

PENERAPAN ALGORITMA K-MEANS CLUSTERING UNTUK MENGELOMPOKKAN KENAKALAN SISWA

Harliana¹⁾

¹⁾ *Teknik Informatika Universitas Muhadi Setiabudi*
email : harliana.hifzhiya@gmail.com¹⁾

Abstraksi

Kenakalan remaja merupakan suatu keadaan dimana adanya perubahan emosional pada diri seorang dari masa kanak-kanak menjadi dewasa. Jenis-jenis kenakalan remaja pada SMK Bina Insani dapat dikategorikan menjadi kenakalan umum dan kenakalan khusus. Dari berbagai macam kenakalan tersebut, maka penelitian ini bertujuan untuk membantu Guru BK dalam melakukan pengelompokan kenakalan siswa melalui pendekatan k-means berdasarkan jumlah pelanggaran yang dilakukannya. Berdasarkan hasil pengujian pada 23 dataset maka didapatkan cluster ke-1 adalah cluster dimana siswa memiliki kenakalan umum paling banyak. Sedangkan cluster ke-3 adalah cluster yang memiliki kenakalan umum ataupun khusus paling sedikit bila dibandingkan dengan cluster lainnya.

Kata Kunci :

k-means clustering, kenakalan siswa, remaja

Abstract

Adolescence is a condition in which there is an emotional change in a person from childhood to adulthood. Types of adolescence in SMK Bina Insani can be categorized into general delinquency and special delinquency. Of the various types of delinquency, this study aims to assist BK teachers in grouping student delinquency through a k-means approach based on the number of violations committed. Based on the results of testing on 23 datasets, it is found that the 1st cluster is a cluster where students have the most common delinquency. While the third cluster is a cluster that has the least general or special delinquency when compared to other clusters.

Keywords :

k-means clustering, teenagers, adolescence

Pendahuluan

Masa remaja merupakan suatu fase dimana seseorang dihadapkan pada berbagai macam tantangan, masalah, ataupun pencarian identitas pengakuan diri dari masa kanak-kanak menjadi dewasa[1]. Saat ini kenakalan remaja bisa terjadi karena adanya faktor internal ataupun eksternal dari *self control* diri remaja itu sendiri[2]. Faktor internal dari kenakalan remaja dapat disebabkan karena pengaruh jenis kelamin, kedudukan, serta kepribadian remaja tersebut, sedangkan faktor eksternal dapat terjadi karena adanya pengaruh lingkungan keluarga, sekolah ataupun masyarakat[3]. Saat ini jenis kenakalan remaja yang terjadi pada SMK Bina Insan Mandiri Kota Cirebon dapat dibagi menjadi jenis kenakalan umum / ringan dan kenakalan khusus / berat. Dimana kategori dari jenis kenakalan umum diantaranya yaitu bolos saat jam pelajaran, tidak patuh kepada guru ataupun orangtua, berkelahi, membuat kegaduhan dikelas, terlambat masuk kelas, berbohong, dan berpakaian tidak pantas selama dilingkungan sekolah. Sedangkan kategori kenakalan khusus yaitu kebut-kebutan dijalan, mencuri, berkelahi antar kelompok / kelas, memiliki / membawa benda yang dapat membahayakan orang

lain, dan melakukan berbagai macam perbuatan yang meresahkan masyarakat.

Ketika seorang siswa melakukan kesalahan, maka siswa tersebut diwajibkan untuk melakukan konsultasi kepada Guru Bimbingan Konseling (Guru BK) atau konselor yang telah difasilitasi oleh pihak sekolah, hal ini bertujuan agar kenakalan yang terjadi pada diri remaja tidak dapat terulang bahkan tidak menjadi lebih buruk untuk kedepannya. Setelah melakukan pendampungan, para siswa yang bermasalah tetap akan dipantau oleh Guru BK selama 3 bulan kedepan, hal ini bertujuan untuk menganalisa sikap yang akan diberikan kepada siswa tersebut apakah akan diberikan sanksi berupa teguran ringan, pembinaan khusus oleh Guru BK, atau bahkan perlu diterapkannya hukuman kepada siswa sebagai efek jera.

Berdasarkan hal tersebut, maka penelitian ini bertujuan untuk membantu Guru BK dalam melakukan pengelompokan kenakalan siswa berdasarkan jumlah pelanggaran yang dilakukannya sehingga mempermudah Guru BK dalam mengambil tindakan yang sesuai. Penelitian ini menggunakan algoritma K-Means dalam pengelompokannya karena algoritma ini mampu melakukan perulangan partisi melalui banyaknya cluster dan data yang

diberikan berdasarkan fungsi jarak terpendek yang dihasilkan[4]. Sedangkan *confusion matrix* digunakan untuk mengetahui prosentase nilai akurasi yang dihasilkan oleh k-means.

Tinjauan Pustaka

Beberapa penelitian terdahulu yang penulis jadikan rujukan dalam penelitian ini diantaranya bahwa algoritma k-means pernah digunakan untuk mengelompokkan tingkat kepuasan siswa terhadap kinerja pelayanan sekolah dengan menggunakan 3 *cluster*, 5 atribut dan 80 dataset[5]. Dalam hal penentuan penerimaan beasiswa, algoritma k-means juga pernah digunakan untuk menentukan mahasiswa yang berhak mendapatkan beasiswa PPA UDINUS melalui 7 atribut dan 441 dataset, hasil penelitiannya menunjukkan bahwa prosentase akurasi yang dihasilkan oleh k-means dalam mengelompokkan mahasiswa yang berhak menerima dan tidak berhak menerima beasiswa PPA sebesar 82%[6]. K-means juga pernah digunakan untuk *clustering* pemilihan peserta olimpiade sains nasional tingkat SMA, dimana hasilnya menunjukkan bahwa iterasi akan berhenti pada iterasi ke-6 dengan jumlah anggota *cluster* ke-1 adalah 7 siswa, anggota *cluster* ke-2 adalah 14 siswa dan anggota *cluster* ke-3 adalah 2 siswa, namun pada penelitian ini tidak diketahui nilai akurasi k-means yang dihasilkan[7].

Nilai akurasi yang tinggi terhadap hasil pengelompokkan juga ditunjukkan oleh k-means dalam mengelompokkan harga motor, dimana dari hasil 15 kali pengujian terhadap 300 dataset jenis motor didapatkan nilai akurasi sebesar 81% dengan rata-rata *precision* 76% dan *recall* 76%[8]. Akurasi yang tinggi juga ditunjukkan oleh k-means dalam menganalisis *customer online reviews* pada toko *online* melalui perbandingan hasil pengelompokkan k-means dengan hasil pengelompokkan real yang mencapai 90% kesamaan[9].

Berdasarkan beberapa rujukan diatas, maka peneliti akan menggunakan k-means untuk mengelompokkan kenakalan remaja di SMK Bina Insan Mandiri Kota Cirebon menjadi 3 kelompok sehingga memudahkan Guru BK dalam memberikan solusi tindakan yang tepat.

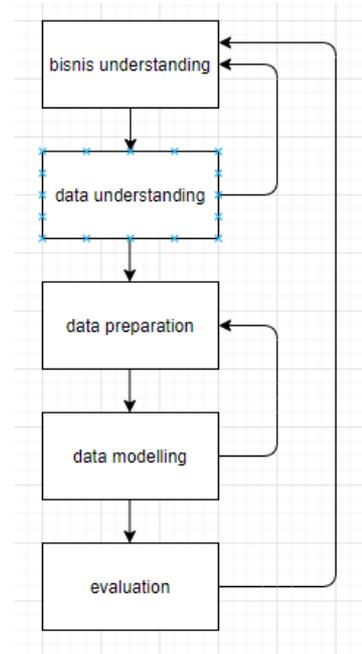
Metode Penelitian

Pada penelitian ini, penulis menggunakan pendekatan secara *crisp-dm* (*cross industry standart process for data mining*) dengan tahapan sesuai Gambar 1.

1. *Bisnis understanding*

Pada tahapan *bisnis understanding*, peneliti akan merumuskan tujuan dari penelitian yang akan dilakukan serta kebutuhan akan permasalahan yang dibahas. Berdasarkan hal tersebut maka penelitian ini bertujuan untuk membantu Guru BK dalam mengelompokkan kenakalan remaja yang terjadi di sekolah, tujuannya untuk

mempermudah Guru BK dalam memberikan tindakan yang akan dilakukan terhadap kenakalan yang dilakukan. Setelah ditemukannya permasalahan, maka selanjutnya peneliti akan melakukan wawancara langsung kepada Guru BK SMK Insan Mandiri Kota Cirebon mengenai beberapa perilaku kenakalan remaja yang terjadi di lingkungan sekolah.



Gambar 1. Tahapan penelitian

2. *Data Understanding*

Pada tahapan *data understanding*, peneliti akan melakukan evaluasi untuk data yang telah didapatkan sebelumnya. Data yang digunakan adalah data hasil konseling siswa selama 3 tahun yaitu tahun 2017 sampai dengan tahun 2019. Tingkatan siswa yang melakukan konseling kepada Guru BK tidak akan dibatasi hanya untuk siswa kelas 12 tetapi juga untuk siswa kelas 10 dan 11 yang pernah melakukan konseling.

3. *Data preparation*

Pada tahapan *preparation*, peneliti melakukan *data selection* terhadap data yang sudah didapatkan dengan cara menentukan atribut yang akan digunakan dalam proses *clustering*. Selanjutnya data akan di *preprocessing* yaitu dibersihkan dari data-data yang tidak valid, tidak sesuai dan *noisy* sehingga didapatkan struktur yang mudah dipahami.

4. *Data modeling*

Pada tahapan ini, data akan dihitung dengan k-means *clustering* untuk mendapatkan kelompok kenakalan remaja di SMK.

5. *Evaluation*

Pada tahapan ini, akan dilakukan pengukuran akurasi dan pengujian terhadap algoritma yang diterapkan.

Hasil dan Pembahasan

Pada penelitian ini akan menggunakan 54 data, namun setelah dilakukan *cleaning* pada tahapan *preparation*, maka data yang akan digunakan hanya berjumlah 23, sedangkan atribut yang digunakan terangkum pada Tabel 1.

Tabel 1. Atribut perhitungan

No	Jenis Atribut
1	Jumlah pelanggaran umum
2	Jumlah pelanggaran khusus
3	Prestasi akademik
4	Pendapat walikelas

Berdasarkan Tabel 1, maka jumlah pelanggaran umum/khusus adalah total dari jumlah pelanggaran yang dilakukan siswa selama 3 bulan. Prestasi akademik adalah prestasi siswa saat sebelum atau selama terjadinya pelanggaran. Sedangkan pendapat walikelas merupakan pendapat dari walikelas terhadap siswa yang melakukan pelanggaran, nilai 0 akan diberikan ketika walikelas berpendapat siswa tersebut sering melakukan pelanggaran dan nilai 1 akan diberikan ketika walikelas berpendapat siswa jarang melakukan pelanggaran.

Pada penelitian ini jumlah cluster yang digunakan adalah 3 ($k = 3$), dan centroid awal yang digunakan adalah titik tengah dari keseluruhan data yaitu $C_1(9,3)$; $C_2(4,2)$; dan $C_3(1,1)$. Selanjutnya akan dihitung jarak dari data ke pusat cluster dengan persamaan (1)

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

jarak data siswa pertama dengan centroid pertama adalah:

$$d(1,1) = \sqrt{(1-9)^2 + (2-3)^2} = 8,06$$

$$d(1,2) = \sqrt{(1-4)^2 + (2-2)^2} = 3,00$$

$$d(1,3) = \sqrt{(1-1)^2 + (2-1)^2} = 1,00$$

dan seterusnya.

Rangkuman mengenai hasil perhitungan terangkum pada Tabel 2.

Tabel 2. Perhitungan jarak terhadap data

Nama	K ₁	K ₂	K ₃	K ₄	C ₁	C ₂	C ₃
1	1	2	1	0	8,06	3,00	1,00
2	5	2	1	0	4,12	1,00	4,12
3	3	1	1	0	6,32	1,41	2,00
4	4	1	1	1	5,39	1,00	3,00
5	7	2	1	1	2,24	3,00	6,08
6	2	1	1	0	7,28	2,24	1,00
7	7	1	1	1	2,83	3,16	6,00
8	4	3	1	0	5,00	1,00	3,61
9	5	2	1	1	4,12	1,00	4,12
10	6	2	1	1	3,16	2,00	5,10
...

Berdasarkan Tabel 2, maka pada nama ke-1 akan masuk kedalam C_3 karena memiliki jarak terkecil dengan pusat cluster yang telah ditentukan. Sedangkan pada nama ke-2 akan masuk kedalam C_2 dan seterusnya.

Langkah selanjutnya yaitu akan dihitung centroid baru untuk masing-masing kriteria yang ditentukan. Centroid baru untuk data siswa yang melakukan pelanggaran umum (K_1) adalah:

$$C_{1,1} = (7+7+7+9)/4 = 7,50$$

$$C_{1,2} = (2+1+3+3)/4 = 2,25$$

Nilai 4 merupakan jumlah data yang masuk kedalam cluster 1.

Sedangkan centroid baru untuk data siswa yang melakukan pelanggaran khusus (K_2) adalah:

$$C_{2,1} = (5+3+4+4+5+5+6+3+5+3+3+5+5+5+3)/15 = 4,27$$

$$C_{2,2} = (2+1+1+3+2+1+2+1+2+1+2+1+2+3+1)/15 = 1,67$$

Dan seterusnya sampai titik centroid baru tidak mengalami perubahan. Pada penelitian ini iterasi akan berhenti pada iterasi ke-6 yaitu saat titik centroid sudah tidak mengalami perubahan. Rangkuman mengenai hasil iterasi ke-6 terangkum pada table 3.

Tabel 3. Hasil akhir saat iterasi ke-6

Nama	C ₁	C ₂	C ₃	C ₁	C ₂	C ₃
1	6,2	3,7	1,5			√
	0	8	5			
2	2,2	0,2	2,6		√	
	1	5	2			
3	4,3	1,9	0,7			√
	7	9	1			
4	3,4	1,1	1,6		√	
	2	8	2			
5	0,2	2,2	4,5	√		
	8	2	9			
6	5,3	2,2	0,6			√
	4	9	3			
7	1,2	2,3	4,5	√		
	2	9	8			
8	3,3	1,3	2,2		√	
	0	6	0			
9	2,2	0,2	2,2		√	
	1	5	6			
10	2,5	0,2	2,5		√	
	1	9	9			
...
23	4,3	1,9	0,7			√
	7	9	1			

Berdasarkan Tabel 3 maka didapatkan bahwa anggota untuk cluster ke-1 yaitu siswa dengan no urut: 5, 7, 11, 15, 18. Sedangkan anggota untuk cluster ke-2 yaitu 2, 4, 8, 9, 10, 14, 19, 20, 21. Sedangkan untuk anggota cluster ke-3 yaitu 1, 3, 6, 12, 13, 16, 17, 22, 23.

Cluster ke-1 merupakan cluster yang memiliki jumlah pelanggaran umum yang tinggi yaitu diantara

7-9 dan pelanggaran khusus 1-3. Sehingga teguran yang akan diberikan dapat berupa hukuman.

Cluster ke-2 merupakan cluster yang memiliki jumlah pelanggaran umum 3-5 dan jumlah pelanggaran khusus 1-3 sehingga perlu dilakukan pembinaan terhadap siswa.

Cluster ke-3 merupakan cluster yang memiliki jumlah pelanggaran umum diantara 2-3 dengan jumlah pelanggaran khusus 1-3. Sehingga dapat dikategorikan menjadi kenakalan remaja yang wajar dengan hanya memerlukan teguran ringan.

Setelah didapatkan hasil pengelompokkan, maka langkah selanjutnya yaitu menghitung akurasi melalui precision dan recall yaitu:

Precision = $0,73 = 73\%$

Recall = $0,69 = 69\%$

Dan accuracy = $0,82 = 82\%$

Kesimpulan dan Saran

Berdasarkan penelitian yang telah dilakukan melalui algoritma k-means pada 23 data siswa dengan $k=3$, atribut 4 dan iterasi yang dihasilkan adalah 6 maka didapat bahwa cluster ke-1 merupakan cluster yang memiliki anggota siswa dengan jumlah pelanggaran umum terbanyak bila dibandingkan dengan cluster lainnya sehingga Guru BK dapat memberikan teguran berupa hukuman, sedangkan untuk cluster ke-2 dan ke-3 merupakan cluster dengan jumlah pelanggaran ringan yaitu kurang dari 5. Sedangkan apabila dilihat dari akurasi yang dihasilkan maka algoritma k-means memiliki akurasi sebesar 82%. Meskipun tingkat akurasi sudah mencapai 82% namun alangkah baiknya penelitian ini dapat dibandingkan dengan berbagai macam algoritma pengelompokkan lainnya.

Daftar Pustaka

- [1] H. A. Bakhrudin, "Model Konseling Kelompok Cognitive Behavior untuk Penanganan Self-esteem Siswa SMK," *J. Bimbingan dan Konseling Terap.*, vol. 1, no. 2, pp. 86–102, 2017, doi: 10.30598/jbkt.v1i2.144.
- [2] E. Indrawati and S. Rahimi, "Fungsi Keluarga dan Self Control Terhadap Kenakalan Remaja," *J. IKRAITH-HUMANIORA*, vol. 3, no. 2, pp. 86–93, 2019.
- [3] S. Prasasti, "Kenakalan remaja dan faktor penyebabnya," *Semin. Nas. Bimbingan dan Konseling*, vol. 1, no. 1, pp. 28–45, 2017, [Online]. Available: prosiding.unipma.ac.id/index.php/SNBK/article/download/110/109.
- [4] M. Robani and A. Widodo, "Algoritma K-Means Clustering Untuk Pengelompokan Ayat Al Quran Pada Terjemahan Bahasa Indonesia," *J. Sist. Inf. Bisnis*, vol. 6, no. 2, p. 164, 2016, doi: 10.21456/vol6iss2pp164-

- 176.
- [5] M. G. Suryanata, D. H. Pane, and M. Hutasuhut, "Implementasi Algoritma K-Means Untuk Mengukur Tingkat Kepuasan Siswa Terhadap Proses Pembelajaran," *J-Sisko Tech J. Teknol. Sist. Inf. dan Sist. Komput. TGD*, vol. 2, no. 2, pp. 118–125, 2019.
- [6] A. Salam, D. Adiatma, and J. Zeniarja, "Implementasi Algoritma K-Means Dalam Pengklasteran untuk Rekomendasi Penerima Beasiswa PPA di UDINUS," *JOINS (Journal Inf. Syst.*, vol. 5, no. 1, pp. 62–68, 2020, doi: 10.33633/joins.v5i1.3350.
- [7] M. Hasanah, S. Defit, and G. W. Nurcahyo, "Implementasi Algoritma K-Means untuk Klasterisasi Peserta Olimpiade Sains Nasional Tingkat SMA," *J. Sistim Inf. dan Teknol.*, vol. 1, no. 3, pp. 31–36, 2019, doi: 10.35134/jsisfotek.v1i3.7.
- [8] R. K. Dinata, S. Safwandi, N. Hasdyna, and N. Azizah, "Analisis K-Means Clustering pada Data Sepeda Motor," *INFORMAL Informatics J.*, vol. 5, no. 1, p. 10, 2020, doi: 10.19184/isj.v5i1.17071.
- [9] R. Nainggolan and E. Purba, "Perbaikan Performa Cluster K-Means Menggunakan SUM Squared Error (SSE) Pada Analisis Online Customer Review Terhadap Produk Toko Online," *J. Times Technol. Inform. Comput. Syst.*, vol. 7, no. 2, pp. 1–8, 2019, doi: 10.1017/CBO9781107415324.004.