

JEPIN

by Novia Hasdyna

Submission date: 07-Dec-2019 07:47PM (UTC-0800)

Submission ID: 1229304017

File name: FIX_sesuai_template_JEPIN_komplit.pdf (642.58K)

Word count: 2967

Character count: 17072



Reduksi Atribut Menggunakan Information Gain Untuk Optimasi Cluster Algoritma K-Means

Rozzi Kesuma Dinata¹, Haried Novriando², Novia Hasdyna³, Sujacka Retno⁴

¹Program Studi Teknik Informatika Universitas Malikussaleh
Jl. Medan-Banda Aceh, Cot Teungku Nie, Reuleut, Krueng Geukueh, Kab. Aceh Utara- Aceh
¹rozzi@unimal.ac.id

²Jurusan Informatika Universitas Tanjungpura
Jl. Prof. Hadari Nawawi, Pontianak, Kalimantan Barat
²haried@informatika.untan.ac.id

³Program Studi Teknik Informatika Universitas Islam Kebangsaan Indonesia
Jl. Medan – Banda Aceh Sp. Alue Awe, Lhokseumawe, Aceh, Indonesia
³noviahasdyna@gmail.com

⁴Program Studi Magister Teknik Informatika Universitas Sumatera Utara
Jl. Universitas No.9, Padang Bulan, Kota Medan, Sumatera Utara
⁴sujackaretno@gmail.com

Abstrak— Proses *clustering* dengan algoritma K-Means pada dataset yang memiliki banyak atribut akan mempengaruhi besarnya jumlah iterasi. Pada penelitian ini, metode *Information Gain* digunakan untuk mereduksi atribut dataset. Dataset yang telah direduksi atribut akan dilanjutkan proses *clustering* dengan K-Means. Dataset yang dianalisis pada penelitian ini adalah data *Hepatitis C Virus* yang diperoleh dari *UCI Machine Learning Repository*, dengan 29 atribut dan 1385 jumlah data. Hasil penelitian ini menunjukkan bahwa rata-rata jumlah iterasi yang diperoleh dari 10 kali pengujian dengan menggunakan K-Means konvensional diperoleh rata-rata sebesar 32 iterasi, sedangkan K-Means dengan reduksi atribut diperoleh rata-rata sebesar 27.7 iterasi. Nilai validitas cluster dihitung menggunakan *Davies-Bouldin Index* (DBI). Nilai DBI pada k-means konvensional sebesar 2.1972, sedangkan DBI pada K-Means yang telah direduksi 1 atribut sampai 5 atribut diperoleh nilai rata-rata DBI masing-masing sebesar 2.0290, 1.8771, 1.8641, 1.8389, dan 1.8117.

Kata kunci— Reduksi Atribut, *Information Gain*, K-Means, *Davies-Bouldin Index*

I. PENDAHULUAN

Clustering adalah proses pengelompokan objek data menjadi beberapa cluster yang terpisah sehingga data yang ada di dalam masing-masing cluster tersebut menjadi sebuah kelompok data yang memiliki kemiripan yang relatif sama [1]. Ada banyak teknik yang dapat digunakan untuk proses *clustering* seperti *Single Linkage*, *Complete Linkage*, *Average Linkage*, *Fuzzy C-Means*, *Kohonen SOM*, *LVQ* dan *K-Means* [2].

Salah satu permasalahan pada proses *clustering* dengan algoritma k-means adalah banyaknya atribut dataset yang menyebabkan bertambahnya jumlah iterasi [3]. Metode

seleksi fitur dapat digunakan untuk mereduksi atribut dataset [4]. Pada penelitian ini, metode seleksi fitur yang digunakan adalah *Information Gain*. Metode ini digunakan untuk mereduksi atribut dataset, yaitu satu atribut, dua atribut, tiga atribut, empat atribut dan lima atribut. Setelah direduksi atribut dataset, akan dilanjutkan proses *clustering* dengan algoritma k-means dan selanjutnya akan dianalisis hasil validasi dengan menggunakan *Davies-Bouldin Index*.

Davies-Bouldin index merupakan salah satu metode yang bertujuan untuk menganalisis peningkatan hasil *clustering* dengan mengevaluasi besarnya nilai kohesi dan separasi [5].

Pada penelitian data yang akan dianalisis yaitu data *Hepatitis C Virus Dataset* yang diperoleh dari *UCI Machine Learning Repository* [6]. Dataset yang direduksi atribut akan dianalisis pengaruhnya terhadap jumlah iterasi dan optimasi hasil evaluasi *clustering* pada algoritma k-means.

II. TINJAUAN PUSTAKA

A. *Information Gain*

Information gain merupakan salah satu metode seleksi fitur yang banyak dipakai oleh peneliti untuk menentukan batas dari kepentingan sebuah atribut [7]. Nilai *information gain* diperoleh dari nilai *entropy* sebelum pemisahan dikurangi dengan nilai *entropy* setelah pemisahan. Pengukuran nilai ini hanya digunakan sebagai tahap awal untuk penentuan atribut yang nantinya akan digunakan atau dibuang. Atribut yang memenuhi kriteria pembobotan yang nantinya akan digunakan dalam proses klasifikasi sebuah algoritma [8]. Pemilihan fitur dengan *information gain* dilakukan dalam 3 tahapan, yaitu:

1. Menghitung nilai *information gain* untuk setiap atribut dalam dataset original.

2. Tentukan batas (*threshold*) yang diinginkan. Hal ini akan memungkinkan atribut yang berbobot sama dengan batas atau lebih besar akan dipertahankan serta membuang atribut yang berada dibawah batas.
3. Dataset diperbaiki dengan pengurangan atribut. Pengukuran atribut ini pertama kali dipelopori oleh Claude Shannon didalam teori informasi (Gallager and Fellow2001) serta dituliskan sebagai berikut:

$$Entropy(s) = \sum_i^c -p_i \log_2 p_i \quad (1)$$

$$Info_A(D) = - \sum_{j=1}^v \frac{D_j}{D} \times I(D_j) \quad (2)$$

$$Gain(A) = I(D) - I(A) \quad (3)$$

Keterangan: $Gain(A)$ = Information atribut A
 $I(D)$ = Total entropy
 $I(A)$ = entropy A

B. Algoritma K-Means

Salah satu metode dalam teknik data mining yang dapat digunakan untuk mengelompokkan data atau *Clustering* sebuah data kedalam bentuk satu *cluster* atau lebih *cluster* adalah k-means[9]. Sarwono mengemukakan secara detail, algoritma *K-Means* adalah sebagai berikut yaitu :

1. Tentukan nilai k sebagai jumlah *cluster* yang diinginkan.
2. Tentukan nilai acak atau random untuk pusat *cluster* awal centroid sebanyak k, dengan menggunakan rumus jarak untuk menghitung jarak *Euclidean Distance* yaitu :

$$d(x_i, \mu_j) = \sqrt{\sum_{i=1}^n (x_i - \mu_j)^2} \quad (4)$$

Dimana: x_i = data kriteria
 μ_j = centroid pada *cluster* ke-js

3. Kelompokkan data berdasarkan nilai jarak terkecil setiap data.
4. Perbaharui centroid baru dari rata-rata *cluster* dengan persamaan:

$$\mu_j(t+1) = \frac{1}{N_{sj}} \sum_{j \in S_j} x_j \quad (5)$$

Keterangan :
 $\mu_j(t+1)$ = centroid baru pada iterasi (t+1)
 N_{sj} = Data pada *cluster* S_j

5. Lakukan perulangan dari langkah 2 hingga 5, sampai anggota tiap *cluster* tidak ada yang berubah.

C. Davies-Bouldin Index

Untuk mengevaluasi hasil *cluster* pada suatu metode *clustering* dapat menggunakan metode Davies-Bouldin Index. Landasan dari metode ini adalah nilai kohesi dan separasi [10].

Pada proses *clustering*, kohesi merupakan jumlah dari kedekatan data terhadap *centroid* dari *cluster* yang diikuti. Sedangkan separasi didasarkan pada jarak antar *centroid* dari *cluster* tersebut [11].

Langkah-langkah untuk menghitung nilai *Davies-Bouldin Index* adalah sebagai berikut [12]:

1. *Sum of Square Within cluster* (SSW)
Sum of Square Within cluster (SSW) adalah rumus yang digunakan untuk mencari matriks kohesi dalam sebuah *cluster* ke-i seperti persamaan berikut :

$$ssw_i = \frac{1}{m_i} \sum_{j=1}^{m_i} d(x_j, c_i) \quad (6)$$

2. *Sum of Square Between cluster* (SSB)
Sum of Square Between cluster (SSB) adalah sebuah persamaan untuk mengenali fungsi pemisah antar *cluster*, seperti persamaan berikut [13] :

$$SSB_{i,j} = d(c_i, c_j) \quad (7)$$

3. *Ratio* (Rasio)
 Perhitungan rasio (R_{ij}) adalah memabandingkan nilai *cluster* i dan *cluster* j, seperti pada persamaan dibawah ini:

$$R_{ij} = \frac{SSW_i + SSW_j}{SSB_{ij}} \quad (8)$$

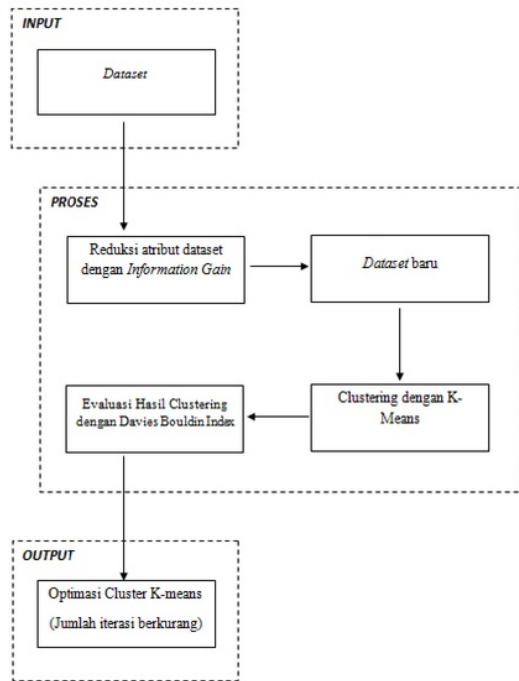
4. *Davies-Bouldin Index* (DBI)
 Setelah didapatkan nilai rasio, akan digunakan untuk mencari nilai *Davies-Bouldin Index* (DBI) dengan rumus berikut [14]:

$$DBI = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} (R_{i,j}) \quad (9)$$

Adapun semakin kecil nilai DBI yang diperoleh (non-negatif ≥ 0), maka akan semakin baik *cluster* algoritma *K-Means* [15].

III. METODE PENELITIAN

Framework reduksi atribut dataset untuk optimasi *cluster* algoritma k-means adalah sebagai berikut:



Gambar 1. Framework Penelitian

Berdasarkan gambar diatas, adapun langkah-langkah dalam penelitian ini yaitu:

1. Input Dataset Original
2. Proses reduksi atribut dengan *Information Gain*
3. Pembentukan Dataset baru setelah direduksi
4. Proses Clustering dengan K-means
5. Evaluasi hasil clustering dengan *Davies Bouldin Index*
6. Optimasi Cluster K-Means (Jumlah iterasi berkurang)

IV. HASIL

A. Nilai *Information Gain* pada Hepatitis C Virus Dataset

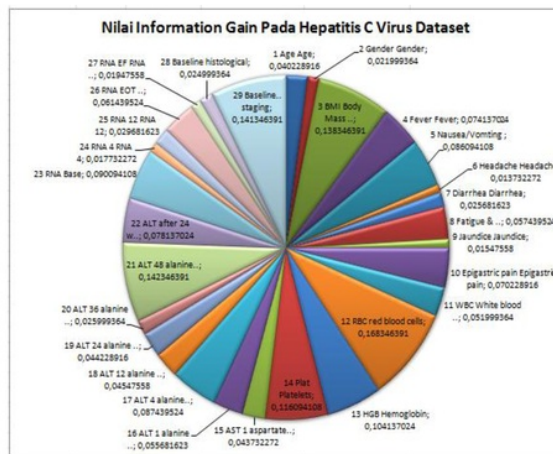
Berdasarkan persamaan (1), (2), dan (3) diperoleh hasil perhitungan information gain seperti pada tabel 1.

Tabel 1. Nilai Information Gain pada Hepatitis C Virus Dataset

No	Atribut	Information Gain	Reduksi
1	Age	0,0019881	
2	Gender	0,0034777	
3	BMI Body Mass ..	0,0046983	
4	Fever	0,0011359	
5	Nausea/Vomting	0,0026763	
6	Headache	0,0000891	Reduksi 1
7	Diarrhea	0,0007394	
8	Fatigue & ..	0,0010000	

9	Jaundice	0,0019145	
10	Epigastric pain	0,0034203	
11	WBC White blood ..	0,0000943	Reduksi 2
12	RBC red blood cells	0,0016022	
13	HGB Hemoglobin	0,0011803	
14	Plat Platelets	0,0008341	
15	AST 1 aspartate..	0,0005072	
16	ALT 1 alanine ..	0,0011256	
17	ALT 4 alanine..	0,0010269	
18	ALT 12 alanine ..	0,0004240	Reduksi 5
19	ALT 24 alanine ..	0,0002474	Reduksi 4
20	ALT 36 alanine ..	0,0001732	Reduksi 3
21	ALT 48 alanine..	0,0004936	
22	ALT after 24 w..	0,0027067	
23	RNA Base	0,3361371	
24	RNA 4 RNA 4	0,3679658	
25	RNA 12 RNA 12	0,0026596	
26	RNA EOT ..	0,0004377	
27	RNA EF RNA ..	0,0041398	
28	Baseline histological	0,0006377	
29	Baseline staging	0,0167970	

Berdasarkan tabel 1, dapat dijabarkan dalam bentuk diagram berikut.



Gambar 2. Diagram Hasil Information Gain Pada Hepatitis C Virus Dataset

Nilai *information gain* terkecil adalah atribut ke-6, kedua adalah atribut ke-11, ketiga adalah atribut ke-20, reduksi keempat adalah atribut ke-19, dan reduksi kelima adalah atribut ke-18. Adapun pada penelitian ini, atribut yang direduksi adalah atribut dengan nilai *information gain* terkecil. Proses pereduksian dengan reduksi satu atribut, dua atribut, tiga atribut, dan empat atribut. Selanjutnya akan di analisis hasil komparasinya dengan DBI.

B. Proses *Clustering dengan K-Means*

Hasil perhitungan clustering dengan algoritma k-means dengan reduksi atribut dapat dilihat pada tabel 1, atribut *Headache*(x_6) direduksi karena memperoleh nilai *Information Gain* terkecil sebesar 0.0000891. Pada pengujian kali ini

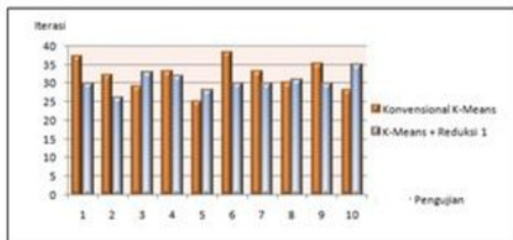
dilakukan sebanyak 10x pengujian yang diperlihatkan pada tabel 2 berikut:

Tabel 2. Pengujian K-means Clustering dengan Reduksi Satu Atribut

Pengujian	Iterasi	
	Konvensional K-Means	K-Means + Reduksi 1
1	37	30
2	32	26
3	29	33
4	33	32
5	25	28
6	38	30
7	33	30
8	30	31
9	35	30
10	28	35
Rata-rata	32	30,5

Berdasarkan Tabel 2, pengujian K-Means konvensional memperoleh jumlah iterasi terbesar untuk menyelesaikan proses clustering sebesar 38 iterasi, pada pengujian ke 5 diperoleh jumlah iterasi terendah sebesar 25 iterasi. Sedangkan menggunakan K-Means dengan mereduksi satu atribut diperoleh jumlah iterasi terbesar pada pengujian ke 10 sebesar 35 iterasi, pada pengujian ke 2 diperoleh jumlah iterasi terendah sebesar 26 iterasi.

Rata-rata jumlah iterasi yang diperoleh dari masing-masing 10x pengujian dengan menggunakan K-Means konvensional diperoleh rata-rata sebesar 32 iterasi, sedangkan dengan k-means dengan reduksi satu atribut diperoleh rata-rata sebesar 30.5 iterasi.



Gambar 3. Grafik Hasil clustering k-means dengan reduksi satu atribut

Atribut *Headache*(x_6) dan *WBC White Blood Cell*(x_{11}) direduksi karena memperoleh nilai *Information Gain* 2 terkecil masing-masing sebesar 0.0000891 dan 0.0000943. Pada pengujian kali ini dilakukan sebanyak 10x pengujian yang diperlihatkan pada tabel 3 berikut:

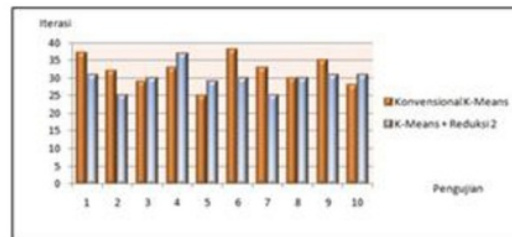
Tabel 3. Pengujian K-means Clustering dengan Reduksi dua Atribut

Pengujian	Iterasi	
	Konvensional K-Means	K-Means + Reduksi 2
1	37	31
2	32	25
3	29	30

4	33	37
5	25	29
6	38	30
7	33	25
8	30	30
9	35	31
10	28	31
Rata-rata	32	29,9

Berdasarkan Tabel 3, pada pengujian K-Means konvensional memperoleh jumlah iterasi terbesar untuk menyelesaikan proses clustering sebesar 38 iterasi, pada pengujian ke 5 diperoleh jumlah iterasi terendah sebesar 25 iterasi. Sedangkan menggunakan K-Means dengan mereduksi dua atribut diperoleh jumlah iterasi terbesar pada pengujian ke 4 sebesar 37 iterasi, pada pengujian ke 2 dan ke 7 diperoleh jumlah iterasi terendah sebesar 25 iterasi.

Rata-rata jumlah iterasi yang diperoleh dari masing-masing 10x pengujian dengan menggunakan K-Means konvensional diperoleh rata-rata sebesar 32 iterasi, sedangkan dengan k-means dengan reduksi dua atribut diperoleh rata-rata sebesar 29.9 iterasi.



Gambar 4. Grafik Hasil clustering k-means dengan reduksi dua atribut

Atribut *Headache*(x_6), *WBC White Blood Cell*(x_{11}) dan *ALT 36*(x_{20}) direduksi karena memperoleh nilai *Information Gain* 3 terkecil masing-masing sebesar 0.0000891, 0.0000943 dan 0.0001732. Pada pengujian kali ini dilakukan sebanyak 10x pengujian diperlihatkan pada tabel 4 berikut:

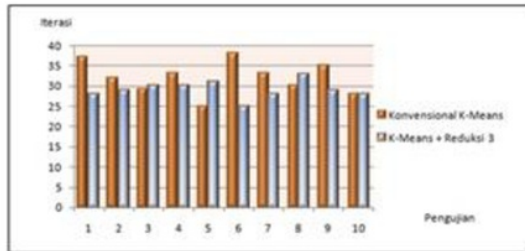
Tabel 4. Pengujian K-means Clustering dengan Reduksi Tiga Atribut

Pengujian	Iterasi	
	Konvensional K-Means	K-Means + Reduksi 3
1	37	28
2	32	29
3	29	30
4	33	30
5	25	31
6	38	25
7	33	28
8	30	33
9	35	29
10	28	28
Rata-rata	32	29,1

Berdasarkan tabel 4, pada pengujian K-Means konvensional memperoleh jumlah iterasi terbesar untuk menyelesaikan proses clustering sebesar 38 iterasi, pada pengujian ke 5 diperoleh jumlah iterasi terendah sebesar 25

iterasi. Sedangkan menggunakan K-Means dengan mereduksi tiga atribut diperoleh jumlah iterasi terbesar pada pengujian ke 8 sebesar 33 iterasi, pada pengujian ke 6 diperoleh jumlah iterasi terendah sebesar 25 iterasi.

Rata-rata jumlah iterasi yang diperoleh dari masing-masing 10x pengujian dengan menggunakan K-Means konvensional diperoleh rata-rata sebesar 32 iterasi, sedangkan dengan k-means reduksi tiga atribut diperoleh rata-rata sebesar 29.1 iterasi.



Gambar 5. Grafik Hasil clustering k-means dengan reduksi tiga atribut

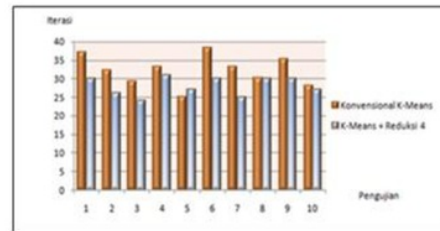
Atribut *Headache*(x_6), *WBC White Blood Cell*(x_{11}), *ALT 36*(x_{20}) dan *ALT 24*(x_{19}) direduksi karena memperoleh nilai *Information Gain* 4 terkecil masing-masing sebesar 0.0000891, 0.0000943, 0.0001732 dan 0.0002474. Pengujian kali ini dilakukan sebanyak 10x pengujian diperlihatkan pada tabel 5 berikut:

Tabel 5. Pengujian K-means Clustering dengan Reduksi Empat Atribut

Pengujian	Iterasi	
	Konvensional K-Means	K-Means + Reduksi 4
1	37	30
2	32	26
3	29	24
4	33	31
5	25	27
6	38	30
7	33	25
8	30	30
9	35	30
10	28	27
Rata-rata	32	28

Berdasarkan Tabel 5, pada K-Means konvensional memperoleh jumlah iterasi terbesar untuk menyelesaikan proses clustering sebesar 38 iterasi, pada pengujian ke 5 diperoleh jumlah iterasi terendah sebesar 25 iterasi, sedangkan menggunakan K-Means dengan mereduksi empat atribut diperoleh jumlah iterasi terbesar pada pengujian ke 4 sebesar 31 iterasi, pada pengujian ke 3 diperoleh jumlah iterasi terendah sebesar 24 iterasi.

Rata-rata jumlah iterasi yang diperoleh dari masing-masing 10x pengujian dengan menggunakan K-Means konvensional diperoleh rata-rata sebesar 32 iterasi, sedangkan dengan k-means reduksi empat diperoleh rata-rata sebesar 28 iterasi.



Gambar 6. Grafik Hasil clustering k-means dengan reduksi empat atribut

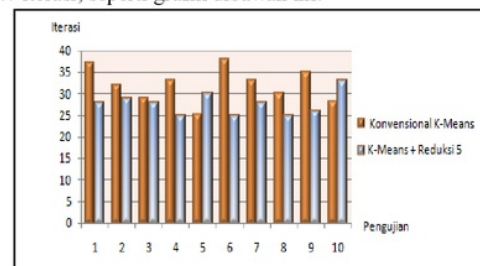
Atribut *Headache*(x_6), *WBC White Blood Cell*(x_{11}), *ALT 36*(x_{20}), *ALT 24*(x_{19}) dan *ALT 12*(x_{18}) direduksi karena memperoleh nilai *Information Gain* 5 terkecil masing-masing sebesar 0.0000891, 0.0000943, 0.0001732, 0.0002474 dan 0.0004240. Pada pengujian kali ini dilakukan sebanyak 10x pengujian diperlihatkan pada tabel 6 berikut:

Tabel 6. Pengujian K-means Clustering dengan Reduksi Lima Atribut

Pengujian	Iterasi	
	Konvensional K-Means	K-Means + Reduksi 5
1	37	28
2	32	29
3	29	28
4	33	25
5	25	30
6	38	25
7	33	28
8	30	25
9	35	26
10	28	33
Rata-rata	32	27.7

Berdasarkan Tabel 6, pada pengujian K-Means konvensional memperoleh jumlah iterasi terbesar untuk menyelesaikan proses clustering sebesar 38 iterasi, pada pengujian ke 5 diperoleh jumlah iterasi terendah sebesar 25 iterasi. Sedangkan menggunakan K-Means dengan mereduksi lima atribut diperoleh jumlah iterasi terbesar pada pengujian ke 10 sebesar 33 iterasi, pada pengujian ke 4, 6 dan 8 diperoleh jumlah iterasi terendah sebesar 25 iterasi.

Rata-rata jumlah iterasi yang diperoleh dari masing-masing 10x pengujian dengan menggunakan K-Means konvensional diperoleh rata-rata sebesar 32 iterasi, sedangkan dengan k-means reduksi lima atribut diperoleh rata-rata sebesar 27.7 iterasi, seperti grafik dibawah ini.



Gambar 7. Grafik Hasil clustering k-means dengan reduksi lima atribut

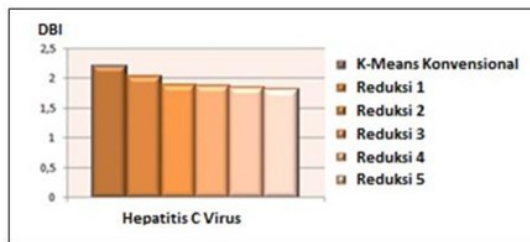
C. Nilai *Davies-Bouldin Index* (DBI) [21]

Hasil perhitungan nilai DBI pada K-Means konvensional dengan K-Means yang direduksi atribut dapat dilihat pada tabel komparasi berikut:

Tabel 7. Komparasi Nilai DBI

Nilai <i>Davies-Bouldin Index</i> (DBI)					
K-Means	K-Means + Reduksi Atribut				
	1	2	3	4	5
2.1972	2.0290	1.8771	1.8641	1.8389	1.8117

Berdasarkan tabel 7. nilai DBI pada k-means konvensional sebesar 2.1972, sedangkan DBI pada K-Means yang telah direduksi 1 atribut sampai 5 atribut diperoleh nilai DBI masing-masing sebesar 2.0290, 1.8771, 1.8641, 1.8389, dan 1.8117, seperti pada grafik dibawah ini.



Gambar 8. Grafik Komparasi Nilai DBI

V. KESIMPULAN

Berdasarkan hasil penelitian yang telah dilakukan, metode reduksi atribut Information Gain mampu mengoptimasi cluster algoritma k-means. Nilai evaluasi clustering *Davies Bouldin Index* terbaik adalah pada reduksi atribut ke lima sebesar 1,8117. Adapun nilai DBI pada k-means sebelum direduksi atribut adalah sebesar 2,1972. Hasil penelitian ini juga menunjukkan bahwa metode *Information Gain* dapat mengurangi jumlah iterasi pada proses clustering K-Means., dengan rata-rata jumlah iterasi pada K-Means sebelum direduksi adalah 32 iterasi. Rata-rata jumlah iterasi pada K-Means dengan reduksi satu sampai lima atribut adalah sebesar 27,7 iterasi.

REFERENSI

[1] Archana, S, Avantika, Y & Ajay, R, *K-Means with Three different Distance Metrics*. IJCA, Vol 67, No 10, 2013.

[2] Arunkumar, N., et al. *K-means clustering and neural network for object detecting and identifying abnormality of brain tumor*. Soft Computing, 2019, 23.19: 9083-9096.

[3] Bates, A & Kalita, J, *Counting Clusters in Twitter Posts*, Proceedings of the 2nd International Conference on Information Technology for Competitive Strategies, 2016, pp. 85.

[4] Bernard J. D. S, *Peningkatan Hasil Evaluasi Cluster Davies-Bouldin Index Dengan Penentuan Titik Pusat cluster awal Algoritma K-Means*, Universitas Sumatera Utara, 2018.

[5] Chandani, Vinita, et al. *Komparasi algoritma klasifikasi Machine Learning dan feature selection pada analisis sentimen review film*. Journal of Intelligent Systems, 2015, 1.1: 56-60.

[6] Dibya, J.B & Anil, K.G, *Effect of Different Distance Measures on the Performance of K-Means Algorithm: An Experimental Study in Matlab*. IJCBIT, Vol 5, No 2, 2014

[7] Dabaghi, Ehsan; Kashanian, Hooman. *Feature dimension reduction of multisensor data fusion using principal component fuzzy analysis*. International Journal of Engineering, 2017, 30.4: 493-499.

[8] Hartono, Ongko, E. & Abdullah, D. *Determining a cluster centroid of k-means clustering using genetic algorithm*, IJCSSE, 4(6), 2015,160-164

[9] Oyelade, O. J, Oladipupo, O. O & Obagbuwa, I.C. *Application of K-Means Clustering Algorithm for Prediction of Students's Academic Performance*, IJCSIS, Vol 7, No 1, 2010.

[10] Prasetyo, Eko. *Reduksi Dimensi Set Data dengan DRC pada Metode Klasifikasi SVM dengan Upaya Penambahan Komponen Ketiga*. Prosiding SNATIF, 2014, 293-300.

[11] Shakeel, P. Mohamed, et al. *Cloud based framework for diagnosis of diabetes mellitus using K-means clustering*. Health information science and systems, 2018, 6.1: 16.

[12] Silwattananusarn, T & Tuamsuk, K, *Data Mining and Its Applications for Knowledge Management: A Literature Review from 2007 to 2012*, IJDKP, Vol 2, No 5, 2012.

[13] Unnati, R.R. & Chaita, J. *Implementing & Improvisation of K-means Clustering Algorithm*, IJCSMC, Vol 5, 191203, Issue 5, 2016.

[14] Waruwu, Y. F., Zarlis, M., Nababan, E. B., & Ziliwu, M. S, *Seleksi Atribut Pada Algoritma Radial Basis Function Neural Network Menggunakan Information Gain*, Seminar Nasional Royal (SENAR), Vol. 1, No. 1, 2018, pp. 21-24.

[15] Zhan, Qiang, *An Improved K-Means Algorithm Based on Structure Features*, Journal of Software, Vol 12, No 1, 2017.

JEPIN

ORIGINALITY REPORT

23%

SIMILARITY INDEX

14%

INTERNET SOURCES

9%

PUBLICATIONS

21%

STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to Universitas Sebelas Maret Student Paper	3%
2	Submitted to Universitas Amikom Student Paper	3%
3	Submitted to Sriwijaya University Student Paper	2%
4	Submitted to iGroup Student Paper	1%
5	jurnal.umk.ac.id Internet Source	1%
6	eprints.dinus.ac.id Internet Source	1%
7	kc.umn.ac.id Internet Source	1%
8	Submitted to Fakultas Ekonomi Universitas Indonesia Student Paper	1%
9	ejournal.polbeng.ac.id	

Internet Source

1%

10

Submitted to Universitas Putera Batam

Student Paper

1%

11

jurnal.fikom.umi.ac.id

Internet Source

1%

12

sites.google.com

Internet Source

1%

13

link.springer.com

Internet Source

1%

14

citeseerx.ist.psu.edu

Internet Source

1%

15

ejournal.uin-suka.ac.id

Internet Source

1%

16

journal.budiluhur.ac.id

Internet Source

1%

17

Submitted to Forum Komunikasi Perpustakaan
Perguruan Tinggi Kristen Indonesia (FKPPTKI)

Student Paper

1%

18

docobook.com

Internet Source

1%

19

ojs.amikom.ac.id

Internet Source

1%

Rachna Jain, Nikita Jain, Akshay Aggarwal, D.

20 Jude Hemanth. "Convolutional neural network based Alzheimer's disease classification from magnetic resonance brain images", Cognitive Systems Research, 2019 1%

Publication

21 Submitted to Universitas Brawijaya 1%

Student Paper

22 Bernad Jumadi Dehotman Sitompul, Opim Salim Sitompul, Poltak Sihombing. "Enhancement Clustering Evaluation Result of Davies-Bouldin Index with Determining Initial Centroid of K-Means Algorithm", Journal of Physics: Conference Series, 2019 1%

Publication

23 es.scribd.com 1%

Internet Source

Exclude quotes Off Exclude matches < 1%
Exclude bibliography On