

PENERAPAN METODE KLASTERISASI *K-MEANS* UNTUK STRATEGI PROMOSI PADA SMK TAMANSISWA SUKADAMAI

Teti Purwanti¹, William Ramdhan^{2*}, Santoso³

^{1,2}Sistem Informasi, STMIK Royal Kisaran

³Manajemen Informatika, STMIK Royal Kisaran

**email*: william_ramdhan052@gmail.com

Abstract: SMK Tamansiswa Sukadamai is one of the private schools in Pulo Bandring District which annually admits new students, this will affect the amount of student data that goes up and down. So far, SMK Tamansiswa Sukadamai has implemented a promotional strategy but it is still not optimal because the number of students registered at SMK Tamansiswa Sukadamai is more dominant in certain areas. Based on this, we need a technique that is able to assist in transforming the data into useful information, namely by applying data mining for promotional strategies, which can be the basis or guideline for promotional strategies based on areas where there are not many students present. The application of the k-means clustering method for promotional strategies can assist in the data grouping process in the form of data grouping results for C1 "Potential" and C2 "Potential", where C1 is the right area for more optimal promotion so that promotion is more effective and efficient. With the development of this clustering system, it can provide input to schools to determine which areas are more optimal for more in-depth promotion.

Keywords: Promotion Strategy; Application of the K-means clustering method

Abstrak: SMK Tamansiswa Sukadamai merupakan salah satu sekolah swasta yang ada di Kecamatan Pulo Bandring yang setiap tahunnya melakukan penerimaan siswa baru, hal ini akan berpengaruh pada jumlah data siswa yang pasang surut. Selama ini SMK Tamansiswa Sukadamai telah melakukan strategi promosi akan tetapi masih belum optimal dikarenakan jumlah siswa yang terdaftar di SMK Tamansiswa Sukadamai lebih dominan berasal di wilayah tertentu saja. Berdasarkan hal tersebut diperlukan suatu teknik yang mampu membantu dalam mentransformasikan data tersebut menjadi informasi yang berguna yaitu dengan menerapkan data mining untuk strategi promosi, yang dapat menjadi dasar atau pedoman untuk strategi promosi berdasarkan wilayah yang tidak banyak dalam menghadirkan siswanya. Penerapan metode klasterisasi k-means untuk strategi promosi dapat membantu dalam proses pengelompokan data dalam bentuk hasil pengelompokan data C1 "Berpotensi" dan C2 "Tidak Berpotensi", dimana C1 merupakan wilayah yang tepat untuk dilakukan promosi yang lebih optimal sehingga promosi lebih efektif dan efisien. Dengan dikembangkannya sistem klasterisasi ini, maka dapat memberikan masukan kepada pihak sekolah untuk menentukan wilayah yang lebih optimal untuk promosi lebih mendalam.

Kata kunci: Strategi Promosi; Penerapan metode Klasterisasi K-Means

PENDAHULUAN

Penerimaan siswa baru merupakan kegiatan yang dilakukan setiap tahunnya oleh setiap sekolah, termasuk SMK Tamansiswa Sukadamai yang ada di Kecamatan Pulo Bandring. Jumlah calon siswa baru setiap tahunnya yang mendaftar ke SMK Tamansiswa Sukadamai cenderung mengalami penurunan. Oleh sebab itu diperlukan suatu teknik yang berguna dalam membantu mentransformasikan data tersebut menjadi informasi yang berguna yaitu dengan menerapkan *data mining* untuk strategi promosi, yang dapat menjadi dasar atau pedoman untuk strategi promosi berdasarkan wilayah yang tidak banyak dalam menghadirkan siswanya.

Penerapan *data mining* pada penerimaan siswa baru di SMK Tamansiswa Sukadamai dapat digunakan dalam menentukan strategi promosi, dalam menentukan daerah atau area potensial untuk melakukan promosi. Hal ini dapat membantu tim Penerima Siswa Baru (PSB) untuk menentukan wilayah promosi serta membantu pihak sekolah untuk mengetahui jumlah siswa berdasarkan wilayahnya.

Selama ini SMK Tamansiswa Sukadamai telah melakukan strategi promosi akan tetapi masih belum optimal dikarenakan jumlah siswa yang terdaftar di SMK Tamansiswa Sukadamai lebih dominan berasal di wilayah tertentu saja. Sehingga dirasa perlu dilakukannya klusterisasi data siswa untuk mengetahui wilayah ataupun desa yang menghasilkan siswa terendah agar dapat dilakukan promosi yang lebih maksimal pada wilayah tersebut sehingga dapat menghemat biaya yang nantinya bisa diarahkan ke wilayah yang belum optimal untuk dilakukannya promosi.

Berdasarkan hal tersebut peneliti mengimplementasi Teknik *K-Means* guna strategi promosi pada SMK Taman Siswa Sukadamai. Sistem ini nantinya dapat menentukan wilayah yang belum optimal dalam melakukan promosi.

METODE

Data mining suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam *database*. *Data mining* juga merupakan teknik yang menggabungkan teknik analisis data dan menemukan pola-pola pada data [1]. *Clustering* adalah proses mengelompokkan objek berdasarkan informasi yang diperoleh dari data yang menjelaskan hubungan antar objek dengan prinsip untuk memaksimalkan kesamaan antar anggota satu kelas dan meminimalkan kesamaan antar kelas/*cluster* [2].

Metode *K-Means* merupakan algoritma klusterisasi yang paling tua dan paling banyak digunakan dalam berbagai aplikasi kecil hingga menengah karena kemudahan implementasinya [3]. Langkah-langkah dalam *Algoritma K-Means* adalah sebagai berikut: (1) Siapkan dataset., (2) Tentukan jumlah *cluster* (K)., (3) Pilih titik *centroid* secara acak., (4) Kelompokkan data sehingga terbentuk K buah *cluster* dengan titik *centroid* dari setiap *cluster*., (5) Perbarui nilai titik *centroid*., (6) Ulangi langkah 3 sampai 5 sampai nilai dari titik *centroid* tidak lagi berubah [4]

Dalam menentuka nilai *centroid* untuk awal iterasi, nilai awal *centroid* dilakukan secara acak. Sedangkan jika iterasi, maka digunakan rumus sebagai berikut:

$$d(x,y)=\sqrt{\sum_{i=1}^{nn} x_i (x_y - y_i)^2} \quad (1)$$

di mana :

d(x,y) = jarak antara data x ke data y
 xi = data testing ke-i
 yi = data training ke-i [5]

$$\mu_k = \frac{1}{N_k} \sum_{i=1}^{N_k} x_i \quad (2)$$

di mana :

μk = jarak centroid dari cluster ke-K
 Nk = banyaknya data pada cluster ke-K
 xi = data ke-i pada cluster ke-K.

HASIL DAN PEMBAHASAN

Menentukan Jumlah Cluster

Jumlah *cluster* merupakan jumlah kelompok yang akan dihasilkan. Dalam penelitian kali ini *cluster* yang akan dibuat adalah 2 *cluster* dimana *cluster 1* untuk kelompok Berpotensi dan *cluster 2* untuk kelompok Tidak Berpotensi. Untuk penentuan awal: (1) Diambil dari data ke-1 sebagai pusat *cluster* ke-1: (0,3,1,2,1), (2) Diambil dari data ke-2 sebagai pusat *cluster* ke-2: (5,6,2,3,5).

Perhitungan Jarak Pusat Cluster

Dengan menggunakan rumus:

$$D_x = \sqrt{(X_i - S_i)^2 + (y_i - t_i)^2} \quad (3)$$

Perhitungan jarak data ke-1 terhadap pusat *cluster*

$$\begin{aligned} C1 &= \sqrt{(0 - 0)^2 + (3 - 3)^2 + (1 - 1)^2 + (2 - 2)^2 + (1 - 1)^2} \\ &= \sqrt{(0)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2} \\ &= \sqrt{0} = 0,00 \end{aligned}$$

$$\begin{aligned} C2 &= \sqrt{(0 - 5)^2 + (3 - 6)^2 + (1 - 2)^2 + (2 - 3)^2 + (1 - 5)^2} \\ &= \sqrt{(-5)^2 + (-3)^2 + (-1)^2 + (-1)^2 + (-4)^2} \\ &= \sqrt{52} = 7,21 \end{aligned}$$

Perhitungan jarak data ke-2 terhadap pusat *cluster*

$$\begin{aligned} C1 &= \sqrt{(5 - 0)^2 + (6 - 3)^2 + (2 - 1)^2 + (3 - 2)^2 + (5 - 1)^2} \\ &= \sqrt{(5)^2 + (3)^2 + (1)^2 + (-1)^2 + (4)^2} \\ &= \sqrt{52} = 7,21 \end{aligned}$$

$$\begin{aligned} C2 &= \sqrt{(5 - 5)^2 + (6 - 6)^2 + (2 - 2)^2 + (3 - 3)^2 + (5 - 5)^2} \\ &= \sqrt{(0)^2 + (0)^2 + (0)^2 + (0)^2 + (0)^2} \\ &= \sqrt{0} = 0,00 \end{aligned}$$

Perhitungan jarak data ke-3 terhadap pusat *cluster*

$$\begin{aligned} C1 &= \sqrt{(10 - 0)^2 + (7 - 3)^2 + (3 - 1)^2 + (13 - 2)^2 + (11 - 1)^2} \\ &= \sqrt{(10)^2 + (4)^2 + (2)^2 + (11)^2 + (10)^2} \end{aligned}$$

$$= \sqrt{341} = 18,46$$

$$C2 = \frac{\sqrt{(10 - 5)^2 + (7 - 6)^2 + (3 - 2)^2 + (13 - 3)^2 + (11 - 5)^2}}{\sqrt{(5)^2 + (1)^2 + (1)^2 + (10)^2 + (6)^2}}$$

$$= \sqrt{163} = 12,76$$

Perhitungan jarak data ke-4 terhadap pusat *cluster*

$$C1 = \frac{\sqrt{(1 - 0)^2 + (1 - 3)^2 + (0 - 1)^2 + (1 - 2)^2 + (0 - 1)^2}}{\sqrt{(1)^2 + (-2)^2 + (-1)^2 + (-1)^2 + (-1)^2}}$$

$$= \sqrt{8} = 2,83$$

$$C2 = \frac{\sqrt{(1 - 5)^2 + (1 - 6)^2 + (0 - 2)^2 + (1 - 3)^2 + (0 - 5)^2}}{\sqrt{(-4)^2 + (-5)^2 + (-2)^2 + (-2)^2 + (-5)^2}}$$

$$= \sqrt{74} = 8,60$$

Perhitungan jarak data ke-5 terhadap pusat *cluster*

$$C1 = \frac{\sqrt{(2 - 0)^2 + (1 - 3)^2 + (0 - 1)^2 + (3 - 2)^2 + (2 - 1)^2}}{\sqrt{(2)^2 + (-2)^2 + (-1)^2 + (1)^2 + (1)^2}}$$

$$= \sqrt{11} = 3,32$$

$$C2 = \frac{\sqrt{(2 - 5)^2 + (1 - 6)^2 + (0 - 2)^2 + (3 - 3)^2 + (2 - 5)^2}}{\sqrt{(-3)^2 + (-5)^2 + (-2)^2 + (0)^2 + (-3)^2}}$$

$$= \sqrt{47} = 6,86$$

Perhitungan jarak dari data ke 1 ke data 18 terhadap pusat *cluster*. Sehingga hasil perhitungan jarak selengkapya adalah:

Tabel 1. Hasil Iterasi Pertama

No.	Wilayah	CENTR 1					CENTR 2				
		0	3	1	2	1	5	6	2	3	5
1.	Aek Gerger	0,00					7,21				
2.	Pagar Bosi	7,21					0,00				
3.	Perhutaan Silau	18,47					12,77				
4.	Riah Naposo	2,83					8,60				
5.	Sayur Matinggi	3,32					6,86				
6.	Sei Balai	4,00					9,49				
7.	Sei Torop	3,32					8,89				
8.	Silau Maraja	2,45					7,87				
9.	Sordang Baru	14,83					8,94				
10.	Sordang Bolon	3,61					9,22				
11.	Sukadamai	5,48					5,10				
12.	Sukadamai Barat	6,71					5,57				
13.	Talun Tanjung	2,24					8,06				
14.	Taman Sari	4,80					5,00				
15.	Tanah Rakyat	3,46					9,70				
16.	Tanjung Rapuan	3,32					9,54				
17.	Ujung Padang	2,45					8,60				
18.	Urung Pane	3,46					7,87				

Struktur Pengelompokan Data

Tabel 2. Pengelompokan Data

Data Ke	C1	C2
1	C1	
2		C2
3		C2
4	C1	
5	C1	
6	C1	
7	C1	
8	C1	
9		C2
10	C1	
11		C2
12		C2
13	C1	
14	C1	
15	C1	
16	C1	
17	C1	
18	C1	

Karena struktur pengelompokan data menghasilkan nilai yang sama dari hasil pengelompokan sebelumnya, maka tidak perlu dilakukan iterasi lagi. Sampai disini hasil klasterisasi sudah mencapai stabil dan konvergen.

SIMPULAN

Berdasarkan hasil analisa dan pembahasan *data mining* menggunakan metode yang dilakukan maka dapat ditarik kesimpulan bahwa dengan jumlah data sebanyak 18 wilayah yang terdiri 5 atribut yaitu tahun 2016, 2017, 2018, 2019 dan 2020 telah menghasilkan 2 *cluster* dimana *cluster* 1 disebut dengan Berpotensi, dan *cluster* 2 disebut dengan Tidak Berpotensi.

Maka kesimpulan dari iterasi pertama sampai dengan ketiga dapat disimpulkan bahwa dimana hasil dari *clustering* kedua dan *clustering* ketiga terdapat data yang sama sehingga tidak perlu dilakukan iterasi/perulangan lagi. Hasil *Clustering* adalah: (1) *Cluster* 1: Wilayah 1,4,5,6,7,8,10,11,13,14,15,16,17,18., (2) *Cluster* 2: Wilayah 2,3,9,12. Dimana *Cluster* pertama terdiri dari Aek Gerger, Riah Naposo, Sayur Matinggi, Sei Balai, Sei Torop, Silau Maraja, Sordang Bolon, Sukadamai, Talun Tanjung, Taman Sari, Tanah Rakyat, Tanjung Rapuan, Ujung Padang, Urung Pane. Untuk *Cluster* kedua yaitu terdiri dari Pagar Bosi, Perhutaan Silau,Sordang Baru, Sukadamai Barat.

DAFTAR PUSTAKA

- [1] N. A. Hasibuan *et al.*, “Implementasi Data Mining Untuk Pengaturan Layout,” vol. 4, no. 4, pp. 6–11, 2017.
- [2] Eric Fammaldo dan Lukman Hakim, “Penerapan Algoritma K-Means Clustering Untuk Pengelompokan Tingkat Kesejahteraan Keluarga Untuk Program Kartu Indonesia Pintar,” *J. Ilm. Teknol. Inf. Terap.*, vol. V, no. 1, pp. 24–32, 2018.
- [3] F. Yunita, “Penerapan Data Mining Menggunakan Algoritma K-Means Clustering Pada Penerimaan Mahasiswa Baru,” *Sistemasi*, vol. 7, no. 3, p. 238, 2018, doi: 10.32520/stmsi.v7i3.388.
- [4] L. U. Sholihannisa and P. Juliawati, “TEMATIK - Jurnal Teknologi Informasi Dan Komunikasi Vol. 3, No. 1 Juni 2016,” *J. Teknol. Inf. Dan Komun.*, vol. 3, no. 1, pp. 86–99, 2016.
- [5] E. Irfiani and S. S. Rani, “Algoritma K-Means Clustering untuk Menentukan Nilai Gizi Balita,” *J. Sist. dan Teknol. Inf.*, vol. 6, no. 4, p. 161, 2018, doi: 10.26418/justin.v6i4.29024.