

SEGMENTASI MAHASISWA BERDASARKAN KARAKTERISTIK POLA BELAJAR MENGUNAKAN METODE K-MEDOIDS CLUSTERING

Rizki Adi Candra^{1*}, Yulison Herry Chrisnanto², dan Puspita Nurul Sabrina³

¹Jurusan Informatika, Fakultas Sains dan Informatika, Universitas Jenderal Achmad Yani,

*Email : rizkiadicandra3@gmail.com, yhc@if.unjani.ac.id, puspita.sabrina@lecture.unjani.ac.id

Abstrak

Data mining kini digunakan diberbagai bidang termasuk pendidikan. Data mining memiliki beberapa teknik, salah satunya yang populer yaitu teknik Clustering. Kajian terkait kedataan sudah banyak dilakukan, salah satu kajian kedataan yang menarik adalah eksplorasi data dalam jumlah yang besar, dan diproses sedimikian rupa untuk diekstrak menjadi kumpulan informasi yang sebelumnya belum diketahui serta memiliki nilai yang tinggi bagi kebutuhan analisis dalam rangka pengambilan keputusan. Kajian kedataan ini dikenal sebagai data mining. Penelitian ini akan melakukan kajian kedataan, khususnya data karakteristik pola belajar mahasiswa di Unjani angkatan 2015/2019. Pendekatan metodologi yang digunakan adalah Clustering Data Mining untuk mendapatkan segmentasi karakteristik pola belajar mahasiswa Unjani. Teknik Clustering yang akan digunakan adalah K-Medoids yaitu pendekatan dalam mempartisi data sesuai dengan kedekatan karakteristik objek data dalam dataset karakteristik pola belajar mahasiswa Unjani. Hasil akhir dari penelitian ini adalah segmentasi karakteristik pola belajar mahasiswa Unjani. Adapun hasil dari penggunaa teknik Clustering dengan algoritma K-Medoids, menunjukkan hasil teknik Clustering dengan algoritma K-Medoids mampu melakukan pengelompokan data terhadap data karakteritik pola belajar. Hasil pengujian cluster didapatkan jumlah $K=2$ yang dimana nilai silhoutte Coefficient yang paling tinggi mendekati $S_i = 1$ pada nilai $K = 2$ yaitu 0.80 yang merupakan struktur kuat Kaufman dan Rousseeuw.

Kata Kunci: Algoritma K-Medoids, Clustering, Data Mining, Karakteristik Pola Belajar

1. PENDAHULUAN

Jenjang pendidikan di Indonesia sangatlah beragam mulai dari pendidikan anak usia dini, taman kanak-kanak, sekolah dasar (SD), sekolah menengah pertama(SMP), sekolah menengah atas (SMA) dan pendidikan perguruan tinggi. Pendidikan pada perguruan tinggi harus mampu menciptakan lulusan yang memiliki skill atau keterampilan yang unggul pada bidangnya dan mampu bersaing di tingkat nasional ataupun internasional(Annizar and Arifin, 2021).Prestasi tersebut dipantau setiap semester dalam bentuk nilai yang kerap disebut sebagai Indeks Prestasi (IP). Nilai ini merupakan nilai yang diberikan dosen sebagai bentuk apresiasi yang dicapai dari setiap matakuliah yang telah dilaksanakan. Sedangkan, akumulasi dari nilai IP setiap semester disebut Indeks Prestasi Kumulatif (IPK)(Murti and Prasetyo, 2018). Namun setiap mahasiswa harus memiliki pengetahuan dan keterampilan sesuai dengan standar yang ditetapkan pada bidang ilmu tertentu. Sehingga perlu dilakukan upaya penanganan secara khusus salah satunya adalah mengidentifikasi karakteristik pola belajar mahasiswa.

Secara umum karakteristik pola belajar bisa juga diartikan sebagai cara atau teknik yang digunakan oleh seseorang untuk berinteraksi, menyerap, memahami, mengorganisasi, dan memproses suatu informasi. Dalam proses kegiatan pembelajaran, Seorang mahasiswa memiliki cara atau karakteristik gaya belajar yang berbeda-beda sehingga hasil belajar yang diperolehnya juga bervariasi(Vermunt and Donche, 2017).

Dalam masalah ini penerapan data mining dapat dilakukan untuk mengumpulkan suatu data informasi dari mahasiswa informatika UNJANI. Sumber data tersebut diolah untuk mendapatkan knowledge yang mungkin bisa menghasilkan teknik data mining. Fokus dalam penelitian ini yaitu segmentasi atau clustering, yang mana data informasi yang di dapatkan akan di kelompokkan sesuai cluster yang telah di tentukan untuk melihat karakteristik pola belajar dan akan di perkuat oleh data nilai

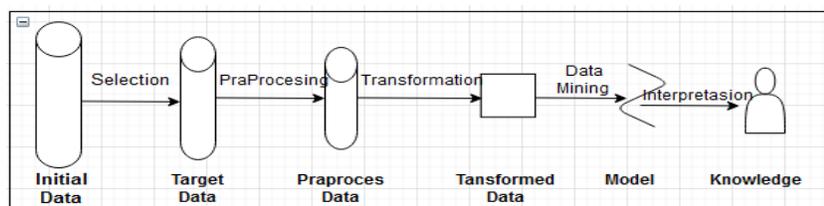
akademik mahasiswa. Di dalam penelitian ini penulis melakukan segmentasi untuk melihat karakteristik gaya belajar mahasiswa menggunakan *teknik clustering*, menggunakan metode K-Medoids.

2. METODOLOGI

Pada bagian ini akan dijelaskan topik - topik tertentu berdasarkan review yang telah dilakukan terhadap beberapa literatur. Materi yang ditinjau dari literatur meliputi informasi tentang Data Mining, Algoritma K-Medoids, mencari K optimal (elbow) dan pengujian Silhouette Coefficient.

2.1. Data Mining

Data mining adalah proses menganalisis data dari perspektif yang berbeda dan mengubahnya menjadi informasi penting yang dapat digunakan untuk meningkatkan keuntungan, mengurangi pengeluaran, atau bahkan keduanya (Rahman *et al.*, 2020). Secara teknis, data mining dapat disebut sebagai proses untuk menemukan korelasi atau pola dari ratusan atau ribuan bidang dari database relasional yang besar (Aldino *et al.*, 2021). *Knowledge Discovery in Database (KDD) dapat dilihat pada Gambar 1.*



Gambar 1 knowledge discovery in database

2.2. Metode Elbow

Metode Elbow ini digunakan untuk menghasilkan informasi dalam menentukan jumlah kluster optimal dengan cara melihat persentase hasil perbandingan antara jumlah kluster yang akan membentuk siku pada suatu titik. dengan nilai k yang telah ditentukan dengan menggunakan metode elbow yang menunjukkan jumlah kluster terbaik (Aditya, Sari and Padilah, 2021). Metode Elbow digunakan untuk memilih jumlah kluster yang terbentuk pada suatu titik di grafik Sum of Square Error (SSE) dan didasarkan pada penurunan SSE yang besar. Jika nilai kluster sebelumnya (k-1) dengan nilai kluster selanjutnya (k) mengalami penurunan terbesar maka jumlah kluster tersebut yang tepat (k) (Syakur *et al.*, 2018). Rumus SSE dapat dituliskan sebagai :

$$SSE = \sum_{K=1}^K \sum_{i=1}^n |x_i - c_k|^2 \quad (1)$$

2.3. Algoritma K-Medoids

Algoritma K-Medoids hadir sebagai varian dari algoritma K-Means Clustering. Ditemukannya kelemahan pada algoritma K-Means yang sensitif terhadap outlier dikarenakan sebuah objek dengan sebuah nilai yang besar mungkin secara substansial menyimpang dari distribusi data. Maka Kmedoids muncul sebagai penanggulangannya, algoritma K-Medoids digunakan dalam menemukan Medoids sebuah cluster. Medoids merupakan objek data dari cluster yang terletak paling terpusat. Pemilihan Medoids dilakukan secara acak dari objek data K untuk membentuk cluster K dan objek data lainnya yang tersisa ditempatkan di dekat Medoids dalam sebuah cluster (Chrisnanto and Abdillah, 2015).

Langkah-langkah untuk algoritma K-Medoids sebagai berikut:

- Secara acak pilih k objek pada sekumpulan data objek sebagai medoid.
- Setiap objek yang bukan medoid dimasukan ke dalam cluster berdasarkan jarak terdekat.
- Perbarui medoid dengan memilih secara acak objek non medoid.
- Hitung total cost S, dari pertukaran medoid.
- Jika $S < 0$ maka tukar objek secara acak untuk membentuk sekumpulan k objek baru sebagai medoid.

- f. Ulangi langkah ke 3 sampai 5 hingga tidak ada perubahan.

Secara umum untuk mendefinisikan jarak antara dua objek x dan y , digunakan rumus jarak Euclidean pada persamaan (2):

$$d_{euclidean}(x, y) = \sqrt{\sum_{k=1}^n (x_i - y_i)^2} \quad (2)$$

Berdasarkan algoritma di atas, setiap objek dihitung jaraknya dengan medoids menggunakan perhitungan jarak Euclidean Distance pada Persamaan 1, jarak terkecil dari objek dalam data dikelompokkan dalam cluster tertentu sesuai dengan medoids-nya. X_i menggambarkan data pada medoids sedangkan y_i menggambarkan data pada obyek (Wira, Budianto and Wiguna, 2019).

2.4. Pengujian Silhouette Coefficient

Silhouette coefficient adalah suatu metode evaluasi cluster untuk melihat kualitas dan kekuatan dari cluster. Fungsi dari Silhouette ini menunjukkan hasil ukuran cluster memiliki rentang dari -1 sampai 1. Untuk nilai 1 menandakan bahwa nilai cluster yang terbentuk merupakan cluster yang baik, sedangkan untuk nilai 0 menunjukkan bahwa hasil cluster yang terbentuk adalah buruk. Tahapan perhitungan silhouette coefficient sebagai berikut : Silhouette Coefficient menjelaskan tahapan untuk melihat penempatan suatu obyek ke dalam suatu cluster.

- Untuk setiap objek i , hitung rata-rata jarak dari obyek ke i dengan seluruh obyek yang berada pada satu cluster dan nilai rata-rata tersebut bernama a_i
- Untuk setiap objek i , hitung rata-rata jarak dari objek ke i dengan objek yang berada di cluster lainnya dan diambil nilai rata-rata terkecil, yang disebut b_i
- Hitung nilai koefisien Silhouette dengan Persamaan (3).

$$S_i = (b_i - a_i) / \max(a_i, b_i) \quad (3)$$

Tabel 1 Silhouette Coefficient

Nilai Silhouette Coefficiente	Interpretasi oleh Kauffman
0,71 – 1,00	Strong Classification
0,51 – 0,70	Good Classification
0,26 – 0,50	Weak Classification
0 – 0,25	Bad Classification

Nilai 1 menunjukkan bahwa nilai cluster yang terbentuk merupakan cluster yang baik, sedangkan untuk nilai -1 menunjukkan bahwa hasil cluster yang terbentuk tidak baik atau buruk (Prakasawati, Chrisnanto and Hadiana, 2019).

3. HASIL DAN PEMBAHASAN

3.1. Data yang digunakan

Data yang digunakan dalam penelitian ini bersifat kuantitatif yaitu dikumpulkan dari hasil kuisioner mahasiswa Jurusan Informatika UNJANI yang berisi tentang karakteristik pola belajar mahasiswa mulai dari angkatan 2015 sampai 2019 Selain itu juga ada data akademik mahasiswa Informatika Unjani yang diambil dari tata usaha jurusan informatika UNJANI. Data yang akan digunakan dalam penelitian ini masing masing berjumlah 196 record.

Tabel 2 Data karakteristik pola belajar

Angkatan	Jurusan	V1	V2	...	A1	...	K10
----------	---------	----	----	-----	----	-----	-----

2017	INFORMATIKA	5	2	...	4	...	5
2017	INFORMATIKA	5	5	...	3	...	3
...
2018	INFORMATIKA	4	4	...	5	...	2
2018	INFORMATIKA	4	4	...	5	...	4

Adapun contoh data nilai akademik yang telah dilakukan seleksi untuk pemilihan variable yang akan digunakan dalam proses clustering. Sebelum selection dilakukan ada 23 variabel data nilai akademik yang digunakan sebagai berikut :

Tabel 3 Data nilai akademik

Nim	IF1141	IF1142	IF2125	IF3226	IF3227
3411151006	3	4	3	3,5	3
3411151008	3,5	3,5	3,5	4	4
.....
3411191139	3	3	3	3	2,5
3411191150	3,5	2,5	3	4	3

3.2. Transformasi Data

Data hasil cleaning yang ditunjukkan pada tabel 3.9 dan 3.10 dilakukan transformasi dengan normalisasi atribut numerik menjadi skala yang lebih kecil, yaitu dengan batas terendah 0 dan batas tertinggi 1. transformasi data kuisine karakteristik pola belajar dan nilai akademik ditunjukkan pada Tabel 3.11 adapun persamaan yang digunakan untuk melakukan normalisasi hasil kuisisioner yaitu dengan min max normalisasi pada persamaan (4) sedangkan nilai akademik dengan persamaan (3) yaitu nilai rata rata :

$$v_i = \frac{v_i - \min_A}{\max_A - \min_A} (new_max_A - new_min_A) + new_min_A \tag{4}$$

$$\bar{x} = \frac{x_1 + x_2 + x_3}{\text{dibagi dengan banyaknya } n} \tag{5}$$

Tabel 4 Normalisasi data kuisisioner

Angkatan	Jurusan	Visual	Auditorial	Kinestik	Nilai Akademik
2017	INFORMATIKA	0.35	0.4	0.34	3.789474
2017	INFORMATIKA	0.8	0.32	0.57	3.722222
.....
2017	INFORMATIKA	0.3	0.16	0.65	3.6875
2017	INFORMATIKA	0.3	0.08	0.57	3.666667

3.3. Penerapan Algoritma K-Medoids

Tahapan algoritma K-Medoids ini merupakan tahapan melakukan perhitungan untuk klasterisasi pola belajar berdasarkan data hasil kuisisioner yang telah melalui proses perubahan menjadi data set. Langkah-langkah untuk algoritma KMedoid sebagai berikut:

- a. Pertama secara acak pilih K objek, Jumlah K yang digunakan pada perhitungan K-medoids ini yaitu K=2. Pada Tabe 5 menunjukkan pengambilan objek secara acak t sebagai medoids awal.

Tabel 5 Normalisasi nilai akadmeik

No Medoids	Visual	Auditorial	Kinestk
13	0.840125	0.63843	0.680162
14	0.827586	0.61553	0.722448

- b. Setiap data (objek) yang bukan medoid dimasukan ke dalam cluster berdasarkan jarakterdekat dengan menggunakan persamaan ukuran jarak euclidian distance.

Tabel 6 iterasi 1

Nim	K1	K2
3411171069	0.049696	0.152525
3411171137	0.170578	0.166034
.....
3411171172	0.250364	0.221582
3411171104	0.261961	0.242959

Pada Tabel 7 menunjukan hasil jarak terdekat dari setiap data pada iterasi ke 1 yang diambil dari jarak terkecil. Sehingga dapat terbentuk pengelompokan data pada setiap cluster.

Tabel 7 Hasil jarak iterasi

Nim	Jarak Kedekatan	Cluster
3411171069	0.049696	1
3411171137	0.166034	2
.....
3411171172	0.221582	2
3411171104	0.242959	2
Total Cost Iterasi 1		38.11

- c. Pada tabel menunjukan medoids yang telah di perbaharui.

Tabel 8 Medoids baru

No Medoids	Visual	Auditorial	Kinestk
144	0.536183	0.226952	0.982685
100	0.43772	0.047077	0.853122

3.4. Perhitungan Metode Elbow

Perhitungan Elbow dalam menentukan klaster ideal, pada rumus persamaan (1) Adapun hasil dari perhitungan total dan selisih pada tabel dibawah ini :

Tabel 9 Hasil elbow

Nilai K	SSE	Selisih
2	38.11	38.11
3	19.86	18.25
4	19.75	1.5

Jadi dari hasil perhitungan elbow mendapatkan nilai K = 2 yang paling optimal karna dari selisih setiap nilai K didapatkan nilai selisih yang sangat signifikan.

3.5. Interpretasi dan Hasil Pengujian

Interpretasi dari hasil pengelompokan pola belajar dengan algoritma K-Medoids maka terbentuk 2 kelompok klaster nilai rata rata akademik dan pola belajar dari jumlah data sebanyak 196 dapat dilihat pada tabel 10

Tabel 10 Hasil pengelompokan metode K-Medoids

Atribut	Hasil Cluster 1	Hasil Cluster 2
Nilai Rata Rata IPK	2.64	3.23
Visual	0.6017	0.4280
Auditorial	0.4859	0.4162
Kinestik	0.4595	0.4795

Berdasarkan tabel diatas, maka dapat diinterpretasikan karakteristik pola belajar mahasiswa khususnya pada karakteristik pola belajar visual, auditorial dan kinestik. menyatakan kumpulan nilai klaster karakteristik pola belajar mahasiswa yaitu di dapatkan nilai tertinggi pada klaster 1 dengan jumlah data mahasiswa 119 pola belajar visual lebih tinggi yaitu (0.60) dibandingkan dengan karakteristik pola belajar auditorial dan kinestik dengan nilai rata rata IPK 2.64. Sedangkan Karakteristik Pola belajar mahasiswa pada klaster 2 dengan jumlah data mahasiswa 76 dan pola belajar kinestik lebih tinggi yaitu (0.48) dibandingkan dengan karakteristik pola belajar visual dan auditorial dengan nilai rata rata IPK 3.23. Adapun hasil pengujian yang dilakukan dengan beberapa nilai K didapatkan hasil nilai yang paling tinggi dan mendekati $Si = 1$ yaitu pada nilai $K = 2$ didapatkan hasil 0.80 dimana nilai Silhouette Coefficient semakin mendekati nilai 1, maka semakin bagus.

4. KESIMPULAN

Pada penelitian ini dapat disimpulkan bahwa karakteristik pola belajar dapat dikelompokkan berdasarkan klaster nya masing – masing. Adapun hasil dari karakteristik pola belajar mahasiswa dengan nilai rata rata IPK 2.64 pada klaster 1 lebih condong pola belajar visual yaitu pola belajar dengan cara melihat dan mengamati . Sedangkan pada klaster 2 dengan nilai rata rata IPK 3.23 polajarnya lebih condong ke kinestik yaitu belajar dengan cara praktik secara langsung.

Adapun hasil dari penggunaan teknik *Clustering* dengan algoritma K-Medoids, menunjukkan hasil teknik *Clustering* dengan *algoritma K-Medoids* mampu melakukan pengelompokan data terhadap data karakteristik pola belajar. Hasil pengujian *cluster* didapatkan jumlah $K=2$ yang di dapat dari elbow menunjukkan nilai $K2$ yan optimal adalah $K2$. Tetapi *silhoutte Coefficient* yang paling tinggi mendekati $Si = 1$ yaitu pada nilai $K2$ yaitu 0.80 yang merupakan struktur kuat *Kaufman* dan *Rousseuw*.

DAFTAR PUSTAKA

- Annizar, A. M. and Arifin, M. (2021) ‘Perbedaan Prestasi Belajar Mahasiswa Ditinjau dari Jalur Seleksi Masuk Perguruan Tinggi’, *SAP (Susunan Artikel Pendidikan)*, 5(3), pp. 197–204. doi: 10.30998/sap.v5i3.8411.
- Arora, P., Deepali and Varshney, S. (2016) ‘Analysis of K-Means and K-Medoids Algorithm for Big Data’, *Physics Procedia*, 78(December 2015), pp. 507–512. doi: 10.1016/j.procs.2016.02.095.
- Chrisnanto, Y. H. and Abdillah, G. (2015) ‘Gambaran Umum Kemampuan Akademik Mahasiswa Unjani Dengan Algoritma Partitioning Around Medoids (PAM) Clustering’, *Seminar Nasional Ilmu Pengetahuan dan Teknologi*, pp. 285–290.
- Harikumar, S. and Surya, P. V. (2015) ‘K-Medoid Clustering for Heterogeneous DataSets’, *Procedia Computer Science*, 70, pp. 226–237. doi: 10.1016/j.procs.2015.10.077.
- Murti, R. W. and Prasetio, A. P. (2018) ‘Pengaruh Kompetensi Dosen terhadap Prestasi Akademik Mahasiswa Fakultas Ekonomi dan Bisnis Telkom University’, *Jurnal Penelitian Pendidikan*, 18(2), pp. 94–102. doi: 10.17509/jpp.v18i2.12950.
- Prakasawati, P. E., Chrisnanto, Y. H. and Hadiana, A. I. (2019) ‘Segmentasi Pelanggan Berdasarkan Produk Menggunakan Metode K- Medoids’, *KOMIK (Konferensi Nasional Teknologi Informasi dan Komputer)*, 3(1), pp. 335–339. doi: 10.30865/komik.v3i1.1610.
- Rahman, F. *et al.* (2020) ‘Application of Data Mining Technique using K-Medoids in the case of Export of Crude Petroleum Materials to the Destination Country’, *IOP Conference Series: Materials Science and Engineering*, 835(1). doi: 10.1088/1757-899X/835/1/012058.

- Sibuea, M. L. and Safta, A. (2017) 'Pemetaan Siswa Berprestasi Menggunakan Metode K-Means Clustering', *Jurteks*, 4(1), pp. 85–92. doi: 10.33330/jurteks.v4i1.28.
- Syakur, M. A. *et al.* (2018) 'Integration K-Means Clustering Method and Elbow Method for Identification of the Best Customer Profile Cluster', *IOP Conference Series: Materials Science and Engineering*, 336(1). doi: 10.1088/1757-899X/336/1/012017.
- Tanamir, M. D. and Dkk (2020) 'Analisis Karakteristik Gaya Belajar Remaja Di Korong Sijangek Kenagarian Sungai Durian Kecamatan Patamuan Kabupaten Padang Pariaman', *Curricula: Journal of Teaching and Learning*, 5(1), pp. 50–59. doi: 10.22216/jcc.2020.v5i1.5035.
- Vermunt, J. D. and Donche, V. (2017) 'A Learning Patterns Perspective on Student Learning in Higher Education: State of the Art and Moving Forward', *Educational Psychology Review*, 29(2), pp. 269–299. doi: 10.1007/s10648-017-9414-6.
- Wira, B., Budiarto, A. E. and Wiguna, A. S. (2019) 'Implementasi Metode K-Medoids Clustering Untuk Mengetahui Pola Pemilihan Program Studi Mahasiswa Baru Tahun 2018 Di Universitas Kanjuruhan Malang', *RAINSTEK: Jurnal Terapan Sains & Teknologi*, 1(3), pp. 53–68. doi: 10.21067/jtst.v1i3.3046.