini hanya beberapa atribut data saja yang digunakan, seperti nama siswa, nilai Ebtanas, Nilai Ujian Seleksi, dan nilai rapot. Berikut ini data siswa baru tahun 2024

3.1 Transformasi Data

Agar data di atas dapat diolah dengan menggunakan metode *k-means clustering* maka data berjenis nominal nilai EBTANAS, nilai hasil Seleksi ,
nilai rapot diinisialisasikan terlebih dahulu dalam bentuk angka.

Untuk melakukan inisialisasi alamat siswa dilakukan dengan langkah-langkah sebagai berikut :

1. Pada data nilai hasil seleksi terlebih dahulu dilakukan pembagian hasil nilai yang menjadi beberapa bagian nilai, yaitu :
   a. Nilai 81-100
   b. Nilai 60-80
   c. Nilai 0-59
2. Kemudian Nilai-nilai tersebut diurutkan dari yang terbesar berdasarkan frekuensi siswa yang di dapat pada nilai hasil seleksi.
3. Setelah itu nilai yang memiliki frekuensi terbesar diberi inisial 1 dan nilai yang memiliki frekuensi terbesar kedua diberi inisial dengan angka, begitu seterusnya hingga nilai dengan frekuensi paling sedikit. Hasil dari inisialisasi nilai seleksi siswa dapat dilihat pada tabel 1.

Tabel 1 Inisialisasi Data Nilai Hasil Seleksi

Nilai Hasil Seleksi	Frekuensi	Inisial
60-80	56	1
81-100	17	2
0-59	15	3

Selain Alamat siswa, nilai EBTANAS juga termasuk ke dalam jenis data nominal sehingga perlu diinisialisasikan. Seperti nilai seleksi siswa, pada nilai EBTANAS juga diberikan inisialisasi nilai tersebut dapat dilihat pada tabel 4.8.

Tabel 2 Inisialisasi Nilai Ebtanas

Nilai EBTANAS	Frekuensi	Inisial
87-92	28	1
75-80	19	2

81-86	14	3
93- 100	12	4

Selain nilai seleksi siswa, nilai EBTANAS, nilai rapot juga termasuk ke dalam jenis data nominal sehingga perlu diinisialisasikan ke dalam bentuk angka. Seperti pada nilai seleksi siswa, nilai EBTANAS, nilai rapot juga diberikan inisialisasi berdasarkan frekuensi siswa pada nilai rapot tersebut. Hasil dari inisialisasi nilai hasil rapot tersebut dapat dilihat pada tabel 3

Tabel 3 Inisialisasi Nilai Rapot Siswa

Nilai Rapot	Frekuensi	Inisial
B	36	1
C	21	2
A	14	3
D	2	4

3.2 Penerapan Algoritma K-Means Clustering

Setelah semua data siswa tahun 2014 di transformasikan ke dalam bentuk angka, maka data-data tersebut bisa dikelompokkan ke dalam algoritma *k-means clustering*. Untuk dapat melakukan pengelompokkan data-data tersebut menjadi beberapa *cluster* perlu dilakukan beberapa langkah, yaitu :

1. Tentukan jumlah *cluster* yang diinginkan. Dalam penelitian ini data-data yang akan dikelompokan menjadi tiga *cluster*.
2. Tentukan titik pusat awal dari setiap cluster. Dalam penelitian ini titik pusat awal ditentukan secara random dan didapat titik pusat dari setiap *cluster* dapat dilihat pada tabel

Tabel 4.10 Titik Pusat Awal Setiap *Cluster*

Titik Pusat Awal	Nama Siswa	NEM	Nilai Seleksi	Nilai Rapot
Cluster 1	Iis Irawan	1	1	1
Cluster 2	Padli Irwansyah	3	1	1
Cluster 3	Gustiya	4	2	4

Cluster 1 : di ambil dari data ke-27
Cluster 2 : di ambil dari data ke-54
Cluster 3 : di ambil dari data ke-24

3. Tempatkan setiap data pada *cluster*. Dalam penilitian ini digunakan metode *hard k-means* untuk mengalokasikan setiap data ke dalam suatu *cluster* yang memiliki jarak paling dekat dengan titik pusat dari setiap *cluster*. Untuk mengetahui *cluster* mana yang paling dekat dengan data, maka perlu dihitung jarak setiap data dengan titik pusat setiap *cluster*. Sebagai contoh, akan dihitung jarak dari data siswa pertama ke pusat *cluster* pertama, menggunakan teori jarak *Euclidean* yang dirumuskan sebagai berikut:

$$d(i,j) = \left[\sum_{k=1}^{p} | x_{ik} - y_{jk} |^2 \right]^2$$

Dimana
dij = Jarak objek antara objek *i* dan *j*
P = Dimensi data
Xik = Koordinat dari objek *i* pada dimensi *k*
Xjk = Koordinat dari obyek *i*

Atau

$$D(i,j) = \sqrt{(x_{1i} - y_1 j)^2 + (x_{2i} - y_2 j)^2 + ... + (x_{ki} - y_{kj})^2}$$

Dimana
D (*i*,*j*) = Jarak data ke *i* ke pusat cluster *j*
X*ki* = Data ke *i* pada atribut data ke *k*
X*k j* = Titik pusat ke *j* pada atribut ke *k*

$$d(1,1) = \sqrt{(1-1)^2 + (1-1)^2 + (2-1)^2}$$

$$= \sqrt{(0) + (0) + (1)}$$

$$= \sqrt{1}$$
$$d(1,1) = 1$$

Dari hasil perhitungan di atas didapatkan hasil bahwa jarak data siswa pertama dengan *cluster* pertama adalah 1

Jarak dari data siswa pertama ke pusat *cluster* kedua :

$$d(1,2) = \sqrt{(1-3)^2 + (1-1)^2 + (2-1)^2}$$
$$= \sqrt{(4) + (0) + (1)}$$
$$= \sqrt{5}$$
$$d(1,2) = 2{,}236$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data siswa pertama dengan pusat *cluster* kedua adalah 2,236

Jarak data siswa pertama ke pusat *cluster* ketiga :

$$d(1,3) = \sqrt{(1-4)^2 + (1-2)^2 + (2-4)^2}$$
$$= \sqrt{(9) + (1) + (4)}$$
$$= \sqrt{14}$$
$$d(1,3) = 3{,}741$$

Dari hasil perhitungan di atas di dapatkan hasil bahwa jarak data siswa pertama dengan pusat *cluster* ketiga adalah 3,741

Berdasarkan hasil ketiga perhitungan di atas didapatkan hasil bahwa jarak data siswa pertama paling dekat adalah dengan *cluster* 1, sehingga data siswa pertama di masukkan ke dalam *cluster* 1 (C1).

a. Menghitung jarak data ke-1 dengan *centroid* pertama (C1)

$$d(1,1) = \sqrt{(1-1)^2 + (1-1)^2 + (2-1)^2}$$
$$= \sqrt{(0) + (0) + (1)}$$
$$= \sqrt{1}$$
$$d(1,1) = 1$$

$$d(2,1) = \sqrt{(1-1)^2 + (1-1)^2 + (1-1)^2} = 0$$
$$d(3,1) = \sqrt{(2-1)^2 + (1-1)^2 + (3-1)^2} = 2.236$$
$$d(4,1) = \sqrt{(1-1)^2 + (1-1)^2 + (2-1)^2} = 1$$
$$d(5,1) = \sqrt{(2-1)^2 + (1-1)^2 + (2-1)^2} = 1{,}414$$
$$d(6,1) = \sqrt{(2-1)^2 + (1-1)^2 + (1-1)^2} = 1$$
$$d(7,1) = \sqrt{(4-1)^2 + (2-1)^2 + (1-1)^2} = 3{,}162$$
$$d(8,1) = \sqrt{(3-1)^2 + (2-1)^2 + (1-1)^2} = 2{,}236$$
$$d(9,1) = \sqrt{(1-1)^2 + (1-1)^2 + (1-1)^2} = 0$$
$$d(10,1) = \sqrt{(4-1)^2 + (1-1)^2 + (1-1)^2} = 3$$

Hasil perhitungan data ke-1 dengan *centroid* pertama (C1) selengkapnya untuk data siswa tahun 2014 bisa dilihat pada tabel 4.11.

b. Menghitung jarak data ke-1 dengan *centroid* kedua (C2)

$$d(1,2) = \sqrt{(1-3)^2 + (1-1)^2 + (2-1)^2}$$
$$= \sqrt{(4) + (0) + (1)}$$
$$= \sqrt{5}$$
$$d(1,2) = 2{,}236$$

$$d(2,2) = \sqrt{(1-3)^2 + (1-1)^2 + (1-1)^2} = 2$$
$$d(3,2) = \sqrt{(2-3)^2 + (1-1)^2 + (3-1)^2} = 2{,}236$$
$$d(4,2) = \sqrt{(1-3)^2 + (1-1)^2 + (2-1)^2} = 2{,}236$$
$$d(5,2) = \sqrt{(2-3)^2 + (1-1)^2 + (2-1)^2} = 1{,}414$$
$$d(6,2) = \sqrt{(2-3)^2 + (1-1)^2 + (1-1)^2} = 1$$
$$d(7,2) = \sqrt{(4-3)^2 + (2-1)^2 + (1-1)^2} = 1{,}414$$
$$d(8,2) = \sqrt{(3-3)^2 + (2-1)^2 + (1-1)^2} = 1$$
$$d(9,2) = \sqrt{(1-3)^2 + (1-1)^2 + (1-1)^2} = 2$$
$$d(10,2) = \sqrt{(4-3)^2 + (1-1)^2 + (1-1)^2} = 1$$

Hasil perhitungan data ke-1 dengan *centroid* kedua (C2) selengkapnya untuk data siswa tahun 2014 bisa dilihat pada tabel 4.11.

c. Menghitung jarak data ke-1 dengan *centroid* ketiga (C3)

$$d(1,3) = \sqrt{(1-4)^2 + (1-2)^2 + (2-4)^2}$$
$$= \sqrt{(9) + (1) + (4)}$$
$$= \sqrt{14}$$
$$d(1,3) = 3{,}741$$
$$d(2,3) = \sqrt{(1-4)^2 + (1-2)^2 + (1-4)^2} = 4{,}358$$
$$d(3,3) = \sqrt{(2-4)^2 + (1-2)^2 + (3-4)^2} = 2{,}449$$
$$d(4,3) = \sqrt{(1-4)^2 + (1-2)^2 + (2-4)^2} = 3{,}741$$
$$d(5,3) = \sqrt{(2-4)^2 + (1-2)^2 + (2-4)^2} = 3$$
$$d(6,3) = \sqrt{(2-4)^2 + (1-2)^2 + (1-4)^2} = 3{,}741$$
$$d(7,3) = \sqrt{(4-4)^2 + (2-2)^2 + (1-4)^2} = 3$$
$$d(8,3) = \sqrt{(3-4)^2 + (2-2)^2 + (1-4)^2} = 3{,}162$$
$$d(9,3) = \sqrt{(1-4)^2 + (1-2)^2 + (1-4)^2} = 4{,}358$$
$$d(10,3) = \sqrt{(4-4)^2 + (1-2)^2 + (1-4)^2} = 3{,}162$$

Hasil perhitungan data ke-1 dengan *centroid* ketiga (C3) selengkapnya untuk data siswa tahun 2014 bisa dilihat pada tabel 4.11.

Tabel 4.11 : Hasil Perhitungan Jarak Setiap Data Pada Iterasi 1

No	Nama Siswa	Nilai EBTANAS	Nilai Seleksi	Nilai Rapot	Jarak Ke			Jarak Terdekat Ke Cluster
					C1	C2	C3	
1	Adnan Rizani	1	1	2	**1**	2,236	3,741	C1
2	Agustina Daulay	1	1	1	**0**	2	4,358	C1
3	Ahmad Irfan Tanjung	2	1	3	**2,236**	**2,236**	2,449	C1 dan C2
4	Ananda Subakti Putra	1	1	2	**1**	2,236	3,741	C1
5	Andri Saputra	2	1	2	**1,414**	**1,414**	3	C1 dan C2

Suatu data akan menjadi anggota dari suatu *cluster* yang memiliki jarak terkecil dari pusat *cluster* nya. Misalkan untuk data pertama, jarak terkecil diperoleh pada *cluster* 1, sehingga data pertama akan menjadi anggota dari *cluster* 1. Demikian juga untuk data kedua, jarak terkecil ada pada *cluster* 3, maka data tersebut akan masuk pada *cluster* 3.

Matriks jarak dengan *centroid* awal seperti tersebut di atas adalah posisi *cluster* selengkapnya dapat dilihat pada tabel 12.

Tabel 12 : Posisi *Cluster* Dengan *Centroid* Awal Pada Iterasi 1

Cluster 1	*Cluster* 2	*Cluster* 3
1	0	0
1	0	0
1	1	0
1	0	0
1	1	0
1	1	0
0	1	0

0	1	0
1	0	0
0	1	0
1	0	0
0	1	0
1	0	0
1	0	0
0	1	0
1	0	0
1	0	0
1	1	0
0	1	0
1	0	0
1	0	0
0	0	1
1	0	0
0	1	0
1	0	0
1	1	0
1	0	0
0	1	0
1	0	0
1	1	0
1	0	0

Lanjutan Tabel 4.12 Posisi *Cluster* Dengan *Centroid* Awal Pada Iterasi 1

Cluster 1	*Cluster* 2	*Cluster* 3
1	0	0
1	1	0
1	0	0
0	1	0
1	0	0
1	1	0
0	0	1
0	1	0
1	1	0

1	1	0
0	1	0
1	0	0
1	1	0
0	1	0
1	1	0
0	0	1
0	1	0
1	1	0
0	1	0
0	1	0
0	1	0
0	1	0
0	0	1
0	0	1
1	0	0
0	1	0
0	1	0
1	0	0
0	1	0
1	0	0
0	1	0
1	0	0
1	0	0
1	1	0
0	1	0
1	1	0

Catatan : tanda angka (1) menyatakan keanggotaan data terhadap suatu kelompok.

4. Setelah semua data ditempatkan ke dalam *cluster* yang terdekat. Hitung kembali pusat *cluster* dengan keanggotaan *cluster* yang sekarang. Pusat *cluster* adalah rata-rata dari semua data/obyek dalam *cluster* tertentu. bisa juga memakai median dari *cluster* tersebut. Jadi rata-rata (*mean*) bukan satu-satunya ukuran yang dapat digunakan. Hitung pusat *cluster* baru.
   Untuk *cluster* 1, ada 46 data yaitu data ke-1, 2, 3, 4, 5, 6, 9, 11, 13, 14, 16, 17, 18, 20, 21, 23, 25, 26, 27, 29, 30, 31, 32, 33, 34, 36, 37, 39, 40, 41, 42, 43, 44, 46, 47, 49, 50, 52, 55, 62, 65, 67, 69, 70, 71, dan data ke-73 sehingga :

$$K_{11} = \frac{1+1+2+1+2+2+1+1+1+1+1+1+2+1+1+1+1+2+1+1+2+1+1+2+1+1+}{46}$$

$$= \frac{2+2+2+1+2+2+1+2+2+1+2+2+2+1+1+1+1+1+2+2}{46} = \frac{65}{46} = 1{,}413$$

$$K_{12} = \frac{1+1+1+1+1+1+1+1+1+1+1+2+1+1+1+1+2+2+2+1+1+1+1+2+1+1+}{46}$$

$$= \frac{1+1+1+1+1+1+1+1+1+1+1+1+1+2+2+1+1+2+2+1}{46} = \frac{55}{46} = 1{,}195$$

$$K_{13} = \frac{2+1+3+2+2+1+1+3+1+2+2+1+2+3+1+1+1+1+1+1+2+2+4+1+3+1+}{46}$$

$$= \frac{1+2+1+1+1+1+2+3+2+1+3+2+1+3+1+1+1+3+1+2}{46} = \frac{78}{46} = 1{,}695$$

Jadi hasil *centroid* baru *cluster* 1 adalah (1,413, 1,195, 1,695)

Untuk *cluster* 2, ada 41 data yaitu data ke-3, 5, 6, 7, 8, 10, 12, 15, 18, 19, 24, 26, 28, 30, 33, 35, 37, 39, 40, 42, 43, 45, 46, 47, 48, 50, 51, 52, 54, 55, 56, 57, 58, 59, 63, 64, 66, 68, 71, 72 dan data ke-73 sehingga :

$$K_{21} = \frac{2+2+2+4+3+4+3+1+2+3+4+2+3+2+2+3+2+2+2+2+2+4+2+2+3+2+}{41}$$

$$= \frac{3+2+3+2+4+3+3+4+3+4+3+3+2+3+2}{41} = \frac{109}{41} = 2{,}658$$

$$K_{22} = \frac{1+1+1+2+2+1+1+2+1+1+2+2+1+1+2+1+1+1+1+1+1+1+1+1+1+1+}{41}$$

$$= \frac{1+1+1+1+1+1+1+1+1+1+2+2+2+1+1}{41} = \frac{50}{41} = 1{,}219$$

$$K_{23} = \frac{3+2+1+1+1+1+2+3+2+2+4+1+2+2+1+2+1+2+1+1+1+1+3+2+1+3+}{41}$$

$$= 1\frac{+2+1+1+1+2+1+1+1+2+1+1+1+1+2}{41} = \frac{65}{41} = 1{,}585$$

Jadi hasil *centroid* baru *cluster* 2 adalah (2,658, 1,219, 1,585)

Untuk *cluster* 3, ada 5 data yaitu data ke-22, 38, 53, 60, dan data ke-61 sehingga :

$$K_{31} = \frac{4+4+4+4+4}{5} = \frac{20}{5} = 4$$

$$K_{32} = \frac{2+2+2+1+2}{5} = \frac{9}{5} = 1{,}8$$

$$K_{33} = \frac{3+3+3+3+3}{5} = \frac{15}{5} = 3$$

Jadi hasil *centroid* baru *cluster* 3 adalah (4, 1.8, 3)

5. Ulangi langkah 3 hingga posisi data sudah tidak mengalami perubahan

Lanjutan Tabel 15 Posisi *Cluster* Dengan *Centroid* Awal Pada Iterasi 3

Cluster 1	*Cluster* 2	*Cluster* 3
0	0	1
0	0	1

0	0	0
0	1	0
0	0	1
1	0	0
0	1	0
1	0	0
0	1	0
1	0	0
1	0	0
0	1	0
0	1	0
1	0	0

Catatan : tanda angka (1) menyatakan keanggotaan data terhadap suatu kelompok.

6. Karena pada iterasi ke-3 posisi *cluster* tidak berubah, maka iterasi dihentikan dan hasil akhir yang diperoleh adalah 3 *cluster.*

Dalam penelitian ini, iterasi *clustering* data siswa terjadi sebanyak 3 kali iterasi. Pada iterasi ke-3 ini, titik pusat dari setiap *cluster* yang tidak berubah dan
tidak ada data yang berpindah antar *cluster.*

Dari hasil *cluster* 1 memiliki pusat (1,413, 1,195, 1,695) adapun yang masuk pada *cluster* 1 yaitu data ke- (1, 2, 3, 4, 5, 9, 13, 14, 15, 16, 17, 18, 21, 23, 25, 27, 29, 30, 31, 32, 36, 39, 41, 44, 46, 47, 49, 50, 52, 62, 65, 67, 69, 70, 73). Terlihat bahwa karakteristik siswa pada *cluster* 1 didominasi oleh nilai EBTANAS berkisar 87-92, berdasarkan nilai seleksi siswa didominasi oleh siswa yang memperoleh nilai berkisar dari 60-80, berdasarkan nilai rapot siswa di dominasi oleh siswa yang memiliki nilai B.

Dari hasil *cluster* 2 memiliki pusat (2,658, 1,219, 1,585) adapun yang masuk pada *cluster* 2 yaitu data ke- (6, 7, 8, 10, 12, 19, 26, 28, 33, 35, 37, 40, 42, 43, 45, 48, 51, 54, 55, 56, 57, 58, 59, 63, 66, 68, 71, 72). Terlihat bahwa karakteristik pada *cluster* 2 didominasi oleh nilai EBTANAS berkisar 81-86, berdasarkan nilai seleksi siswa didominasi dengan nilai berkisar dari 60- 80, berdasarkan nilai rapot siswa di dominasi oleh nilai B.
Dari hasil *cluster* 3 memiliki pusat (4, 1.8, 3) adapun yang masuk pada *cluster* 3 yaitu data ke- (11, 20, 22, 24, 34, 38, 53, 60, 61, 64). Terlihat bahwa karakteristik siswa pada *cluster* 3 didominasi oleh nilai EBTANAS berkisar 93-100, berdasarkan nilai seleksi siswa di dominasi nilai 80-100, berdasarkan nilai rapot siswa di dominasi nilai A.

4. KESIMPULAN

Pembagian kategori siwa baru di SMA Negeri 1 Kualuh Hilir belum efektif dan efisien, dikarenakan belum adanya metode khusus untuk mendukung keberhasilan dari proses pembagian kategori, dan juga pengkategorian siswa masih dilakukan dengan manual. Melakukan pembagian kategori berdasarkan data-data nilai ebtanas, nilai seleksi dan nilai rapot yang paling banyak didominasi siswa kemudian akan dikategorikan menjadi siswa pintar, sedang dan biasa, dengan melakukan penyelarasan menggunakan *promotion mix* dengan melihat nilai –nilai yang didapat pada setiap *cluster.*Algoritma *K-Means Clustering* pada *rapidminer* dimulai dengan penginputan data siswa menjadi database pada *Ms.Excel* dengan ekstensi csv. Mengelompokkan data dengan algoritma *k-measn clustering,* dilakukan dengan cara menentukan jumlah *cluster,* hitung jarak terdekat dengan pusat *cluster.* Data dengan jarak terdekat menyatakan anggota *cluster* tersebut, dilakukan perhitungan kembali sampai data tidak berpindah pada *cluster* lain, untuk meminimalkan fungsi objektif.

REFERENCES

[1] E. Ndruru dan E. N. Purba, “Penerapan Metode ARAS Dalam Pemilihan Lokasi Objek Wisata Yang Terbaik Pada Kabupaten Nias Selatan,” *METHOMIKA J. Manaj. Inform. Komputerisasi Akunt.*, vol. 3, no. 2, hal. 151–159, 2019.

[2] S. Saefudin dan S. DN, “Penerapan Data Mining Dengan Metode Algoritma Apriori Untuk Menentukan Pola Pembelian Ikan,” *JSiI (Jurnal Sist. Informasi)*, vol. 6, no. 2, hal. 36, 2019, doi: 10.30656/jsii.v6i2.1587.

[3] S. F. Rodiyansyah, "Algoritma Apriori untuk Analisis Keranjang Belanja pada Data Transaksi Penjualan," *Infotech*, vol. 1, no. 1, hal. 36–39, 2015.

[4] K. Tampubolon, H. Saragih, B. Reza, K. Epicentrum, A. Asosiasi, dan A. Apriori, "Implementasi Data Mining Algoritma Apriori Pada Sistem Persediaan Alat-Alat Kesehatan," hal. 93–106, 2013.

[5] D. H. Kamagi dan S. Hansun, "Implementasi Data Mining dengan Algoritma C4.5 untuk Memprediksi Tingkat Kelulusan Mahasiswa," *J. Ultim.*, vol. 6, no. 1, hal. 15–20, 2014, doi: 10.31937/ti.v6i1.327.

[6] A. Pratama, R. C. Wihandika, dan D. E. Ratnawati, "Implementasi algoritme support vector machine (SVM) untuk prediksi ketepatan waktu kelulusan mahasiswa," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 2, no. April, hal. 1704–1708, 2018.

[7] S. Syarli dan A. Muin, "Metode Naive Bayes Untuk Prediksi Kelulusan (Studi Kasus: Data Mahasiswa Baru Perguruan Tinggi)," *J. Ilm. Ilmu Komput.*, vol. 2, no. 1, hal. 22–26, 2016.

[8] A. E. Wicaksono, "… Dalam Pengelompokan Data Peserta Didik Di Sekolah Untuk Memprediksi Calon Penerima Beasiswa Dengan Menggunakan Algoritma K-Means (Studi Kasus Sman 16 …," *J. Ilm. Teknol. dan Rekayasa*, vol. 21, no. 3, 2017.