



Komparasi Algoritme X-Means dan K-Medoids pada Klasterisasi Data Akademik Mahasiswa

Cepi Ramdani^{1*}
¹cepi@itttelkom-pwt.ac.id

¹Institut Teknologi Telkom Purwokerto, Fakultas Informatika, Program Studi Sistem Informasi, Jl. DI. Panjaitan No. 128 Purwokerto, Jawa Tengah, Indonesia

Informasi Artikel

Diterima: 06-08-2023
Direvisi: 14-08-2023
Diterbitkan: 23-08-2023

Kata Kunci

Klasterisasi, K-Means, K-Medoids, Davies-Bouldin Index, Data Akademik

*** Author Korespondensi**

Abstrak

Data akademik mahasiswa dapat diolah menjadi Informasi yang sangat penting dalam pengambilan kebijakan maupun keputusan bagi pengelola program studi (Prodi) dalam upaya peningkatan kualitas pelayanan, keberhasilan pembelajaran, dan perbaikan kurikulum. Salah satu cara pengolahan data adalah klasterisasi. Berbagai algoritme klasterisasi telah banyak dikembangkan diantaranya Algoritme *X-Means* dan *K-Medoids*. Penelitian ini bertujuan untuk mendapatkan klasterisasi optimal kinerja akademik mahasiswa berdasarkan data akademik dengan membandingkan hasil klasterisasi algoritme *X-Means* dan *K-Medoids*. Dataset didapatkan dari Sistem Informasi Akademik Kampus XYZ Program Studi (Prodi) S1 Sistem Informasi, diperoleh data sejumlah 483 dari angkatan 2017 hingga 2020 yang terdiri dari 11 atribut. Komparasi hasil klasterisasi kedua algoritme dilakukan dengan menghitung nilai *Davies-Bouldin Index* (DBI). Berdasarkan hasil perhitungan DBI didapati bahwa Klasterisasi Algoritme *K-Medoids* lebih baik daripada *X-Means* dengan nilai DBI *K-Medoids* 0.061 dan *X-Means* 0.091. Hasil klasterisasi *K-Medoids* menunjukkan bahwa 97% mahasiswa memiliki kinerja akademik baik dan sisanya belum baik sebanyak 3%.

Abstract

Student academic data can be processed into information that is very important in making policies and decisions for study program (Prodi) managers in an effort to improve service quality, learning success, and curriculum improvement. One way of processing data is clustering. Various clustering algorithms have been developed including X-Means and K-Medoids algorithms. This study aims to obtain optimal clustering of student academic performance based on academic data by comparing the clustering results of X-Means and K-Medoids algorithms. The dataset is obtained from the Academic Information System of XYZ Campus, S1 Information Systems Study Program (Prodi), obtained data totaling 483 from batch 2017 to 2020 consisting of 11 attributes. Comparison of clustering results of both algorithms is done by calculating the Davies-Bouldin Index (DBI) value. Based on the DBI calculation results, it is found that the K-Medoids Clustering Algorithm is better than X-Means with a DBI value of K-Medoids 0.061 and X-Means 0.091. K-Medoids clustering results show that 97% of students have good academic performance and the remaining 3% are not good.

1. Pendahuluan

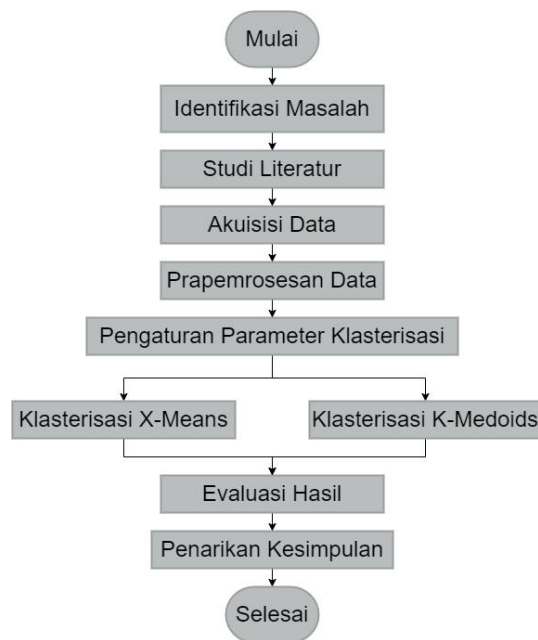
Data akademik mahasiswa berpotensi menghasilkan Informasi yang bernilai dan bermanfaat bagi Pengelola Program Studi (Prodi). Pengolahan data akademik mahasiswa dapat menghasilkan Informasi penting yang dapat menjadi salah satu aspek pertimbangan penyusunan strategi dan pengambilan keputusan dalam upaya peningkatan kualitas pelayanan, keberhasilan pembelajaran, dan perbaikan kurikulum (Aziz et al., 2018). Salah satu cara pengolahan data adalah klusterisasi. Pengelompokan data kedalam kluster-kluster berdasarkan tingkat kemiripannya. Semakin dekat tingkat kemiripannya maka data-data tersebut akan dikelompokkan kedalam kluster yang sama, sebaliknya semakin jauh tingkat kemiripannya maka data-data tersebut akan berada pada kluster yang berbeda (Priyatman et al., 2019). Pengelompokan kluster dilakukan pada data yang tidak memiliki label atau kategori. Berbagai algoritme klusterisasi telah banyak dikembangkan diantaranya Algoritme *X-Means* dan *K-Medoids*. Kedua algoritme berbasis klusterisasi partisi, data dibagi kedalam beberapa kluster, selama proses optimasi kluster berlangsung, data akan berpindah dari satu kluster ke kluster lain menyesuaikan dengan kedekatan jarak pada pusat kluster (Musfiani, 2019) (Nirmal, 2019). Algoritme *X-Means* merupakan pengembangan dari Algoritme *K-Means* yang populer digunakan karena tingkat efisiensinya. Algoritme *X-Means* mengatasi kelemahan dalam penentuan jumlah kluster optimal. Selama ini jumlah kluster ditentukan secara acak, hal tersebut mempengaruhi kualitas hasil klusterisasi (Mughnyanti et al., 2020). Algoritme *X-Means* melakukan perhitungan dalam penentuan jumlah kluster, sehingga dihasilkan jumlah kluster yang paling optimal, hal ini menjadikan penentuan jumlah kluster lebih terarah, cerdas, dan meminimalisir penentuan secara acak. Algoritme *K-Medoids* mirip dengan *K-Means*, perbedaannya terletak pada data yang menjadi pusat kluster, pada *K-Means* data pusat kluster yang digunakan adalah rata-rata dari seluruh titik data yang ada dalam kluster, sementara data pusat kluster *K-Medoids* adalah salah satu titik data yang merupakan anggota kluster (*Medoids*). *K-Medoids* lebih tahan terhadap data yang memiliki pencilan (*Outlier*) (Kamila et al., 2019).

Studi terkait algoritme *X-Means* dan *K-Medoids* telah banyak dilakukan, namun komparasi klusterisasi kedua algoritme tersebut masih jarang dilakukan, berikut hasil studi literatur mengenai penelitian komparasi berbagai algoritme klusterisasi. Penelitian (Kaligis & Yulianto, 2022), melakukan klusterisasi dengan menggunakan 3 algoritme klusterisasi yakni *K-Means*, *K-Medoids*, dan *X-Means* pada data kinerja pegawai di Sekretariat DPRD Provinsi Sulawesi Utara yang terdiri dari 5 atribut data yakni orientasi pelayanan, integritas, komitmen, disiplin, dan kerjasama. Tujuan dari penelitian tersebut untuk mendapatkan algoritme optimal dalam pengelompokan data kinerja pegawai dengan melakukan komparasi terhadap hasil klusterisasi ketiga algoritme yang digunakan. Hasil komparasi menemukan bahwa algoritme *K-Means* lebih optimal dari kedua algoritme lainnya untuk klusterisasi data kinerja pegawai, hal ini berdasarkan hasil perhitungan nilai DBI. Algoritme *K-Means* memiliki nilai DBI paling kecil. Secara berturut-turut nilai DBI dari algoritme *K-Means*, *K-Medoids*, dan *X-Means* yakni -0.377, -0.930 dan -0.497. Penelitian (Utomo, 2021), melakukan perbandingan algoritme klusterisasi *K-Means* dan *K-Medoids* dengan menggunakan data yang bersumber dari Kementerian Kesehatan Indonesia mengenai kasus *Covid-19* yang terdiri dari 4 atribut data yakni status terkonfirmasi, status pengobatan, status pemulihan, dan status kematian. Klusterisasi dilakukan dengan pengaturan jumlah kluster pada rentang 2 – 9. *Peak point K-Means* berada pada jumlah kluster 5 dengan nilai DBI 0.064, sedangkan *K-Medoids* pada jumlah kluster 2 dengan nilai DBI 0.411. Kesimpulan yang diambil, *K-Means* lebih baik dari *K-Medoids* untuk klusterisasi data Kasus *Covid-19*. Penelitian (Sholikhah, 2022), dilatarbelakangi kurangnya kesadaran masyarakat Kabupaten Bojonegoro dalam melengkapi dokumen kependudukan maka diperlukan pengelompokan Kecamatan di Kabupaten Bojonegoro. Data yang digunakan berasal dari Dinas Kependudukan dan Catatan Sipil Kabupaten Bojonegoro tahun 2020 yang terdiri dari 3 atribut data, yakni kelengkapan Kartu Tanda Penduduk (KTP), Kartu Keluarga (KK), dan Akta Kelahiran (AK). Tujuan dari penelitian ini membandingkan metode-metode klusterisasi dalam rangka mendapatkan metode terbaik dalam pengelompokan Kecamatan di Kabupaten Bojonegoro. Algoritme klusterisasi yang digunakan yaitu *K-Means*, *K-Medoids*, *X-Means*, dan *DBSCAN*. Kesimpulan penelitian didapati algoritme paling optimal untuk klusterisasi kecamatan yakni *K-Means* dengan nilai DBI -0,697. Penelitian (Aryuni et al., 2018), melakukan segmentasi konsumen *Internet Banking* berdasarkan nilai FRM (*Recency, Frequency and Monetary*), konsumen yang tersegmentasi memiliki kriteria yang mayoritas sama,

informasi tersebut digunakan untuk mengetahui tingkat loyalitas konsumen dan mendukung pada Manajemen Hubungan Pelanggan (*Customer Relationship Management/CRM*) Bank. Segmentasi dilakukan dengan menerapkan algoritme *K-Means* dan *K-Medoids*. Dataset penelitian menggunakan data perilaku konsumen *Internet Banking* dengan atribut *Recency*, *Frequency* dan *Monetary*. Hasil segmentasi dibandingkan untuk mengetahui hasil yang paling optimal dengan melakukan evaluasi menggunakan metode perhitungan jarak intra kluster (AWC) dan DBI. Kesimpulan penelitian, algoritme *K-Means* lebih baik dari *K-Medoids*. Penelitian(Herviany et al., 2021), Pemetaan resiko bencana alam, khususnya bencana tanah longsor sangat penting dalam penataan penanggulangan bencana yang terarah dan tepat di Jawa Barat. Karena itu diperlukan pengelompokan daerah kabupaten/kota yang sering terjadi bencana alam. Klasterisasi dilakukan dengan menggunakan algoritme *K-Means* dan *K-Medoids*. Dataset yang digunakan adalah data tanah longsor Provinsi Jawa Barat pada tahun 2019 dengan 2 atribut data yakni atribut nama daerah dan jumlah kejadian bencana. Kesimpulan penelitian menunjukkan bahwa *K-Means* lebih baik dari *K-Medoids* berdasarkan hasil evaluasi DBI dengan jumlah kluster optimal sebanyak 6 kluster.

Berdasarkan latarbelakang permasalahan dan peninjauan penelitian terdahulu, maka pada penelitian ini dilakukan komparasi klasterisasi data akademik mahasiswa untuk mengetahui sebaran kinerja akademik mahasiswa Prodi S1 Sistem Informasi menggunakan algoritme *X-Means* dan *K-Medoids*. Hasil klasterisasi kedua algoritme dibandingkan untuk mendapatkan hasil klasterisasi yang optimal dengan menggunakan metode evaluasi perhitungan DBI. Data akademik mahasiswa didapatkan dari Sistem Informasi Akademik Kampus XYZ Prodi S1 Sistem Informasi sejumlah 483 data dari Angkatan 2017 hingga 2020, yang terdiri dari 11 atribut yakni (1) Jumlah Satuan Kredit Semester (SKS) lulus semester 1, (2) Jumlah Satuan Kredit Semester (SKS) lulus semester 2, (3) Jumlah Satuan Kredit Semester (SKS) lulus semester 3, (4) Jumlah Satuan Kredit Semester (SKS) lulus semester 4, (5) Jumlah Satuan Kredit Semester (SKS) lulus semester 5, (6) Indeks Prestasi (IP) semester 1, (7) Indeks Prestasi (IP) semester 2, (8) Indeks Prestasi (IP) semester 3, (9) Indeks Prestasi (IP) semester 4, (10) Indeks Prestasi (IP) semester 5, dan (11) Sisa SKS yang harus diambil mahasiswa untuk lulus (144 SKS – Total SKS Lulus). Proses klasterisasi dilakukan dengan pengaturan parameter-parameter yang sama untuk kedua algoritme seperti jumlah kluster, jumlah iterasi, dan algoritme perhitungan jarak. Aplikasi *Rapidminer 10.1* digunakan sebagai media bantu dalam penelitian ini.

2. Metode Penelitian



Gambar 1. Alur Penelitian

Alur penelitian ini disajikan pada Gambar 1, dengan penjelasan setiap tahapannya sebagai berikut,

2.1 Identifikasi Masalah

Tahapan ini merupakan tahapan awal penelitian, pencarian permasalahan penelitian dengan cara observasi dilingkungan Prodi, didapati permasalahan terkait data akademik mahasiswa yang belum diberdayakan secara maksimal untuk peningkatan berkelanjutan Prodi. Disamping itu dilakukan studi terhadap pustaka-pustaka ilmiah, ditemukan bahwa penelitian terkait komparasi algoritme klasterisasi *X-Means* dengan *K-Medoids* masih jarang dilakukan terutama pada data akademik.

2.2 Studi Literatur

Tahapan berikutnya setelah permasalahan ditemukan adalah pencarian wawasan, pengetahuan, dan solusi terkait klasterisasi dengan menelaah pustaka-pustaka ilmiah diantaranya Jurnal, Buku, Video Edukasi.

2.3 Akuisisi Data

Data penelitian untuk keperluan klasterisasi diakuisisi dari Sistem Informasi Akademik Kampus XYZ dibatasi hanya data akademik mahasiswa Prodi S1 Sistem Informasi. Data yang diakuisisi sejumlah 483 data, merupakan gabungan dari data akademik mahasiswa angkatan 2017 hingga 2020 dari rentang semester 1 hingga semester 5, terdiri dari 11 atribut data yaitu (1) Jumlah Satuan Kredit Semester (SKS) lulus semester 1, (2) Jumlah Satuan Kredit Semester (SKS) lulus semester 2, (3) Jumlah Satuan Kredit Semester (SKS) lulus semester 3, (4) Jumlah Satuan Kredit Semester (SKS) lulus semester 4, (5) Jumlah Satuan Kredit Semester (SKS) lulus semester 5, (6) Indeks Prestasi (IP) semester 1, (7) Indeks Prestasi (IP) semester 2, (8) Indeks Prestasi (IP) semester 3, (9) Indeks Prestasi (IP) semester 4, (10) Indeks Prestasi (IP) semester 5, dan (11) Sisa SKS yang harus diambil mahasiswa untuk lulus (144 SKS – Total SKS Lulus).

2.4 Prapemrosesan Data

Data yang telah dikumpulkan dilakukan prapemrosesan data, yakni menghapus item data yang memiliki atribut yang bernilai kosong, total data setelah pemrosesan ini adalah 483 data. Tahap berikutnya dilakukan normalisasi data untuk menghindari pengaruh dominan dari suatu atribut (Syahputra et al., 2022). Menerapkan metode normalisasi *Min-Max*, dengan batasan rentang data minimal 0 dan maksimal 1, rumus *Min-Max* terdapat pada rumus (1). Tabel 1 merupakan contoh data sebelum normalisasi dan Tabel 2 adalah contoh data setelah normalisasi.

$$d' = \frac{d - \min(e)}{\max(e) - \min(e)} \tag{1}$$

Keterangan :

d' = data hasil normalisasi

d = data sebelum normalisasi

$\min(e)$ = data minimum pada kolom/atribut sebelum normalisasi

$\max(e)$ = data maximum pada kolom/atribut sebelum normalisasi

Tabel 1. Data Sebelum Normalisasi

Mahasiswa	SKS 1 Lulus	SKS 2 Lulus	SKS 3 Lulus	SKS 4 Lulus	SKS 5 Lulus	SKS Kurang	IP 1	IP 2	IP 3	IP 4	IP 5
Mahasiswa 1	17	20	17	15	14	61	2.45	2.4	2.82	2.25	2.50
Mahasiswa 2	20	20	24	23	23	34	3.7	3.75	3.67	3.72	3.74
Mahasiswa 3	20	20	24	23	23	34	3.5	3.5	3.44	3.00	3.26
Mahasiswa 4	20	20	24	18	20	42	3.05	3	3.38	2.46	3.53
...
...

...
Mahasiswa 481	17	20	23	22	21	41	2.65	3.4	2.89	3.09	3.45
Mahasiswa 482	20	20	23	23	23	35	3.9	3.9	3.89	3.57	3.91
Mahasiswa 483	13	7	2	17	12	93	2	0.9	0.58	2.85	1.74

Tabel 1. merupakan dataset penelitian sebelum dilakukan normalisasi. Data sebanyak 483 data yang terdiri dari 11 atribut dengan rentang data dari 0 hingga 111.

Tabel 2. Data Setelah Normalisasi

Mahasiswa	SKS 1 Lulus	SKS 2 Lulus	SKS 3 Lulus	SKS 4 Lulus	SKS 5 Lulus	SKS Kurang	IP 1	IP 2	IP 3	IP 4	IP 5
Mahasiswa 1	0.57	1.00	0.71	0.50	0.58	0.36	0.25	0.51	0.69	0.49	0.63
Mahasiswa 2	1.00	1.00	1.00	0.94	0.96	0.01	0.86	0.92	0.91	0.92	0.93
Mahasiswa 3	1.00	1.00	1.00	0.94	0.96	0.01	0.76	0.85	0.85	0.71	0.82
Mahasiswa 4	1.00	1.00	1.00	0.67	0.83	0.12	0.54	0.69	0.83	0.55	0.88
...
...
...
Mahasiswa 481	0.57	1.00	0.96	0.89	0.88	0.10	0.35	0.82	0.70	0.73	0.86
Mahasiswa 482	1.00	1.00	0.96	0.94	0.96	0.03	0.95	0.97	0.97	0.87	0.98
Mahasiswa 483	0.00	0.19	0.08	0.61	0.50	0.77	0.04	0.05	0.09	0.66	0.43

Tabel 2. merupakan dataset penelitian setelah dilakukan normalisasi. Rentang data dari 0 hingga 1.

2.5 Pengaturan Parameter Klasterisasi

Tahap berikutnya setelah data dinormalisasi adalah melakukan pengaturan terhadap parameter-parameter klasterisasi. Pengaturan tersebut yakni, Jumlah klaster pada rentang 2 - 4, Jumlah maksimal proses klasterisasi 100, Jumlah maksimal iterasi setiap proses klasterisasi 500, Jenis pengukuran jarak *Numerical Measures* dengan metode *Euclidean Distance*. Rumus(2) merupakan rumus *Euclidean Distance*(Syahputra et al., 2022).

$$d(A,B) = \sqrt{(b_1 - a_1)^2 + (b_2 - a_2)^2 + \dots + (b_n - a_n)^2} \quad (2)$$

Keterangan :

d(A,B) = Jarak data A ke data B
 b_1 = nilai atribut pertama data B
 a_1 = nilai atribut pertama data A
 b_n = nilai atribut ke n data B
 a_n = nilai atribut ke n data A

2.6 Klasterisasi

Pada tahap ini dilakukan klasterisasi data akademik mahasiswa menerapkan algoritme *X-Means* dan *K-Medoids* dengan pengaturan parameter-parameter klasterisasi sesuai tahap sebelumnya. Proses klasterisasi

memanfaatkan aplikasi *Rapidminer versi 10.1*. Aplikasi ini merupakan salah satu aplikasi penambangan data (*data mining*) yang bersifat *open source* berlisensi AGPL (*GNU Affero General Public License*) (Fimawahib & Rouza, 2021).

2.6.1 Klasterisasi *X-Means*

Algoritme *X-Means* merupakan pengembangan dari algoritme *K-Means* yang mengatasi kelemahan dalam penentuan jumlah kluster. *X-Means* secara otomatis menentukan jumlah kluster yang paling optimal. Tahapan dari Algoritme *X-Means* sebagai berikut (Hakim et al., 2022) (Ge et al., 2021),

1. Tetapkan rentang jumlah kluster minimal dan maksimal [k_{min} , k_{max}]. Pada penelitian ini jumlah minimal kluster diatur sebanyak 2 kluster dan jumlah maksimal kluster sebanyak 9 kluster;
2. Inisialisasi pusat kluster k_{min} , dan jalankan algoritme *K-Means* untuk $k = k_{min}$ dalam ruang sampel untuk mendapatkan pusat-pusat kluster k_{min} ;
3. Membagi setiap pusat kluster yang diperoleh pada langkah sebelumnya menjadi dua pusat kluster, dan jalankan algoritme *K-Means* untuk $k=2$ pada ruang sampel masing-masing kluster untuk mendapatkan model baru;
4. Hitung nilai *Bayesian Information Criterion* (BIC) dari model baru dan model asli. Jika nilai BIC dari model baru lebih besar dari model asli, hapus pusat kluster model asli dan tambahkan pusat-pusat kluster setelah pemisahan. Jika tidak, pertahankan pusat-pusat kluster model asli;
5. Tentukan apakah jumlah pusat kluster saat ini kurang dari k_{max} . Jika kurang dari k_{max} , tentukan apakah jumlah pusat kluster saat ini sama dengan jumlah pusat kluster pada iterasi sebelumnya. Jika benar, maka selesai; jika tidak, kembali ke langkah 2. Jika jumlah pusat kluster lebih besar atau sama dengan k_{max} , maka selesai.

2.6.2 Klasterisasi *K-Medoids*

Algoritme *K-Medoids* hampir serupa dengan Algoritme *K-Means*, namun pusat kluster dalam *K-Medoids* diwakili oleh salah satu titik data yang merupakan anggota kluster (titik *medoids*) sementara pada *K-Means*, pusat kluster adalah rerata dari seluruh data dalam suatu kluster. Berikut adalah langkah-langkah dari algoritme *K-Medoids* (Hidayati et al., 2020) (Supriyadi et al., 2021).

1. Menentukan jumlah kluster yang diinginkan
2. Inisialisasi sejumlah kluster titik data sebagai pusat kluster awal (*medoids*) secara acak
3. Menghitung jarak titik data dengan *medoids* pada setiap kluster dan tempatkan setiap titik data tersebut ke *medoids* terdekat
4. Menghitung total jarak antara setiap titik data dalam kluster dengan *medoid*nya. Total jarak ini mengukur seberapa baik *medoids* mewakili titik-titik data dalam kluster tersebut.
5. Pada setiap kluster, pilih titik data yang memiliki total jarak terkecil dengan seluruh titik data lain dalam kluster sebagai *medoids* baru untuk kluster tersebut
6. Perbarui *medoid* dan ulangi langkah 3 hingga langkah 5 sampai *medoid* stabil atau tidak ada perubahan yang signifikan dalam *medoids*.

2.7 Evaluasi Hasil

Hasil klasterisasi kedua algoritme dievaluasi dengan melakukan perhitungan DBI. *Davies-Bouldin Index* (DBI) merupakan salah satu instrumen evaluasi yang digunakan untuk mengukur kualitas hasil klasterisasi. Semakin rendah nilai DBI, semakin baik kualitas hasil klasterisasinya. Tahapan perhitungan nilai DBI yakni (1) perhitungan *Sum Of Square Within Cluster*, menghitung rata-rata jarak setiap titik data dengan pusat klasternya, (2) perhitungan *Sum Of Square Between Cluster*, setiap pasangan kluster hitung jarak antara pusat klasternya, (3) hitung nilai DBI untuk setiap kluster, (4) hitung nilai DBI untuk seluruh kluster (Septiani et al., 2022). Evaluasi dengan metode DBI menerapkan aplikasi *Rapidminer* menggunakan operator *Cluster Distance Performance*, pengaturan kriteria utama *Davies Bouldin*.

2.8 Penarikan Kesimpulan

Kesimpulan penelitian dihasilkan dengan melakukan analisis hasil evaluasi terhadap hasil klusterisasi algoritme *X-Means* dan *K-Medoids* dengan menghitung nilai DBI.

3. Hasil dan Pembahasan

Hasil evaluasi DBI klusterisasi algoritme *X-Means* dan *K-Medoids* dengan pengaturan jumlah kluster 2, 3, dan 4 ditunjukkan pada Tabel 3 dibawah.

Tabel 3. Nilai DBI Setiap Algoritme

Jumlah Kluster	DBI X-Means	DBI K-Medoids
Kluster = 2	0.091	0.061
Kluster = 3	0.102	0.089
Kluster = 4	0.114	0.108

Berdasarkan data pada Tabel 3, untuk 3 skenario klusterisasi dengan jumlah kluster yang berbeda, nilai DBI *K-Medoids* semuanya lebih kecil dari nilai DBI *X-Means*. Nilai DBI terkecil dari *K-Medoids* yakni 0.061 dengan jumlah kluster sebanyak 2 dan nilai DBI terkecil dari *X-Means* yaitu 0.091 dengan jumlah kluster sebanyak 2 juga. Pusat kluster dari masing-masing algoritme untuk jumlah kluster sebanyak 2, dapat dilihat pada Tabel 4 untuk algoritme *X-Means* dan Tabel 5 untuk *K-Medoids*.

Tabel 4. Pusat Kluster Algoritme X-Means 2 Kluster

Atribut	Kluster 1	Kluster 2
SKS 1 Lulus	0.982831	0.738095
SKS 2 Lulus	0.994189	0.760417
SKS 3 Lulus	0.923658	0.69246
SKS 4 Lulus	0.908793	0.568783
SKS 5 Lulus	0.933579	0.65377
SKS Kurang	0.054887	0.360501
IP 1	0.719231	0.390419
IP 2	0.82142	0.480769
IP 3	0.881963	0.646026
IP 4	0.888247	0.57829
IP 5	0.916987	0.608989

Tabel 4. merupakan pusat kluster yang dihasilkan oleh Algoritme X-Means dengan jumlah kluster sebanyak 2. Pusat Kluster 1 memiliki nilai yang lebih besar dari Pusat Kluster 2 pada setiap atributnya.

Tabel 5. Pusat Kluster Algoritme K-Medoids 2 Kluster

Atribut	Kluster 1	Kluster 2
SKS 1 Lulus	1	0
SKS 2 Lulus	1	0.1875
SKS 3 Lulus	0.958333333	0.083333333
SKS 4 Lulus	0.944444444	0.611111111
SKS 5 Lulus	0.958333333	0.5
SKS Kurang	0.025641026	0.769230769

IP 1	0.951807229	0.036144578
IP 2	0.969230769	0.046153846
IP 3	0.971014493	0.088888889
IP 4	0.872122762	0.661764706
IP 5	0.97826087	0.434782609

Tabel 5. merupakan pusat kluster yang dihasilkan oleh Algoritme K-Medoids dengan jumlah kluster sebanyak 2. Sama halnya dengan X-Means, Pusat Kluster 1 memiliki nilai yang lebih besar dari Pusat Kluster 2 pada setiap atributnya.

Hasil klusterisasi algoritme *X-Means* dan *K-Medoids* terhadap data akademik mahasiswa untuk jumlah kluster sebanyak 2 dapat dilihat pada Tabel 6. Untuk menentukan kategori kelompok mahasiswa dari masing-masing kluster dapat dilakukan dengan melihat dari hasil titik pusat yang terbentuk (Syahputra et al., 2022). Kluster 1 diinterpretasikan sebagai kluster mahasiswa yang memiliki kinerja akademik baik, sedangkan kluster 2 adalah kelompok mahasiswa yang memiliki kinerja akademik kurang. Hal ini didasarkan pada nilai setiap atribut pada pusat kluster. Pusat Kluster 1 memiliki nilai atribut lebih besar dari pusat Kluster 2 baik pada *X-Means*, maupun *K-Medoids* kecuali untuk atribut “SKS Kurang”, karena semakin kecil nilai atribut ini maka semakin baik kinerja akademik mahasiswa.

Tabel 6. Hasil Klusterisasi *X-Means* dan *K-Medoids* 2 Kluster

Algoritme	Kluster 1	Kluster 2	Persentase Kluster 1	Persentase Kluster 2
X-Means	441	42	91%	9%
K-Medoids	470	13	97%	3%

Tabel 6, menunjukkan rekap pengelompokan mahasiswa Prodi S1 Sistem Informasi Kampus XYZ, yakni sebanyak 91% mahasiswa memiliki kinerja akademik baik (kluster 1) sisanya sebanyak 9% berkinerja kurang (kluster 2) menurut algoritme X-Means. Sementara berdasarkan K-Medoids, mahasiswa yang memiliki kinerja akademik baik sebanyak 97% dan memiliki kinerja akademik kurang sebanyak 3%.

4. Kesimpulan

Evaluasi nilai DBI terhadap hasil klusterisasi algoritme *X-Means* dan *K-Medoids* pada penelitian ini, memberikan kesimpulan bahwa *K-Medoids* lebih baik dari *X-Means*, meskipun selisih nilainya tidak terlalu jauh, DBI *K-Medoids* dan *X-Means* untuk jumlah kluster optimal (kluster = 2) secara berturut-turut 0.061 dan 0.091. Hasil klusterisasi data akademik mahasiswa baik dengan *K-Medoids* dan *X-Means*, menunjukkan bahwa mayoritas mahasiswa Prodi S1 Sistem Informasi Kampus XYZ (diatas 90%) tergolong kedalam kelompok mahasiswa yang memiliki kinerja akademik baik. Jumlah dataset pada penelitian ini masuk tergolong sedikit, akan lebih baik apabila jumlah dataset ditambah. Model klusterisasi dapat diimplementasikan menjadi sistem yang dapat digunakan untuk mengelompokkan kinerja mahasiswa secara *real time*.

5. Referensi

- Aryuni, M., Didik Madyatmadja, E., & Miranda, E. (2018). Customer Segmentation in XYZ Bank Using K-Means and K-Medoids Clustering. *Proceedings of 2018 International Conference on Information Management and Technology, ICIMTech 2018, September, 412-416.* <https://doi.org/10.1109/ICIMTech.2018.8528086>
- Aziz, F. N. R. F. J., Setiawan, B. D., & Arwani, I. (2018). Implementasi Algoritma K-Means untuk Klusterisasi Kinerja Akademik Mahasiswa. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer, 2(6), 2243-2251.*
- Fimawahib, L., & Rouza, E. (2021). Penerapan K-Means Clustering pada Penentuan Jenis Pembelajaran di Universitas Pasir Pengaraian. *INOVTEK Polbeng - Seri Informatika, 6(2), 234.*

<https://doi.org/10.35314/isi.v6i2.2096>

- Ge, J., Sun, H., Xue, C., He, L., Jia, X., He, H., & Chen, J. (2021). LPX: Overlapping community detection based on X-means and label propagation algorithm in attributed networks. *Computational Intelligence*, 37(1), 484–510. <https://doi.org/10.1111/coin.12420>
- Hakim, I., Rafid, M., & Anggraini, F. (2022). Pemanfaatan Machine Learning dengan Algoritma X-Means untuk Pemetaan Luas Panen, Produktivitas, dan Produksi Padi. *Building of Informatics, Technology and Science (BITS)*, 4(3). <https://doi.org/10.47065/bits.v4i3.2654>
- Herviany, M., Putri Delima, S., Nurhidayah, T., & Kasini, K. (2021). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Daerah Rawan Tanah Longsor Pada Provinsi Jawa Barat. *MALCOM: Indonesian Journal of Machine Learning and Computer Science*, 1(1), 34–40. <https://doi.org/10.57152/malcom.v1i1.60>
- Hidayati, N., Rizmayanti, A. I., Dewi, C. B. S., Fatmasari, R., & Gata, W. (2020). Penerapan Algoritma Klusterisasi dan Klasifikasi pada Tingkat Kepentingan Sistem Pembelajaran di Universitas Terbuka. *Swabumi*, 8(2), 134–142. <https://doi.org/10.31294/swabumi.v8i2.8385>
- Kaligis, G. B., & Yulianto, S. (2022). Analisa Perbandingan Algoritma K-Means, K-Medoids, Dan X-Means Untuk Pengelompokan Kinerja Pegawai. *IT-Explore: Jurnal Penerapan Teknologi Informasi Dan Komunikasi*, 1(3), 179–193. <https://doi.org/10.24246/itexplore.v1i3.2022.pp179-193>
- Kamila, I., Khairunnisa, U., & Mustakim, M. (2019). Perbandingan Algoritma K-Means dan K-Medoids untuk Pengelompokan Data Transaksi Bongkar Muat di Provinsi Riau. *Jurnal Ilmiah Rekayasa Dan Manajemen Sistem Informasi*, 5(1), 119. <https://doi.org/10.24014/rmsi.v5i1.7381>
- Mughnyanti, M., Efendi, S., & Zarlis, M. (2020). Analysis of determining centroid clustering x-means algorithm with davies-bouldin index evaluation. *IOP Conference Series: Materials Science and Engineering*, 725(1). <https://doi.org/10.1088/1757-899X/725/1/012128>
- Musfiani, M. (2019). Analisis Cluster Dengan Menggunakan Metode Partisi Pada Pengguna Alat Kontrasepsi Di Kalimantan Barat. *Bimaster : Buletin Ilmiah Matematika, Statistika Dan Terapannya*, 8(4), 893–902. <https://doi.org/10.26418/bbimst.v8i4.36584>
- Nirmal, S. (2019). Comparative study between k-means and k-medoids clustering algorithms. *International Research Journal of Engineering and Technology*, 839, 839–844. <https://www.irjet.net/archives/V6/i3/IRJET-V6I3154.pdf>
- Priyatman, H., Sajid, F., & Haldivany, D. (2019). Klusterisasi Menggunakan Algoritma K-Means Clustering untuk Memprediksi Waktu Kelulusan Mahasiswa. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, 5(1), 62. <https://doi.org/10.26418/jp.v5i1.29611>
- Septiani, I. W., Fauzan, A. C., & Huda, M. M. (2022). Implementasi Algoritma K-Medoids Dengan Evaluasi Davies-Bouldin-Index Untuk Klusterisasi Harapan Hidup Pasca Operasi Pada Pasien Penderita Kanker Paru-Paru. *Jurnal Sistem Komputer Dan Informatika (JSON)*, 3(4), 556. <https://doi.org/10.30865/json.v3i4.4055>
- Sholikhah, N. A. (2022). Studi Perbandingan Clustering Kecamatan di Kabupaten Bojonegoro Berdasarkan Keaktifan Penduduk Dalam Kepemilikan Dokumen Kependudukan. *Jurnal Statistika Dan Komputasi*, 1(1), 42–53. <https://doi.org/10.32665/statkom.v1i1.443>
- Supriyadi, A., Triayudi, A., & Sholihati, I. D. (2021). Perbandingan Algoritma K-Means Dengan K-Medoids Pada Pengelompokan Armada Kendaraan Truk Berdasarkan Produktivitas. *JUPI (Jurnal Ilmiah Penelitian Dan Pembelajaran Informatika)*, 6(2), 229–240. <https://doi.org/10.29100/jipi.v6i2.2008>
- Syahputra, I., Ilhamsyah, I., Rahmayuda, S., & Febrianto, F. (2022). Sistem Klusterisasi Data Kesehatan Penduduk Untuk Menentukan Rentang Derajat Kesehatan Daerah Menggunakan K-Means. *Jurnal Khatulistiwa Informatika*, 10(1), 66–73. <https://doi.org/10.31294/jki.v10i1.12872>

Utomo, W. (2021). The comparison of k-means and k-medoids algorithms for clustering the spread of the covid-19 outbreak in Indonesia. *ILKOM Jurnal Ilmiah*, 13(1), 31-35.
<https://doi.org/10.33096/ilkom.v13i1.763.31-35>