



PENGELOMPOKAN MENGGUNAKAN ALGORITMA K-MEDOID UNTUK EVALUASI PERFORMA SISWA

Yoga Religia¹, Rifki Tia Bayu Jaya²

**Program Studi Teknik Informatika, Universitas Pelita Bangsa
Kabupaten Bekasi, Indonesia**

Korespondensi email: yoga.religia@pelitabangsa.ac.id

Abstrak

Pendidikan merupakan salah satu pondasi dalam kemajuan suatu bangsa, semakin baik kualitas pendidikan yang diselenggarakan oleh suatu bangsa, maka akan diikuti dengan semakin baiknya kualitas bangsa tersebut. Saat ini, siswa pada jenjang SMA banyak mendapat sorotan. Palsanya tingkat kenakalan remaja banyak terjadi pada masa-masa SMA. Hal tersebut memerlukan evaluasi performa siswa untuk meningkatkan prestasi siswa dan menghindarkan siswa dari tindakan negatif. Salah satu data yang dapat digunakan terkait performa siswa adalah menggunakan Student Academics Dataset. Student Academics terdiri dari 21 atribut, dimana atribut-atribut didalamnya adalah berupa gender hingga Home to College Travel Time dalam bentuk unsupervised learning. Salah satu algoritma yang cukup baik digunakan untuk proses clustering adalah algoritma k-medoid. Algoritma k-medoid merupakan bagian dari partitioning clustering dimana k-medoids cukup efisien untuk diterapkan pada dataset yang kecil. Pada penelitian ini pengujian dilakukan menggunakan waktu rata-rata eksekusi dan davies bouldin indek. Hasil dari pengujian menunjukkan pengelompokan data Student Academics kedalam 3 status performa siswa menggunakan algoritma k-medoid memiliki akumulasi waktu rata-rata eksekusi selama 22,8 detik dan nilai Index Davies dari algoritma k-medoid untuk pengelompokan performa siswa sebesar 0,367.

Informasi Artikel

Diterima: 31-02-2020

Direvisi: 05-03-2020

Dipublikasikan: 25-04-2020

Kata Kunci

Data mining, pengelompokan, k-medoid, Euclidean distance

I. Pendahuluan

Pendidikan merupakan salah satu pondasi dalam kemajuan suatu bangsa, semakin baik kualitas pendidikan yang diselenggarakan oleh suatu bangsa, maka akan diikuti dengan semakin baiknya kualitas bangsa tersebut. Pendidikan SMA adalah sebuah lembaga pendidikan yang sangat penting untuk anak-anak bangsa. Saat ini, siswa pada jenjang SMA banyak mendapat sorotan. Palsunya tingkat kenakalan remaja banyak terjadi pada masa-masa SMA [1]. Hal tersebut memerlukan evaluasi performa siswa untuk meningkatkan prestasi siswa dan menghindarkan siswa dari tindakan negatif.

Pada situs UCI dataset terdapat data-data yang dapat diakses secara terbuka untuk digunakan dalam hal penelitian. Salah satu data yang dapat digunakan terkait performa siswa adalah menggunakan *Student Academics* Dataset. *Student Academics* terdiri dari 21 atribut, dimana atribut-atribut didalamnya adalah berupa *gender* hingga *Home to College Travel Time* dalam bentuk *unsupervised learning* [2]. Salah satu teknik yang dapat digunakan untuk mengolah data *unsupervised learning* adalah Teknik data mining.

Adanya jumlah data yang besar dan dapat diakses secara mudah mendorong para peneliti untuk menjadikan data tersebut menjadi suatu informasi yang dapat dimanfaatkan [3]. Metode yang dapat digunakan untuk mengolah data *unsupervised learning* dengan Teknik data mining adalah dengan metode *clustering* (pengelompokan). Salah satu algoritma yang cukup baik digunakan untuk proses *clustering* adalah algoritma k-medoid.

Algoritma k-medoid merupakan bagian dari *partitioning clustering* dimana k-medoids cukup efisien untuk diterapkan pada dataset yang kecil [4]. Algoritma k-medoid hadir untuk mengatasi kelemahan Algoritma k-means yaitu sensitif terhadap outlier karena suatu objek dengan suatu nilai yang besar mungkin secara substansial menyimpang dari distribusi data [5] [6]. Algoritma k-medoids tidak menentukan nilai rata-rata

dari objek dalam *cluster* sebagai titik acuan, tetapi menggunakan medoid (median), yang merupakan objek yang paling terletak dipusat sebuah *cluster*. Dengan demikian, metode partisi masih dapat dilakukan berdasarkan prinsip meminimalkan jumlah dari ketidaksamaan antara setiap objek dan titik acuan yang sesuai (medoid) [7]. Selain itu k-medoid juga memiliki waktu eksekusi yang relative cepat, sehingga dapat dimanfaatkan untuk pengelompokan data *Student Academics*.

Berdasarkan uraian yang sudah disampaikan pada paragraph-paragraf sebelumnya, pada penelitian ini akan dilakukan “Pengelompokan Menggunakan Algoritma K-Medoid untuk Evaluasi Performa Siswa”.

II. Metodologi

2.1. Data Mining

Data mining merupakan salah satu bidang paling penting dalam penelitian yang bertujuan untuk memperoleh informasi dari data set. Data mining mulai ada sejak 1990-an sebagai cara yang efektif untuk mengambil pola dan informasi yang sebelumnya tidak diketahui dari suatu data set [8]. Teknik data mining digunakan untuk menemukan hubungan antara data untuk melakukan pengklasifikasian yang memprediksikan nilai-nilai dari beberapa variabel (klasifikasi), atau untuk membagi data yang diketahui menjadi kelompok-kelompok yang mempunyai kesamaan karakteristik (*clustering*). Data mining merupakan bagian dari proses penemuan pengetahuan dari basis data (*Knowledge Discovery in Databases*) [9].

2.2. Clustering

Clustering merupakan bagian dari pembelajaran *unsupervised* yang digunakan sebagai alat yang efektif dalam data mining [10]. *Clustering* tidak memiliki atribut yang dapat digunakan untuk memandu proses pembelajaran (tidak memiliki label), sehingga seluruh atribut yang dimiliki diperlakukan sama. Metode *clustering*

memungkinkan untuk mengungkapkan hubungan dan struktur yang sebelumnya tidak jelas dari *data-set*. Tujuan dari *clustering* adalah untuk mengelompokkan data yang memiliki kesamaan karakteristik kedalam kelompok yang sama dan data yang berbeda karakteristik kedalam kelompok yang lain. Algoritma *clustering* dapat dikategorikan kedalam empat metode yaitu *Partitional-based*, *Hierarchical-based*, *Density-based* dan *Grid-based* [11].

2.3. Algoritma k-medoid

Algoritma k-medoids tidak menentukan nilai rata-rata dari objek dalam *cluster* sebagai titik acuan, tetapi menggunakan medoid (median), yang merupakan objek yang paling terletak dipusat sebuah *cluster*. Dengan demikian, metode partisi masih dapat dilakukan berdasarkan prinsip meminimalkan jumlah dari ketidaksamaan antara setiap objek dan titik acuan yang sesuai (medoid). Hal ini merupakan dasar dari metode k-medoids. Strategi dasar dari algoritma k-medoids adalah untuk menemukan *cluster* k pada objek n dengan terlebih dahulu menemukan objek awal (medoid) secara acak sebagai perwakilan untuk setiap *cluster* [12]. Setiap objek yang tersisa dikelompokkan dengan medoid yang paling mirip.

Adapun secara diskriptif langkah-langkah dari algoritma k-medoids adalah sebagai berikut [13]:

1. Secara acak memilih k objek dari D sebagai titik acuan;

Repeat:

2. Tetapkan setiap objek yang tersisa ke dalam *cluster* dengan medoid yang paling dekat;
3. Pilih secara acak objek non-medoid : O_{random} ;
4. Hitung total jarak poin S dari pertukaran objek O_j dengan O_{random} ;
5. Jika, $S < 0$ kemudian tukar O_j dengan O_{random} untuk membentuk set baru k medoid;
6. **Until:** tidak ada perubahan

2.4. Davies Bouldin Index (DBI)

Davies Bouldin Index (DBI) merupakan cara validasi *cluster* yang dibuat oleh D.L. Davies. DBI adalah fungsi rasio dari jumlah distribusi didalam *cluster* untuk pemisahan antar *cluster* [14]. Pengukuran menggunakan DBI bertujuan untuk memaksimalkan jarak *inter-cluster*. Dalam penelitian ini, DBI digunakan untuk melakukan validasi data pada setiap *cluster*. DBI dapat dihitung menggunakan persamaan:

$$R_i = \max_{j=1 \dots k, i \neq j} R_{ij} \quad (1)$$

$$\text{var}(x) = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2 \quad (2)$$

$$R_{ij} = \frac{\text{var}(C_i) + \text{var}(C_j)}{\|c_i - c_j\|} \quad (3)$$

$$DB = \frac{1}{k} \sum_{i=1}^k R_i \quad (4)$$

Keterangan:

- R : jarak antar *cluster*
- Var : *variance* dari data
- x : data ke-i
- \bar{x} : rata-rata dari tiap *cluster*
- DB : validasi Davies Bouldin

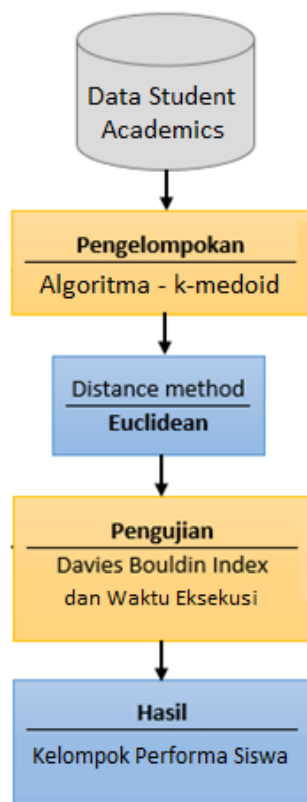
Dengan menggunakan *Davies Bouldin Index* suatu *cluster* akan dianggap memiliki skema *clustering* yang optimal jika memiliki *Index Davies Bouldin* minimal.

2.5. Model Penelitian

Dalam penelitian ini metode yang digunakan adalah metode data mining dengan memakai teknik *clustering*/pengelompokan menggunakan algoritma k-medoid dengan perhitungan jarak Euclidean. Kemudian dari metode tersebut akan digunakan untuk mengelompokkan performa siswa berdasarkan data *Student Academics*. Dengan menggunakan algoritma k-medoid, data *Student Academics* akan dibagi kedalam 3 *cluster* yaitu *cluster 0*, *cluster 1*,

dan *cluster* 2. Penentuan 3 *cluster* ini dikarenakan ukuran performa siswa akan di kategorikan kedalam 3 kelas yaitu Bagus, Rata-rata, dan Kurang. Adapun teknik *clustering* yang digunakan dalam penelitian ini dapat dilihat pada gambar 1.

Pada gambar 1 menunjukkan bahwa teknik *clustering* diawali dari *input* berupa data *Student Academics* yang belum dikelompokkan kedalam kategori performa siswa. Kemudian dari data *input* akan diproses menggunakan algoritma k-medoid.



Gambar 1. Teknik clustering menggunakan k-medoid

Adapun didalam algoritma k-medoid akan menggunakan metode perhitungan jarak untuk menentukan *centroid* dari setiap *cluster*. Metode perhitungan jarak yang digunakan yaitu Euclidean. Dari proses pengelompokan akan menghasilkan 3 *cluster* performa siswa, kemudian hasil akan di uji menggunakan *Davies Bouldin Index* dan waktu eksekusi untuk mengetahui performa algoritma.

III. Hasil dan Pembahasan

3.1 Hasil Cluster

Dari pengujian yang telah dilakukan diperoleh nilai *centroid* dan jumlah data pada setiap *cluster*. Dari setiap *cluster* yang diperoleh memiliki nilai *centroid*, dimana *centroid* merupakan “titik tengah” dari *cluster*. Sehingga pada penelitian ini untuk menentukan status performa siswa dilakukan dengan mengitung jumlah dari *centroid* untuk setiap *cluster*. Selanjutnya penentuan status performa siswa akan diurutkan berdasarkan nilai penjumlahan *centroid* dari setiap indikator pada masing-masing *cluster*, yang mana nilai penjumlahan terendah akan diinisialisasikan sebagai status performa siswanya kurang dan nilai penjumlahan tertinggi akan diinisialisasikan sebagai status performa siswanya bagus.

Adapun nilai *centroid* dan jumlah data pada setiap *cluster* dari pengujian metode perhitungan jarak Euclidean pada algoritma k-medoid yang telah dilakukan adalah sebagai berikut:

Tabel 1. Centroid pada setiap cluster

Atribut	C ₀	C ₁	C ₂
ge	1	0	0
cst	0	1	0
tnp	1	0	0
twp	0	2	0
iap	1	0	0
...
...
tt	0	0	0
Jumlah	13	19	14

Dari penggunaan algoritma k-medoid dengan metode perhitungan jarak Euclidean *distance* untuk mengelompokan data *Student Academics* yang berjumlah 909 *record*, diperoleh jumlah data pada setiap *cluster* sebagai berikut:

- *Cluster* 0: 426 siswa
- *Cluster* 1: 292 siswa
- *Cluster* 2: 191 siswa

Apabila dilihat dari jumlah *centroid* pada setiap *cluster* yang dihitung jumlah centroid pada setiap *cluster*, dapat diperoleh status performa siswa dari pengelompokan k-medoid dengan metode perhitungan jarak Euclidean *distance* yang ditunjukkan pada Tabel 2.

Tabel 2. Status performa siswa dan Jumlah siswa pada setiap cluster

Cluster	Jumlah Centroid	Status performa siswa	Jml Siswa
C ₀	13	Kurang	426
C ₁	19	Bagus	292
C ₂	14	Rata-rata	191

3.2 Hasil Pengujian Waktu Eksekusi

Akumulasi waktu dilakukan dengan melakukan eksekusi sebanyak 5 kali. Dari 5 kali eksekusi tersebut kemudian akan dirata-rata untuk memperoleh rata-rata waktu eksekusi. Dari pengujian yang telah dilakukan diperoleh lama waktu eksekusi yang berbeda, adapun lama waktu eksekusi dari pengujian pengelompokan performa siswa menggunakan algoritma k-medoid dengan metode perhitungan jarak Euclidean yang telah dilakukan dapat dilihat pada Tabel 3.

Tabel 3. Lama Waktu Eksekusi

Pengujian	Waktu Eksekusi K-Medoid
1	20 detik
2	19 detik
3	28 detik
4	25 detik
5	22 detik
Rata-rata	22,8 detik

Pada Tabel 3 dapat dilihat bahwa waktu eksekusi dari algoritma k-medoid untuk pengujian 1 sampai pengujian 5 secara berturut-turut yaitu 20 detik, 19 detik, 28 detik, 25 detik dan 22 detik, sehingga apabila diambil rata-rata waktu eksekusi dari algoritma k-medoid untuk pengelompokan performa siswa adalah 22,8 detik. Hasil tersebut menunjukkan bahwa

algoritma k-medoid cukup cepat untuk digunakan dalam mengelompokkan status performa siswa.

3.3 Hasil Pengujian DBI

Dalam penelitian ini, *Davies Bouldin Index* (DBI) digunakan untuk melakukan validasi data pada setiap *cluster*. Pengukuran menggunakan DBI bertujuan untuk memaksimalkan jarak *inter-cluster*. Dengan menggunakan DBI suatu *cluster* akan dianggap memiliki skema *clustering* yang optimal jika yang memiliki *Index Davies* minimal. Adapun dari pengujian yang sudah dilakukan diperoleh nilai *Index Davies* dari algoritma k-medoid untuk pengelompokan performa siswa sebesar 0,367. Dengan nilai DBI sebesar 0,367 artinya *cluster* yang dihasilkan memiliki kerapatan antar *cluster* sebesar 63.3%. hasil tersebut sudah cukup baik karena kerapatan *cluster* sudah lebih dari 50% meskipun terdapat banyak atribut yang digunakan.

IV. Kesimpulan dan Saran

4.1 Kesimpulan

Dari pengujian algoritma k-medoid untuk pengelompokan performa siswa yang telah dilakukan dihasilkan:

1. Model pengujian yang digunakan dapat berjalan dengan baik dan menunjukkan hasil berupa nilai *centroid* untuk setiap *cluster* dari algoritma k-medoid, sehingga status performa siswa dapat ditentukan dari jumlah *centroid* pada setiap *cluster*.
2. Akumulasi waktu yang diperoleh dari pengujian yang telah dilakukan menunjukkan bahwa penggunaan algoritma k-medoid untuk pengelompokan performa siswa memiliki waktu eksekusi cukup efisien yaitu dengan rata-rata waktu eksekusi selama 22,8 detik.
3. Nilai *Index Davies* dari algoritma k-medoid untuk pengelompokan performa siswa sebesar 0,367. Dengan nilai DBI sebesar 0,367 artinya *cluster* yang dihasilkan memiliki kerapatan antar *cluster* sebesar 63.3%. hasil

tersebut sudah cukup baik karena kerapatan antar *cluster* sudah lebih dari 50% meskipun terdapat banyak atribut yang digunakan.

Dari pengujian yang sudah dilakukan dapat diketahui bahwa pengelompokan data *Student Academics* menggunakan algoritma k-medoid memiliki akumulasi waktu rata-rata eksekusi yang cukup cepat dan memiliki kualitas *cluster* yang baik, sehingga dari penelitian ini diperoleh *cluster* status performa siswa sebagai berikut:

- *Cluster* Status performa siswa **kurang** sebanyak 426 siswa
- *Cluster* Status performa siswa **rata-rata** sebanyak 191 siswa
- *Cluster* Status performa siswa **bagus** sebanyak 292 siswa

4.2 Saran

Dari pengelompokan menggunakan algoritma k-medoid untuk evaluasi performa siswa ada hal yang perlu diperhatikan supaya menjadi lebih baik kedepannya, antara lain yaitu: dari penelitian ini diketahui bahwa setiap *cluster* yang dihasilkan memiliki jarak dari *cluster* yang lain, dengan hasil *cluster* yang diperoleh diharapkan dapat diukur jarak setiap atribut antar *cluster* agar dapat ditentukan atribut mana yang perlu diperhatikan sebagai prioritas evaluasi agar status performa siswa dapat meningkat.

Daftar Pustaka

- [1] ujiansma.com, "ujiansma.com," 16 January 2013. [Online]. Available: <http://ujiansma.com/pentingnya-pendidikan-lanjutan-bagi-siswa-sma>. [Accessed 04 Maret 2020].
- [2] U. Dataset, "Student Academics Data Set," Maret 2020. [Online]. Available: archive.ics.uci.edu/ml/datasets/Student+Academics+Performance.
- [3] S. Hussain, N. A. Dahan, F. M. Ba-Alwi and N. Ribata, "Educational Data Mining and Analysis of Students' Academic Performance Using WEKA," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 9, no. 2, pp. 447-459, 2018.
- [4] D. Asima S, A. Perdana W, D. Hartama and Sumarno, "Penerapan Metode K-Medoid pada Pengelompokan Rumah Tangga Dalam Perlakuan Memilah Sampah Menurut Provinsi," in *Seminar Nasional Sains & Teknologi Informasi (SENSASI)*, Medan, Indonesia, 2019.
- [5] T. Wiwit A, "Algoritma K-Medoids Untuk Penentuan Strategi Pemasaran," *Jurnal Teknik Industri*, vol. 6, no. 1, p. 183-188, 2015.
- [6] I. Parlina, A. Perdana W, A. Wanto and M. R. Lubis, "Memanfaatkan Algoritma K-Means dalam Menentukan Pegawai yang Layak Mengikuti Assessment Center untuk Clustering Program SDP," *Journal of Computer Engineering, System and Science*, vol. 3, no. 1, p. 87-93, 2018.
- [7] S. Defiyanti, N. Rohmawati W and M. Jajuli, "K-Medoid Algorithm in Clustering Student Scholarship Applicants," *Scientific Journal of Informatics*, vol. 4, no. 1, pp. 27-33, 2017.
- [8] D. Tomar and S. Agarwal, "A survey on Data Mining approaches for Healthcare," *International Journal of Bio-Science and Bio-Technology*, vol. 5, no. 5, pp. 241-266, 2013.
- [9] M. Cannataro, P. H. Guzzi and A. Carica, "Data mining and life sciences applications on the grid,"

- WIREs Data Mining and Knowledge Discovery*, p. 216–238, May 2013.
- [10] S. Ding, F. Wu, Q. Jun, H. Jia and F. Jin, "Research on data stream clustering algorithms," *Artificial Intelligence Review*, vol. 43, no. 4, pp. 593-600, 2013.
- [11] A. K. Mann and N. Kaur, "Review Paper on Clustering Techniques," *Global Journal of Computer Science and Technology Software & Data Engineering*, vol. 13, no. 5, 2013.
- [12] S. Defiyanti¹, M. Jajuli and N. Rohmawati W, "Optimalisasi K-Medoid Dalam Pengklasteran Mahasiswa Pelamar Beasiswa Dengan Cubic Clustering Criterion Clustering Criterion," *Teknosi*, vol. 3, no. 1, pp. 211-218, 2017.
- [13] J. Han and M. P. J. Kamber, "Jiawei Han Micheline Kamber Jian Pei," in *Data Mining: Concepts and Techniques*, Morgan Kaufmann, Elsevier, 2011, p. 744.
- [14] B. K. Mishra, A. Rath, N. R. Nayak and S. Swain, "Far Efficient K-Means Clustering Algorithm," in *International Conference on Advances in Computing, Communications and Informatics*, Chennai, 2012.
- [15] T. Sajana, C. M. S. Rani and K. V. Narayana, "A Survey on Clustering Techniques for Big Data Mining," *Indian Journal of Science and Technology*, vol. 9, no. 3, pp. 1-12, 2016.