

Perbandingan Algoritma K-Means, X-Means Dan K-Medoids Untuk Klasterisasi Awak Kabin Lion Air

Ahmad Jurnal Wahidin¹, Dana Indra Sensuse²

¹Program Studi Teknologi Informasi, Universitas Bina Sarana Informatika, Indonesia

²Program Studi Teknologi Informasi, Universitas Indonesia, Indonesia

Email: ¹ahmad.ajn@bsi.ac.id, ²dana@cs.ui.id

INFORMASI ARTIKEL

Histori artikel:

Naskah masuk, 8 November 2021

Direvisi, 9 November 2021

Diiterima, 9 November 2021

ABSTRAK

Abstract- Lion Air is part of PT. Lion Group, which is a airline in Indonesia and a low-cost airline based in Jakarta, Lion Air throughout the year experienced an increase in its fleet and an increase in the number of flights, the greater the need for cabin crew, in addition to the recruitment process which must be selective, it is also necessary to monitor the crew. cabin crew so that cabin crew performance will continue to be well maintained so that several groups are formed called cabin crew monitoring groups, currently cabin crew clustering is done randomly so as to produce group members with different characteristics, clustering should be computerized by utilizing data mining clustering. The purpose of this study is to compare three algorithms by calculating the value of the Davies-Bouldin Index (DBI), at the data processing stage by eliminating missing values and determining attributes, it produces 100 data, at the modeling stage the most optimum results obtained using the k-means algorithm are 4 clusters and 6 attributes are shown with a DBI value of 0.792, while the DBI value of the x-means algorithm is 0.812 and the k-medoids algorithm is 1.700 so that the k-means algorithm is the best algorithm in this study.

Kata Kunci:

Algoritma K-Means,
Algoritma X-Means,
Algoritma K-Medoids,
Awak Kabin

Abstrak- Lion Air merupakan bagian dari PT. Lion Group yang merupakan sebuah maskapai swasta di Indonesia dan penerbangan bertarif rendah yang berpangkalan di Jakarta, Lion Air sepanjang tahun mengalami penambahan armada dan penambahan jumlah penerbangan maka semakin besar juga kebutuhan awak kabin, selain pada proses *recruitment* yang harus selektif diperlukan juga proses monitoring terhadap awak kabin agar performa awak kabin akan terus terjaga baik sehingga dibentuk beberapa kelompok yang disebut dengan group monitoring awak kabin, saat ini klasterisasi awak kabin dilakukan secara acak sehingga menghasilkan anggota kelompok dengan karakteristik berbeda, klasterisasi seharusnya terkomputerisasi dengan memanfaatkan data mining *clustering*. Tujuan dari penelitian ini adalah membandingkan tiga algoritma dengan menghitung nilai *Davies-Bouldin Index* (DBI), pada tahapan pengolahan data dengan menghilangkan missing value dan menentukan atribut maka menghasilkan 100 data, pada tahapan pemodelan hasil paling optimum yang didapat dengan menggunakan algoritma *k-means* adalah 4 klaster dan 6 atribut ditunjukkan dengan nilai DBI sebesar 0.792, sedangkan nilai DBI algoritma *x-means* sebesar 0.812 dan algoritma *k-medoids* sebesar 1,700 sehingga *k-means* menjadi algoritma terbaik pada penelitian ini.

Copyright © 2021 LPPM - STMIK IKMI Cirebon
This is an open access article under the CC-BY license

Penulis Korespondensi:

Ahmad Jurnal Wahidin

Program Studi Teknologi Informasi,
Universitas Bina Sarana Informatika
Jl. Kramat Raya No.98, RT.2/RW.9, Kwitang, Kec. Senen,
Kota Jakarta Pusat, Daerah Khusus Ibukota Jakarta, Indonesia
Email: ahmad.aj@bsi.ac.id

1. Pendahuluan

Dimasa sekarang ini berkembangnya teknologi informasi tumbuh secara pesat, jumlah data yang besar menimbulkan bertumpuknya data sehingga sulit untuk memperoleh informasi, oleh karena itu dibutuhkan data mining sebagai teknik komputasi pengolahan data. *Data mining* dapat diterapkan untuk menggali nilai tambah dari suatu kumpulan data berupa pengetahuan yang selama ini tidak diketahui secara manual [1], dalam data mining terdapat metode *clustering*, *clustering* adalah salah satu sub kategori data mining dan merupakan proses di mana sampel yang sama dibagi menjadi kelompok-kelompok yang disebut cluster [2]. Sayangnya, sebagian besar pengembang sistem berbasis data mining mengalami kesulitan dalam menentukan metode yang paling sesuai, jika salah maka akan mengalami kesulitan pada tahap desain dan implementasi. Algoritma *k-means* merupakan algoritma yang paling banyak digunakan dan sering digunakan sebagai algoritma perbandingan, *k-means* memiliki kelebihan ketelitian yang cukup tinggi terhadap ukuran objek, sehingga algoritma ini relatif lebih terukur dan efisien untuk pengolahan objek dalam jumlah besar [3], namun *k-means* juga memiliki kelemahan antara lain mengharapakan pengguna untuk menentukan banyak cluster terlebih dahulu dan juga sifat kaku (*hard*) yang dimiliki [4].

Berdasarkan pengukuran akurasi clustering *Silhouette Coefficient*, terbukti bahwa *x-means* memiliki tingkat akurasi yang lebih baik dibandingkan *k-means*, meskipun perbedaannya tidak terlalu signifikan[5], terdapat juga algoritma *k-medoids* yang lebih baik dari *k-means* dalam melakukan pengelompokan jika dibandingkan dengan melihat nilai validitasnya [6] dan memiliki kelebihan mampu mengatasi kelemahan dari algoritma *k-means* yang sensitif terhadap *outlier* [7].

Lion Air merupakan sebuah maskapai swasta terbesar di Indonesia dan sepanjang tahun operasionalnya mengalami penambahan armada dan jumlah penerbangan secara signifikan. Semakin bertambahnya jumlah armada dan jumlah penerbangan maka semakin besar juga kebutuhan awak kabin, selain pada proses *recruitment* yang harus selektif diperlukan juga proses monitoring terhadap awak kabin agar performa awak kabin akan

terus baik agar menghasilkan pelayanan yang maksimal terhadap penumpang, sehingga dibentuk beberapa kelompok yang disebut dengan group monitoring awak kabin yang berguna untuk memonitor setiap perkembangan dari awak kabin, proses monitoring dengan melakukan pengecekan terhadap awak kabin dengan melihat tinggi badan, berat badan, jam terbang, perlengkapan terbang serta penampilan. Saat ini klasterisasi awak kabin dilakukan secara acak sehingga menghasilkan anggota kelompok yang tidak mirip, dan penelitian ini bertujuan membandingkan kinerja algoritma *k-means*, *x-means* dan *k-medoids* yang digunakan pada klastering awak kabin pada maskapai lion air.

2. Metode

Pada penelitian ini, proses *clustering* awak kabin lion air dilakukan menggunakan tiga algoritma yaitu *k-means*, *x-means* dan *k-medoids*, dengan menggunakan data awak kabin yang didapat langsung dari divisi *Flight Operation Support (FOS)* dengan metode pengumpulan data observasi, wawancara dan studi pustaka. Data yang didapat masih terdapat *missing value* dan atribut yang tidak sesuai dengan pembahasan, sehingga perlu dilakukan tahap-tahap data mining untuk menghasilkan data yang siap untuk diproses atau disebut *data set*.

Untuk mengetahui algoritma terbaik dilakukan proses perbandingan antara ketiga algoritma *clustering* dengan melihat nilai *Davies Bouldin Indeks* yang merupakan salah satu metode validasi klaster untuk evaluasi kuantitatif dari hasil *clustering*. Pengukuran ini bertujuan meminimumkan jarak *inter-cluster* dan memaksimalkan jarak *intra-cluster*. Semakin kecil nilai DB Index menunjukkan skema klaster yang paling optimal [8].

Untuk menghitung nilai *Davies Bouldin Indeks* penelitian ini menggunakan *framework* data mining yaitu Rapidminer Studio. Rapidminer merupakan solusi untuk melakukan analisis terhadap data mining [9].

3. Hasil dan Pembahasan

Data yang digunakan adalah data awak kabin pada maskapai lion air sebanyak 100 data, dan algoritma terbaik akan diketahui melalui perbandingan nilai *Davis-Bouldin Index* (DBI).

3.1 Pengolahan Data

Data awak kabin yang didapat masih dengan 14 atribut ditampilkan pada tabel 1.

Tabel 1 Data Awak Kabin

Emp Code	52127806	Rating	B739	Height	160
Name	INTAN OKTAFIKA RUSDI	Birth Date	1Oct93	Weight	51
License	10020	Gender	F	Flight Hour	100
Religion	Islam	Position	SFA	Address	Mess Balaraja
Join	16Aug12	Educational Background	SMA		

Tahap awal pengolahan data adalah pembersihan data, pembersihan data merupakan proses menghilangkan noise dan data yang tidak konsisten atau data tidak relevan [10]. Dari Teknik Analisa, penulis mendapatkan kebutuhan-kebutuhan sistem yang akan digunakan dalam pemecahan masalah, kemudian dilakukan tahap seleksi data (*data selection*) untuk pemilihan atribut yang sesuai dengan kebutuhan dari penelitian sehingga didapat 7 atribut yaitu *EmpCode*, *Position*, *BirthDate*, *Join*, *Height*, *Weight* dan *Flight Hour*.

Tabel 2 Data Dengan 7 Atribut

ID	Position	Join	Usia	Height	Weight	Flight Hour
131382	SFA	17-Apr-16	01-Aug-97	162	60	85
131430	SFA	18-Apr-16	09-Feb-98	164	53	80
131504	SFA	19-Apr-16	04-Nov-96	160	48	100
133561	SFA	05-Sep-16	11-Oct-96	165	60	90
134701	SFA	04-Oct-16	05-Jul-96	158	52	100
134778	SFA	04-Oct-16	30-Jun-95	182	89	73
134986	SFA	21-Jan-16	18-Jan-95	172	56	110
135694	SFA	17-Mar-16	15-May-94	184	85	82
135700	SFA	22-Apr-16	17-Sep-96	173	71	45
135866	FA	13-May-16	20-Oct-96	160	51	82

Inisialisasi data yaitu melakukan perubahan data yang belum berjenis *numeric*, atribut *BirthDate* dirubah menjadi *Usia*, *Join* juga dirubah menjadi *Masa Kerja* dan *Position* dirubah (*SFA*=1, *FA*=2). Kemudian mengubah atribut *EmpCode* menjadi *ID*, *height* menjadi *tinggi badan*, *weight* menjadi *berat badan* dan *flight hour* menjadi *jam terbang*. Dan atribut *EmpCode* akan dijadikan label pada penelitian ini.

Setelah proses inisialisasi dan menghasilkan data yang siap untuk diproses dan disebut dataset yang ditampilkan pada tabel 3.

Table 3 Data Set

ID	Posisi	Masa Kerja	Usia	Tinggi Badan	Berat Badan	Jam Terbang
131382	1	5	24	162	60	85
131430	1	5	23	164	53	80
131504	1	5	25	160	48	100
133561	1	5	25	165	60	90
134701	1	5	25	158	52	100
134778	1	5	26	182	89	73
134986	1	5	26	172	56	110
135694	1	5	27	184	85	82
135700	1	5	25	173	71	45
135866	2	5	25	160	51	82

3.2 Pemodelan dengan Algoritma K-Means

Proses pemodelan pertama yaitu menggunakan algoritma k-means, pemodelan dilakukan dengan bantuan framework data mining yaitu Rapidminer untuk mengetahui nilai DBI.

Pemodelan awal bertujuan untuk menentukan jumlah kluster yang memiliki nilai DBI paling kecil, berdasarkan pemodelan dari beberapa percobaan maka didapat nilai DBI seperti dibawah ini:

Table 4 Hasil Nilai DBI Penentuan Jumlah Cluster

Percobaan ke-	Jumlah k=	Nilai Davies Bouldin Index
1	3	0,809
2	4	0,792
3	5	0,827
4	6	0,947
5	7	0,956
6	8	1,029
7	9	1,072
8	10	1,106

Berdasarkan pada tabel diatas nilai DBI paling kecil adalah percobaan ke-2 yaitu 0,792, karena semakin kecil nilai DBI maka semakin optimum kluster yang dihasilkan.

Selain pemodelan dengan percobaan jumlah kluster juga dilakukan percobaan dengan merubah jumlah atribut untuk mengetahui jumlah atribut paling optimum. Percobaan dilakukan dengan 6 atribut, 5 dan 4 atribut yaitu posisi, masa kerja, usia, tinggi badan, berat badan dan jam terbang, kemudian menghilangkan atribut jam terbang lalu usia.

Table 5 Nilai DBI Dengan Perubahan Jumlah Atribut

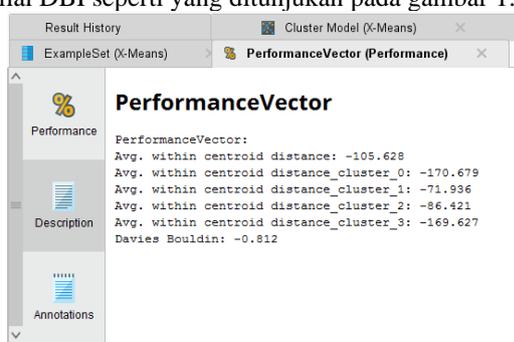
Jumlah k=	Nilai DBI dengan 6 Atribut	Nilai DBI dengan 5 Atribut	Nilai DBI dengan 4 Atribut
3	0,809	0,917	0,820
4	0,792	1,006	0,809
5	0,827	1,012	0,976
6	0,947	0,994	0,948
7	0,956	1,025	0,970
8	1,029	1,026	0,949

9	1,072	1,003	0,892
10	1,106	1,043	0,884

Tabel diatas menunjukkan bahwa nilai DBI yang paling kecil adalah dengan 6 atribut dan 4 klaster yaitu sebesar 0.792, sehingga klaster yang dihasilkan lebih optimum. sehingga pada penelitian ini klaster yang akan dibuat sebanyak 4 klaster.

3.3 Pemodelan dengan Algoritma X-Means

Selain pemodelan dengan algoritma k-means, juga dilakukan pemodelan dengan algoritma *data mining* clustering lainnya yaitu algoritma x-means sebagai pembanding, pemodelan dilakukan dengan jumlah klaster yang sama yaitu 4 klaster dan jumlah atribut yang sama yaitu 6 atribut dan menghasilkan nilai DBI seperti yang ditunjukkan pada gambar 1.

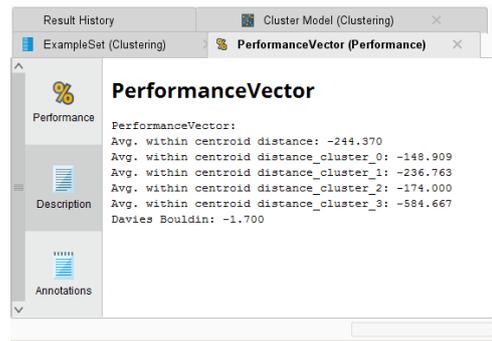


Gambar 1 Nilai *Davies Bouldin Index* pada algoritma X-Means

Pada hasil *PerformanceVector* diatas menunjukkan nilai *Davies Bouldin Index* yang dihasilkan sebesar 0,812, nilai DBI yang dihasilkan lebih tinggi dari nilai DBI pada percobaan menggunakan 4 klaster dalam algoritma k-means.

3.4 Pemodelan dengan Algoritma K-Medoids

Dua pemodelan yang sudah dilakukan sebelumnya yaitu menggunakan algoritma *k-means* dan *x-means* dengan menggunakan data yang sama, atribut serta jumlah klaster yang sama menghasilkan nilai DBI yang berbeda. Akan dilakukan juga pemodelan menggunakan algoritma *k-medoids* sebagai pembanding ketiga dengan jumlah klaster dan atribut sama yaitu 4 klaster dan 6 atribut.



Gambar 2 Nilai *Davies Bouldin Index* pada Algoritma K-Medoids

Pada hasil *PerformanceVector* diatas menunjukkan nilai *Davies Bouldin Index* yang dihasilkan sebesar 1,700, nilai DBI yang dihasilkan lebih tinggi dari nilai DBI pada percobaan menggunakan empat klaster dalam algoritma k-means dan x-means.

Sehingga hasil dari pemodelan dengan menggunakan 6 atribut dan 4 klaster yang telah dilakukan yaitu dengan algoritma k-means, algoritma x-means dan algoritma k-medoids menghasilkan nilai DBI dari masing-masing algoritma seperti yang ditunjukkan pada tabel 6.

Table 6 Perbandingan *Davies-Bouldin Index* Dari Tiga Algoritma

Algoritma	Davies Bouldin Index
K-Means	0,792
X-Means	0,812
K-Medoids	1,700

Dari tabel diatas hasil terbaik berdasarkan nilai *Davies-Bouldin Index* terdapat pada algoritma *k-means* dengan nilai DBI sebesar 0,792, nilai tersebut dikatakan baik karena hasil tersebut merupakan nilai terkecil yang diperoleh, karena semakin kecil nilai DBI atau semakin mendekati nilai nol maka semakin optimum klaster yang dihasilkan.

4. Kesimpulan

Dari pengukuran kinerja dengan melakukan perbandingan nilai DBI dari tiga algoritma yang telah dilakukan dengan menggunakan 100 data, maka pada penentuan nilai k atau jumlah klaster yang dilakukan menggunakan k-means nilai terbaik diperoleh pada percobaan kedua dengan nilai *Davies-Bouldin Index* (DBI) sebesar 0,792 yang menghasilkan empat klaster. Dengan pengurangan atribut menjadi 5 dan 4 pada data yang diolah terbukti tidak menurunkan nilai DBI, sehingga dengan menggunakan 6 atribut akan menghasilkan klaster yang lebih optimum. Algoritma k-means menghasilkan nilai DBI lebih kecil yaitu 0,792 dibandingkan dengan nilai DBI algoritma x-means

sebesar 0.812 dan algoritma k-medoids sebesar 1,700, sehingga k-means menjadi algoritma terbaik untuk klasterisasi awak kabin lion air.

Selanjutnya perlu dilakukan penelitian dengan membandingkan algoritma k-means dengan algoritma data mining *clustering* yang lainnya agar dapat mengetahui apakah terdapat algoritma yang lebih baik dari k-means.

Daftar Pustaka

- [1] A. N. Khormarudin, "Teknik Data Mining: Algoritma K-Means Clustering," *J. Ilmu Komput.*, pp. 1–12, 2016.
- [2] A. Fauzi, "Data Mining dengan Teknik Clustering Menggunakan Algoritma K-Means pada Data Transaksi Superstore," 2017.
- [3] A. Bastian, "Penerapan algoritma k-means clustering analysis pada penyakit menular manusia (studi kasus kabupaten Majalengka)," *J. Sist. Inf.*, vol. 14, no. 1, pp. 28–34, 2018.
- [4] K. Fitriah, "Perbandingan K-Means Clustering, Adaptive K-Means Clustering Dan Fuzzy K-Means Clustering Berdasarkan Internal Validation dan External Validation (Studi Kasus: Pengelompokan Kabupaten/Kota)." 2017.
- [5] D. Arundina, Shaufiah, and T. Suharto, "Implementasi Klasterisasi Menggunakan Algoritma X-Means," *Telkom Univ.*, 2010.
- [6] D. Marlina, N. F. Putri, A. Fernando, and A. Ramadhan, "Implementasi Algoritma K-Medoids dan K-Means untuk Pengelompokan Wilayah Sebaran Cacat pada Anak," *J. CoreIT*, vol. 4, no. 2, p. 64, 2018.
- [7] D. F. Pramesti, M. T. Furqon, and C. Dewi, "Implementasi Metode K-Medoids Clustering Untuk Pengelompokan Data Potensi Kebakaran Hutan/Lahan Berdasarkan Persebaran Titik Panas (Hotspot)," *J. Pengemb. Teknol. Inf. dan Ilmu Komput. e-ISSN*, vol. 2548, p. 964X, 2017.
- [8] W. Widiarina and R. S. Wahono, "Algoritma cluster dinamik untuk optimasi cluster pada algoritma k-means dalam pemetaan nasabah potensial," *J. Intell. Syst.*, vol. 1, no. 1, pp. 33–36, 2015.
- [9] M. Mardalius, "Pemanfaatan Rapid Miner Studio 8.2 Untuk Pengelompokan Data Penjualan Aksesoris Menggunakan Algoritma K-Means," *JURTEKSI (Jurnal Teknol. dan Sist. Informasi)*, vol. 4, no. 2, pp. 123–132, 2018.
- [10] F. Marisa, "Educational Data Mining (Konsep dan Penerapan)," *J. Teknol. Inf. Teor. Konsep, dan Implementasi*, vol. 4, no. 2, pp. 90–97, 2013.