

**Indeks Calinski – Harabasz Analisis Fuzzy C – Means dan K – Means Cluster Kabupaten/Kota di Provinsi Jambi Menurut Potensi Pertambangan, Penggalan, Pengadaan Listrik, dan Gas**

*Calinski – Harabasz Index Fuzzy C – Means and K – Means Cluster Regencies/Cities in Jambi Province According to Mining, Quarrying, Electricity Procurement, and Gas Potentials*

**Bunga Mardhotillah<sup>1\*</sup>, Amril Fadli<sup>2</sup>, Edi Elisa<sup>3</sup>, Zurweni<sup>4</sup>**

<sup>1</sup>Program Studi Matematika Universitas Jambi

<sup>2</sup>Kementerian ESDM RI

<sup>3</sup>Universitas Pendidikan Ganesha

<sup>4</sup>Universitas Jambi

\* Email: [bunga.mardhotillah@gmail.com](mailto:bunga.mardhotillah@gmail.com)

**Abstrak**

Penelitian ini bertujuan untuk membandingkan Analisis *Fuzzy C – Means* dan *K – Means Cluster* dengan menghitung Indeks Calinski – Harabasz, di mana semakin tinggi Indeks Calinski – Harabasz suatu analisis *cluster*, semakin baik *cluster* yang terbentuk. Analisis Data menggunakan software JASP, data yang digunakan adalah data potensi pertambangan, penggalan, pengadaan listrik, dan gas berupa data kontribusi sektor – sektor tersebut dalam PDRB Kabupaten/Kota di Provinsi Jambi. Hasil penelitian menunjukkan dengan Analisis *Cluster Fuzzy C – Means*, terbentuk dua *clusters*, sedangkan dengan Analisis *K – Means* terbentuk tiga *clusters*. Indeks Calinski – Harabasz *K – Means* lebih tinggi dibandingkan dengan *Fuzzy C – Means*. Hasil penelitian ini menyimpulkan bahwa, berdasarkan perbandingan Indeks Calinski – Harabasz, Analisis *Cluster K – Means* lebih baik dibandingkan dengan *Fuzzy C – Means Cluster*.

Kata Kunci: Indeks Calinski – Harabasz, Fuzzy C Means, K – Means, Analisis Cluster

**Abstract**

*This study aims to compare the Fuzzy C-Means and K-Means Cluster Analysis by calculating the Calinski-Harabasz Index, where the higher the Calinski-Harabasz Index of a cluster analysis, the better the cluster formed. Data analysis using JASP software, the data used is data on the potential for mining, quarrying, electricity and gas procurement in the form of data on the contribution of these sectors in the GRDP of the Regency/City in Jambi Province. The results showed that with the Fuzzy C-Means Cluster analysis, two clusters were formed, while the K-Means analysis formed three clusters. Calinski – Harabasz K – Means index is higher than Fuzzy C – Means. The results of this study conclude that, based on the comparison of the Calinski - Harabasz Index, the K - Means Cluster Analysis is better than the Fuzzy C - Means Cluster.*

Keywords: Calinski – Harabasz Index, Fuzzy C – Means, K – Means, Cluster Analysis

**Pendahuluan**

Sektor Pertambangan, Penggalian, Pengadaan Listrik, dan Gas merupakan sektor yang memiliki potensi krusial serta berkontribusi signifikan dalam Produk Domestik Regional Bruto (PDRB) Kabupaten/Kota, sehingga diperlukan pengelompokan Kabupaten/Kota di Provinsi Jambi berdasarkan empat sektor tersebut, untuk kemudian ditetapkan kategori cluster berdasarkan potensi tiap Kabupaten/Kota. Adapun pengelompokannya menggunakan Analisis Cluster K – Means dan Fuzzy C – Means Cluster. Kemudian ditentukan metode terbaik dari kedua metode tersebut dengan membandingkan Indeks Calinski – Harabasz. K – Means merupakan salah satu Teknik cluster non – hirarki untuk mempartisi obyek yang ada ke dalam beberapa cluster berdasar karakteristiknya. Tujuan peneglompokan menggunakan K – Means adalah meminimalkan variasi dalam satu cluster dan memaksimalkan variasi antar cluster. Sedangkan dalam Fuzzy C – Means, pengelompokan didasarkan pada logika Fuzzy. Logika Fuzzy merupakan logika yang memiliki nilai samar atau kekaburan (fuzziness) antara benar atau salah. Pengelompokan didasarkan pada potensi pertambangan, penggalian, pengadaan listrik dan gas 11 Kabupaten/Kota di Provinsi Jambi berdasarkan data kontribusi empat sektor tersebut dalam PDRB Tahun 2021 dengan data yang bersumber dari BPS Provinsi Jambi, seperti pada tabel 1.

**Tabel 1.** Potensi pertambangan, penggalian, pengadaan listrik dan gas di provinsi Jambi.

Kab/Kota	Potensi Pertambangan dan Penggalian	Persentase Kontribusi Pertambangan dan Penggalian dalam PDRB	Potensi Pengadaan Listrik dan Gas	Persentase Kontribusi Pengadaan Listrik dan Gas dalam PDRB
Kota Jambi	477,45 Miliar	1,50%	75,26 Miliar	0,24%
Kota Sungai Penuh	62,57 Miliar	0,78%	2,03 Miliar	0,03%
Kab Batang Hari	1874,44 Miliar	9,96%	10,60 Miliar	0,06%
Kab Tanjung Jabung Timur	8482,18 Miliar	38,76%	5,03 Miliar	0,02%
Kab. Tanjung Jabung Barat	12,26 Miliar	29,15%	10,92 Miliar	0,03%
Kab. Muaro Jambi	2320,61 Miliar	8,36%	12,82 Miliar	0,05%
Kab. Bungo	2765,57 Miliar	13,72%	12,51 Miliar	0,06%
Kab. Tebo	845,62 Miliar	4,99%	14,04 Miliar	0,08%
Kab. Merangin	439,13 Miliar	2,46%	12,17 Miliar	0,07%
Kab. Sarolangun	3208,33 Miliar	19,29%	5,63 Miliar	0,03%
Kab. Kerinci	198,82 Miliar	1,76%	4,54 Miliar	0,04%

Menentukan metode terbaik dari dua metode dalam pengelompokan salah satunya adalah dengan menggunakan Indeks Calinski – Harabasz (CH) yang disebut juga kriteria rasio varians dalam analisis klaser. Indeks Calinski – Harabasz (Calinski and Harabasz, 1974) ditentukan dengan menjumlahkan ukuran dispersi inter – cluster dan menjumlahkan ukuran disperse cluster keseluruhan. Dengan formulasi sebagai berikut:

$$CH = \frac{\frac{BGSS}{K-1}}{\frac{WGSS}{N-K}} = \frac{BGSS}{WGSS} \times \frac{N-K}{K-1}$$

Keterangan:

*BGSS: Between Group Dispersion Sum of Squares*

*WGSS: Within Group Dispersion Sum of Squares*

*N: Total Number of Observations*

*K: Total Number of Clusters*

Algoritma K-Means merupakan salah satu teknologi data mining yang menggunakan teknik klustering yang berulang-ulang. Algoritma ini dimulai dengan pemilihan secara acak K, yang merupakan banyaknya kluster yang ingin dibentuk. Kemudian K-Means merupakan salah satu metode clustering non hirarki yang berusaha mempartisi data yang ada ke dalam bentuk satu atau lebih cluster. Metode ini mempartisi data ke dalam cluster sehingga data yang memiliki karakteristik yang sama dikelompokkan ke dalam satu cluster yang sama dan data yang mempunyai karakteristik yang berbeda di kelompokkan ke dalam cluster yang lain. Sementara Fuzzy C – Means (FCM) diterapkan dalam penelitian Phukon, K.K., memodifikasi FCM dimana data yang akan dikelompokkan disajikan dalam graf. Secara spesifik algoritma klasik FCM disandingkan dengan graf, dengan harapan informasi yang diperoleh akan lebih banyak dibandingkan dengan FCM klasik.

### **Metode Penelitian**

Penelitian ini merupakan penelitian kuantitatif, menggunakan analisis statistika berupa perbandingan Analisis Cluster FCM dan K – Means menggunakan Indeks Calinski – Harabasz. Data berupa potensi pertambangan, penggalian, pengadaan listrik dan gas 11 Kabupaten/Kota di Provinsi Jambi, diolah menggunakan software JASP. Adapun Langkah – Langkah dalam penelitian ini sebagai berikut:

1. Pengumpulan dan entry Data.
2. Analisis Cluster dengan FCM.
3. Analisis Cluster dengan K – Means.
4. Perbandingan Indeks Calinski – Harabasz FCM dan K – Means.
5. Metode Terbaik berdasarkan Indeks CH.

### **Hasil dan Pembahasan**

Berikut adalah hasil analisis data menggunakan Software Open Source JASP: *Fuzzy C-Means Clustering* seperti pada tabel 2 – 5 dan gambar 1 dan 2.

**Tabel 2.** *Fuzzy C-Means Clustering*

<i>Clusters</i>	<i>N</i>	<i>R*</i>	<i>AIC</i>	<i>BIC</i>	<i>Silhouette</i>
2	11	0.540	41.420	45.400	0.620

**Tabel 3.** Cluster Information

Cluster	1	2
Size	1	10
Explained proportion within-cluster heterogeneity	0.000	1.000
Within sum of squares	0.000	21.421
Silhouette score	0.000	0.685
Centroid kontribusi_sektor1	-0.416	-0.011
Centroid kontribusi_sektor2	2.774	-0.284
Centroid persentase_sektor1	-0.677	0.013
Centroid persentase_sektor2	2.695	-0.273
Centroid transformasi_persentasesektor2	2.695	-0.273

Note. The Between Sum of Squares of the 2 Cluster Model is 25.1

Note. The Total Sum of Squares of the 2 Cluster Model is 46.52

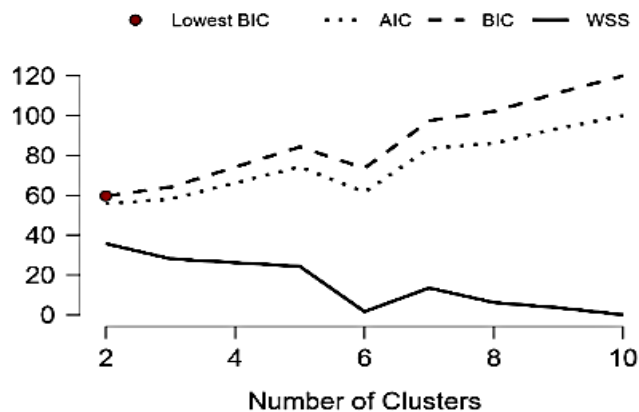
**Tabel 4.** Cluster Means

	Kontribusi Sektor1	Kontribusi Sektor 2	Persentase Sektor1	Persentase Sektor2
Cluster 1	-0.529	2.940	-0.793	2.866
Cluster 2	0.053	-0.294	0.079	-0.287

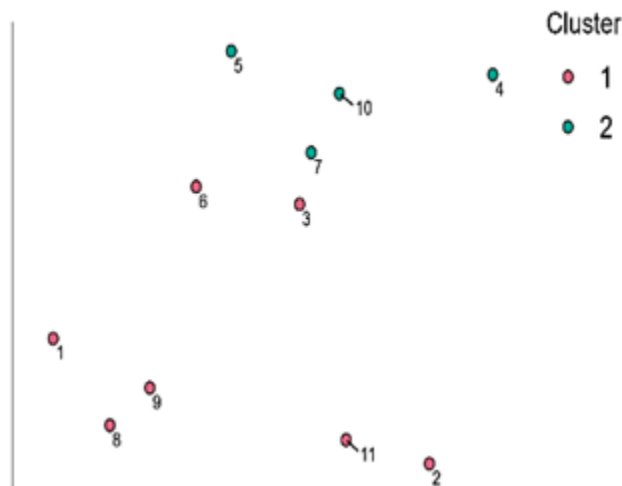
**Tabel 5.** Evaluation Metric

	Value
Maximum diameter	4.552
Minimum separation	4.937
Pearson's y	0.810
Dunn index	1.085
Entropy	0.305
Calinski-Harabasz index	12.008

Note. All metric are based on the euclidean distance



**Gambar 1.** Elbow Method Plot



Gambar 2. *t-SNE Cluster Plot*

Selanjutnya dilakukan Analisis K – Means Cluster Algoritma Hartigan Wong berbantuan software JASP, seperti pada tabel 6 – 10 dan gambar 3 dan 4.

Tabel 6. K-Means Clustering

Clusters	N	R <sup>2</sup>	AIC	BIC	Silhouette
3	11	0.803	31.880	36.650	0.530

Tabel 7. Cluster Information

Cluster	1	2	3
Size	9	1	1
Explained proportion within-cluster heterogeneity	1.000	0.000	0.000
Within sum of squares	7.879	0.000	0.000
Silhouette score	0.645	0.000	0.000
Centroid kontribusi_sektor1	-0.236	2.652	-0.529
Centroid kontribusi_sektor2	-0.284	-0.384	2.940
Centroid persentase_sektor1	-0.153	2.168	-0.793
Centroid persentase_sektor2	-0.238	2.168	2.866

Note. The Between Sum of Squares of the 3 cluster model is 32.12

Note. The Total Sum of Squares of the 3 cluster model is 40

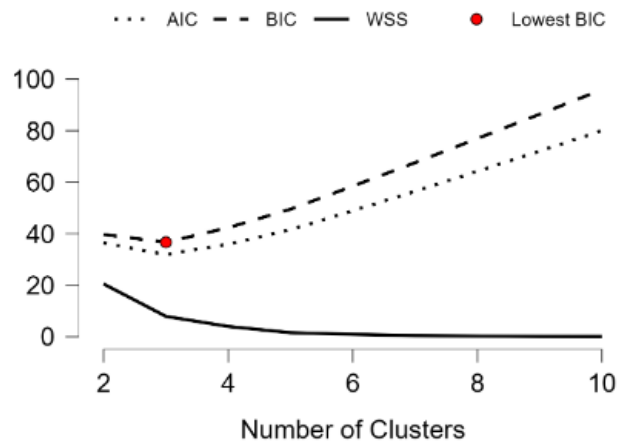
Tabel 8. Cluster Means

	Kontribusi Sektor1	Kontribusi Sektor 2	Persentase Sektor1	Persentase Sektor2
Cluster 1	-0.236	-0.284	-0.153	-0.238
Cluster 2	2.652	-0.384	2.168	-0.728
Cluster 3	-0.529	2.940	-0.793	2.866

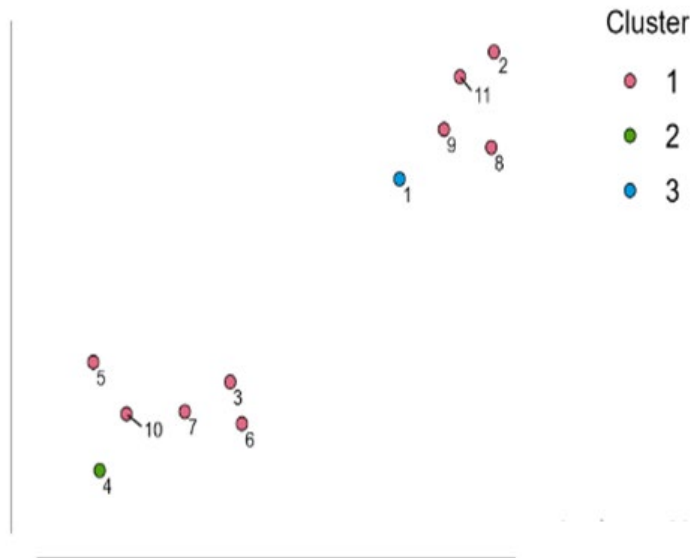
**Tabel 5.** Evaluation Metric

	<i>Value</i>
<i>Maximum diameter</i>	2.326
<i>Minimum separation</i>	2.642
<i>Pearson's y</i>	0.906
<i>Dunn index</i>	1.136
<i>Entropy</i>	0.600
<i>Calinski-Harabasz index</i>	16.308

*Note.* All metric are based on the euclidean distance



**Gambar 1.** Elbow Method Plot



**Gambar 4.** t-SNE Cluster Plot

Hasil analisis cluster menggunakan Fuzzy C – Means menghasilkan 2 clusters, cluster 1 adalah Kota Jambi, Kota Sungai Penuh, Kabupaten Batang Hari, Kabupaten Muaro Jambi,

Kabupaten Tebo, Kabupaten Merangin, dan Kabupaten Kerinci. Cluster 2 adalah Kabupaten Tanjung Jabung Timur, Kabupaten Bungo, Kabupaten Tanjung Jabung Barat, dan Sarolangun. Sedangkan hasil analisis cluster menggunakan K – Means membentuk 3 clusters. Cluster 1 terdiri dari Kota Sungai Penuh, Kabupaten Batang Hari, Kabupaten Tanjung Jabung Barat, Kabupaten Muaro Jambi, Kabupaten Bungo, Kabupaten Tebo, Kabupaten Merangin, Kabupaten Sarolangun, dan Kabupaten Kerinci. Cluster 2: Kabupaten Tanjung Jabung Timur, dan Cluster 3: Kota Jambi. Metode terbaik menurut Indeks Calinski – Harabasz adalah Analisis Cluster K – Means, dengan Indeks CH senilai 16,308, sedangkan Analisis Cluster FCM dengan Indeks CH senilai 12,008.

## **Kesimpulan**

Berdasarkan perbandingan Indeks Calinski – Harabasz, Analisis Cluster K – Means lebih baik dibandingkan dengan Analisis Cluster FCM, sehingga hasil pengelompokan yang digunakan adalah hasil Analisis cluster K – Means, yaitu Cluster 1 terdiri dari Kota Sungai Penuh, Kabupaten Batang Hari, Kabupaten Tanjung Jabung Barat, Kabupaten Muaro Jambi, Kabupaten Bungo, Kabupaten Tebo, Kabupaten Merangin, Kabupaten Sarolangun, dan Kabupaten Kerinci. Cluster 2: Kabupaten Tanjung Jabung Timur, dan Cluster 3: Kota Jambi.

## **Daftar Rujukan**

- [1] Anderson, T.W. 2003. An Introduction to Multivariate Statistical Analysis. Wiley Interscience.
- [2] Badan Pusat Statistik Provinsi Jambi, 2022, Provinsi Jambi dalam Angka 2022, Jambi: BPS Provinsi Jambi
- [3] Calinski, T., and Harabasz, J., 1974, A Dendrite Method for Cluster Analysis, Communications in Statistics Theory and Methods 3(1): 1 - 27
- [4] Gelley, N., and Roger, J., 2000, Fuzzy Logic Toolbox, Mathork, Inc., USA.
- [5] Hair, J.F., dkk. 2010. Multivariate Data Analysis, 7th Edition, Prentice- Hall.
- [6] J. Valente, W. Pedrycs, 2007, Advances in Fuzzy Clustering and its Applications, John Wiley & Sons, Ltd.
- [7] Johnson, R.A and Wichern, D.W. 2007. Applied Multivariate Statistical Analysis, 7th Edition, Prentice Hall.
- [8] K. K., Phukon, H. K. Baruah, Extension of The Fuzzy C Means Clustering Algorithm to Fit with The Composite Graph Model for Web Document Representation, IJCRSEE International Journal of Cognitive Research in Science, Engineering and Educations Vol. 1 No. 2, 2013.