
Metode K-Means dan DBSCAN pada Pengelompokan Data Dasar Kompetensi Laboratorium ITS Tahun 2017

Laila Qadrini

Universitas Sulawesi Barat

Jalan Prof. Dr. H. Baharuddin Lopa, S.H, Kelurahan Baurung,

Kecamatan Banggae Timur, Kabupaten Majene, Provinsi Sulawesi Barat 91412

e-mail: laila.qadrini@unsulbar.ac.id

ABSTRAK

Analisis Kelompok telah digunakan diberbagai bidang ilmu pengetahuan, dengan tujuan mengelompokkan objek/observasi. Hal penting dalam analisis kelompok adalah memperoleh nilai simpangan baku dalam kelompok yang minimum dan nilai simpangan baku antar kelompok yang maksimum, untuk memperoleh kelompok yang optimum didapatkan dengan menggunakan kriteria *Elbow* dan untuk menentukan validitas kelompok yaitu koefisien *Silhouette* dan Indeks Dunn. Pada penelitian ini dilakukan pengelompokan variabel dasar capaian laboratorium ITS dengan menggunakan beberapa analisis kelompok, yaitu Metode K-Means, dan DBSCAN. berdasarkan hasil analisis pengelompokan diperoleh Metode DBSCAN menghasilkan nilai koefisien *Silhouette* dan Indeks Dunn lebih besar dibandingkan dengan Metode K-Means. Dan Jumlah dari masing-masing Metode membentuk 4 kelompok.

Kata kunci : *K-Means, DBSCAN, Silhouette, Dunn, Laboratorium ITS*

ABSTRACT

Clusters analysis is significantly used to classify objects/observations in various fields of science. This strategy is essentially used to obtain the minimum and maximum standard deviation values between clusters, the optimum cluster obtained using the Elbow criteria, and to determine the cluster validity, namely the Silhouette coefficient and the Dunn Index. In this study, the basic ITS laboratory performance variables were analyzed using several cluster strategies, such as the K-Means and DBSCAN methods. Based on the clustering analysis results, the DBSCAN Method produced a greater value of the Silhouette coefficient and Dunn Index than the K-Means Method. Furthermore, the sum of each method forms 4 clusters.

Keywords : *K-Means, DBSCAN, Silhouette, Dunn, Laboratorium ITS*

1. PENDAHULUAN

Analisis kelompok merupakan analisis untuk menggabungkan obyek-obyek pengamatan dalam suatu kelompok, dimana obyek dalam kelompok memiliki varians homogen namun antar kelompok memiliki varians yang heterogen. Analisis kelompok adalah Metode pengelompokan kumpulan data kedalam kelompok yang menampilkan karakteristik serupa. Pengelompokan dilakukan berdasarkan similaritas atau disimilaritas (Elvira, 2019). Hal ini berarti bahwa klaster yang baik akan mempunyai homogenitas (kesamaan) yang tinggi antar anggota dalam satu klaster (*within-cluster*) dan heterogenitas (perbedaan) yang tinggi antar klaster yang satu dengan yang lainnya (*between-cluster*).

Prosedur pengelompokan terdiri dari dua yaitu Metode hierarki dan Metode non hierarki (Musfiani, 2019). Metode hierarki digunakan jika jumlah kelompok belum ditentukan sedangkan Metode non hierarki digunakan jika jumlah kelompok telah ditentukan. Pada Metode K-Means pengelompokan data non-hierarki (*partitioning*) dilakukan dengan cara mempartisi data ke dalam bentuk dua kelompok atau lebih, sehingga data yang berkarakteristik sama dimasukkan ke dalam satu kelompok yang sama. Sedangkan Metode *Density-Based Spatial Clustering Algorithm with Noise* (DBSCAN) merupakan suatu algoritma pengelompokan yang didasarkan pada kepadatan (*density*) data. Setiap Metode memiliki kelebihan dan kekurangannya

masing-masing. Hasil kelompok yang optimal dapat dipengaruhi oleh Metode pengelompokan yang digunakan, karakteristik dataset, struktur kepadatan data, ukuran data, jumlah kelompok yang digunakan. Metode *Elbow* digunakan untuk menentukan jumlah kelompok yang terbaik yang dapat digunakan untuk menghasilkan hasil pengelompokan yang terbaik dan dapat memaksimalkan kualitas hasil pengelompokan. (Dewa & Dewa, 2019). Penelitian yang dilakukan Bena dkk, adalah penelitian dengan tujuan untuk melakukan pengelompokan data penjualan *online* retail menggunakan Metode K-Means dan DBSCAN. Percobaan dilakukan menggunakan dataset sebanyak 500 data dan memiliki 3 atribut: deskripsi, kuantitas barang per transaksi dan harga barang per unit. Hasil percobaan pada penelitian ini menunjukkan, penerapan K-Means menghasilkan 3 kelompok dengan masing-masing berjumlah 103 anggota, 261 anggota dan 134 anggota. Sedangkan menggunakan DBSCAN dengan epsilon 1,005 dan minimum points 11 menghasilkan 3 kelompok dengan masing-masing berjumlah 30 anggota, 47 anggota dan 347 anggota, serta 74 data *noise*. (Bena *et al.*, 2019). Penelitian yang dilakukan oleh Dewa Ayu dan Dewa Ayu membandingkan Metode *Elbow* dan koefisien *Silhouette* untuk menentukan jumlah kelompok yang tepat sehingga menghasilkan kualitas kelompok yang optimal. Metode yang digunakan untuk menguji hasil kelompok adalah Metode *Davies Bouldin Index* (DBI). Hasil pengujian pengelompokan dengan Metode *Elbow* menggunakan nilai DBI menghasilkan nilai DBI sebesar 1,10. Sedangkan pada uji coba pengelompokan dengan koefisien *Silhouette* menghasilkan nilai DBI sebesar 1,06. Hal ini menunjukkan bahwa hasil pengelompokan k-medoid dengan koefisien *Silhouette* menghasilkan kualitas kelompok lebih baik karena memiliki nilai DBI lebih rendah daripada pengelompokan k-medoid dengan Metode *Elbow* (Dewa & Dewa, 2019). Selanjutnya Penelitian oleh Ninik Tri Hartanti, yaitu Penerapan algoritma K-Means untuk membentuk kelompok-kelompok siswa berdasarkan nilai rata-rata ketiga mata pelajaran UN (matematika, Bahasa Indonesia, dan Bahasa Inggris). Penentuan jumlah optimal

kelompok menggunakan Metode *Elbow*, yang menunjukkan banyaknya kelompok optimal adalah 3, kemudian dihasilkan 3 kelompok yang terdiri dari cluster kategori “Siap”, “Cukup Siap” dan Tidak Siap” (Ninik, 2020). Belum ada penelitian yang memadukan Metode K-Means dan DBSCAN dengan menggunakan Metode *Elbow* untuk menentukan Jumlah Kelompok yang optimal. Penelitian ini menggunakan kedua Metode tersebut karena data penelitian berukuran cukup besar dan memiliki tipe variabel kontinu. Berdasarkan kelebihan dari Metode-Metode tersebut maka tepat digunakan untuk mencari pengelompokan terbaik pada kasus dasar Laboratorium ITS.

2. METODE PENELITIAN

Dalam kelompok ada banyak Metode yang bisa digunakan dan pada penelitian ini menggunakan 2 Metode yaitu sebagai berikut:

2.1. K-Means

K-Means merupakan Metode pengelompokan yang paling terkenal dan banyak digunakan diberbagai bidang karena sederhana, mudah diimplementasikan, memiliki kemampuan untuk mengelompok data yang besar, mampu menangani data outlier, dan kompleksitas waktunya linear $O(nKT)$ dengan n adalah jumlah dokumen, K adalah jumlah kelompok, dan T adalah jumlah iterasi. K-Means merupakan Metode pengkelompokan secara *partitioning* yang memisahkan data ke dalam kelompok yang berbeda. Dengan *partitioning* secara iteratif, K-Means mampu meminimalkan rata-rata jarak setiap data ke kelompoknya.. K-Means ditemukan oleh MacQueen pada Tahun 1967. Menurut Dean dalam Yohannes (2016) K-Means adalah salah satu teknik pengelompokan yang paling banyak digunakan karena kesederhanaan dan kecepatan. Algoritma ini mempartisi data ke dalam k kelompok dengan menetapkan setiap objek kelompok *centroid* terdekatnya (nilai rata-rata dari variabel untuk semua objek dalam kelompok tertentu) berdasarkan ukuran jarak yang digunakan. Hal ini lebih kuat untuk berbagai jenis variabel.

Langkah-langkah algoritma K-Means :

1. Tentukan jumlah kelompok k
2. Pilih pusat kelompok k (pemilihan objek k dari data set dilakukan secara random)

3. Hitung jarak setiap objek terhadap pusat kelompok.
4. Hitung kembali jarak terdekat setiap objek terhadap pusat kelompok.
5. Ulangi langkah 3 dan 4 sampai (selisih perhitungan jarak) atau iterasi maksimum tercapai.

2.2. DBSCAN

DBSCAN adalah algoritma pengelompokan yang didasarkan pada kepadatan (*density*) data (Suyanto, 2017). Konsep kepadatan yang dimaksud dalam DBSCAN adalah jumlah data yang berada dalam radius MinPts (Jumlah minimal data dalam radius ϵ), data tersebut masuk dalam kategori kepadatan yang diinginkan, jumlah data dalam radius tersebut termasuk data inti itu sendiri. Konsep kepadatan seperti ini melahirkan tiga macam status dari setiap data, yaitu inti (*core*), batas (*border*), dan *noise*. Sebuah data akan dimasukkan sebagai inti jika jumlah data tetangga dan dirinya sendiri pada radius ϵ berjumlah \geq MinPts. Nilai radius ϵ dan MinPts ini ditetapkan secara mandiri. Untuk data yang jumlah tetangga dan dirinya sendiri dalam radius ϵ kurang dari MinPts, tetapi tetangganya menjadi inti karena kehadirannya, data tersebut dikategorikan sebagai batas. Jika jumlah tetangga dan dirinya sendiri dalam radius ϵ kurang dari MinPts dan tidak ada tetangga yang menjadi inti karena kehadirannya, data tersebut dikategorikan sebagai *noise* (Diah *et al.*, 2017). Mendefinisikan kelompok sebagai satu set poin dengan kepadatan maksimal yang terhubung. DBSCAN menentukan sendiri jumlah kelompok yang akan dihasilkan, tapi memerlukan 2 input lain, yaitu:

1. MinPts : minimal banyak items dalam suatu kelompok
2. Eps : nilai untuk jarak antar-items yang menjadi dasar pembentukan *neighborhood* dari suatu titik item.

Beberapa istilah mengenai DBSCAN:

1. Lingkungan dalam radius Eps suatu objek tertentu disebut objek *e-neighborhood*.
2. Jika *e-neighborhood* dari sebuah objek berisi setidaknya jumlah minimal, MinPts, objek, maka objek disebut objek inti (*core object*).
3. Ada 2 jenis titik (*points*) dalam suatu kelompok: titik tengah di dalam kelompok

(*core points*) dan di tepian kelompok (*border points*).

4. *Neighborhood* dari *border points* berisi jauh lebih sedikit items daripada *neighborhood* dari *core points*. Suatu *border point* bisa jadi termasuk ke dalam lebih dari 1 *core object*. (Qadrini, 2020).

2.3. Metode Elbow

Metode *Elbow* merupakan salah satu Metode untuk menentukan jumlah kelompok yang tepat melalui persentase hasil perbandingan antara jumlah kelompok yang akan membentuk siku pada suatu titik (Dewa & Dewa, 2019). Jika nilai kelompok pertama dengan nilai kelompok kedua memberikan sudut dalam grafik atau nilainya mengalami penurunan paling besar maka jumlah nilai kelompok tersebut yang tepat. Untuk mendapatkan perbandingannya adalah dengan menghitung *Sum of Square Error* (SSE) dari masing-masing nilai kelompok. Karena semakin besar jumlah nilai kelompok k , maka nilai SSE akan semakin kecil. Rumus SSE sesuai dengan Persamaan 1.

$$SSE = \sum_{k=1}^k \sum_{x_i} [x_i - c_k]^2 \quad (1)$$

- k : Kelompok ke- c
 x_i : Jarak data objek ke- i
 c_k : Pusat kelompok ke- i

2.4. Validitas Kelompok

Validasi kelompok adalah prosedur yang mengevaluasi hasil analisis kelompok secara kuantitatif dan objektif. Terdapat tiga pendekatan untuk mengeksplorasi validitas kelompok yaitu validasi eksternal, internal dan relatif. Dalam penelitian ini akan digunakan validasi internal, yaitu dengan menggunakan Metode *Silhouette* dan Indeks Dunn (Jain & Dubes, 1988).

2.4.1 Indeks Dunn

Indeks Dunn adalah rasio jarak terkecil antara observasi pada kelompok yang berbeda dengan jarak terbesar pada masing-masing cluster data. Indeks Dunn diperoleh dari hasil pembagian antara d_{min} dengan d_{max} .

$$C = \frac{d_{min}}{d_{max}} \quad (2)$$

d_{min} : Jarak terkecil antara observasi pada kelompok yang berbeda

d_{max} : Jarak terbesar pada masing-masing kelompok data

Nilai DI yang semakin besar menandakan hasil pengelompokan yang semakin baik. (Nurul & Edy,2017).

2.4.2 Indeks Silhouette

Indeks *Silhouette* merupakan gabungan dari Metode separasi dan kohesi. (Kodinariva & Makwana, 2013). Untuk menghitung nilai *silhouette coefficient* diperlukan jarak antar objek dengan menggunakan rumus *Euclidean Distance*. Setelah itu tahapan untuk menghitung nilai *Silhouette Coefficient* adalah sebagai berikut :

1. Untuk setiap objek i , hitung rata-rata jarak dari objek i dengan seluruh objek yang berada dalam satu kelompok. Akan didapatkan nilai rata-rata yang disebut a_i .
2. Untuk setiap objek i , hitung rata-rata jarak dari objek i dengan objek yang berada di kelompok lainnya. Dari semua jarak rata-rata tersebut ambil nilai yang paling kecil. Nilai ini disebut b_i .
3. Setelah itu maka untuk objek i memiliki nilai *Silhouette Coefficient* :

$$S_i = \frac{(b_i - a_i)}{\max(a_i, b_i)} \tag{3}$$

hasil perhitungan nilai *Silhouette Coefficient* dapat bervariasi antara -1 hingga 1. Hasil pengelompokan dikatakan baik jika nilai *Silhouette Coefficient* bernilai positif ($a_i < b_i$) dan a_i mendekati 0, sehingga akan menghasilkan nilai *Silhouette Coefficient* yang maksimum yaitu 1 saat $a_i = 0$. Maka dapat dikatakan, jika $S_i = 1$ berarti objek i sudah berada dalam kelompok yang tepat. Jika nilai $S_i = 0$ maka objek i berada di antara dua kelompok sehingga objek tersebut tidak jelas harus dimasukkan ke dalam kelompok A atau kelompok B. Akan tetapi, jika $S_i = -1$ artinya struktur kelompok yang dihasilkan *overlapping*, sehingga objek i lebih tepat dimasukkan ke dalam kelompok yang lain. Nilai rata-rata *Silhouette Coefficient* dari tiap objek dalam suatu kelompok adalah suatu ukuran yang menunjukkan seberapa ketat data dikelompokkan dalam kelompok tersebut. Berikut adalah nilai *Silhouette Coefficient* berdasarkan (Kaufman & Rousseeuw, 1990) adalah:

Tabel 1. Nilai *Coeffisien Silhouette*

Nilai SC	Keterangan
$0,7 < SC \leq 1$	Struktur Kuat
$0,5 < SC \leq 0,7$	Struktur Medium
$0,25 < SC \leq 0,5$	Struktur Lemah
$SC \leq 0,25$	Tidak ada Struktur

2.5 Teknik Pengolahan Data

Teknik pengolahan data dalam penelitian ini adalah pengelompokan dengan Metode K-Means dan Metode DBSCAN menggunakan software R Statistic 3.6.3. Tahap analisis data yang digunakan dalam penelitian ini sebagai berikut :

1. Melakukan *preprocessing* data
2. Menentukan Jumlah Kelompok yang optimal
3. Melakukan pengelompokan data menggunakan Metode K-Means
4. Melakukan pengelompokan data menggunakan Metode DBSCAN
5. Pengujian validitas kelompok menggunakan uji Koefisien *silhouette* dan Indeks Dunn pada kelompok yang telah dibentuk menggunakan Metode K-Means dan DBSCAN

3. HASIL DAN PEMBAHASAN

3.1 Sumber Data

Sumber data yang digunakan merupakan data Dasar Capaian Laboratorium Semester 1 Instituit Teknologi Sepuluh Nopember Surabaya Tahun 2017. Data yang digunakan adalah Laboratorium Dasar FTSP, FTIF, FBMT, Vokasi, dan FTI. Metode yang digunakan dalam penelitian ini adalah metode pengelompokan K-Means, dan DBSCAN. Untuk kriteria kebaikan pengelompokan yang diterapkan adalah *Silhouette Coefficient*. Adapun persamaan untuk menghitung kebaikan model tersebut adalah sebagai berikut:

$$a(0) = \frac{\sum_{o' \in C_i, o \neq o'} \text{dist}(o, o')}{|C_i| - 1} \tag{4}$$

Dan

$$b(0) = \min_{C_j: 1 \leq j \leq k, j \neq i} \frac{\sum_{o' \in C_j, o \neq o'} \text{dist}(o, o')}{|C_j|} \tag{5}$$

Untuk mengukur kualitas kelompok dalam suatu pengelompokan, kita dapat menghitung rata-rata *Silhouette Coefficient* dari semua objek dalam kelompok tersebut. Sementara itu, untuk mengukur kualitas suatu pengelompokan, kita dapat menggunakan rata-rata *silhouette coefficient* dari semua objek dalam kelompok.

$$s(o) = \frac{b(o)-a(o)}{\max\{a(o),b(o)\}} \quad (6)$$

3.2. Analisis Deskriptif

Ada beberapa Metode yang digunakan pada penelitian ini, tetapi sebelum membahas Metode yang digunakan terlebih dahulu akan diteliti untuk karakteristik data yang digunakan dalam penelitian ini menggunakan analisis deskriptif. Penjelasan tentang karakteristik data pada penelitian ini dapat dilihat pada tabel berikut ini :

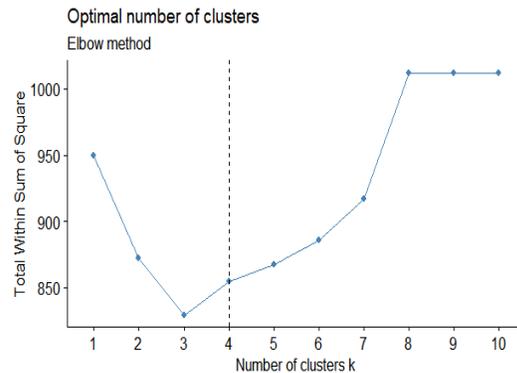
Tabel 2. Descriptive Statistics

	N	Min	Max	Mean	Std. Deviation
X1	93	0	33	7.05	7.030
X2	93	0	14	4.75	3.031
X3	93	0	11	3.46	2.362
X4	93	0	7	.61	1.344
X5	93	0	9	1.04	1.841
X6	93	0	29	2.13	3.987
X7	93	0	11	1.18	2.131
X8	93	0	1569	297.91	378.832
X9	93	0	7	.70	1.397
X10	93	0	30	2.38	4.046
X11	93	0	55	4.34	8.513

Jumlah data yang diteliti sebanyak 94 (Lab) dengan nilai terkecil adalah 0 terdapat pada beberapa variabel yang menjelaskan bahwa pada setiap Lab mempunyai penelitian yang tidak dipublikasikan. Nilai maksimum dan rata-rata terbesar terdapat pada variabel X8 (Jumlah Sitasi Dosen dalam *Google Scholar*) sebesar 1569 pada *Logistics and Supply Chain*

Management Lab. dan rata-rata sebesar 297,91 dan nilai standar deviasi sebesar 378,832.

3.3 Menentukan Jumlah Kelompok yang Optimal



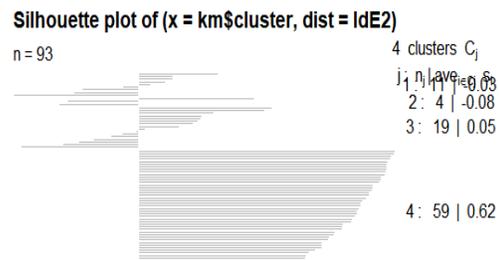
Gambar 1. Jumlah Kelompok Optimal

Berdasarkan Gambar 1 sesuai *output* pada *Software R 3.5.5*, diperoleh keterangan sesuai hasil analisis Metode *Elbow* untuk pengelompokan K-Means dan DBSCAN, jumlah kelompok optimal untuk mengelompokkan kompetensi laboratorium ini adalah 4 kelompok

3.4 Metode Pengelompokan Terbaik

3.4.1 Hasil Pengelompokan K-Means

Berdasarkan hasil analisis dari metode *Elbow*, diperoleh bahwa jumlah kelompok yang terbentuk adalah 4 dengan nilai *Average Silhouette* sebesar 0,4, sesuai Tabel 1 nilai *Silhouette Coefficient* ≤ 0,5 yang berarti termasuk struktur lemah, atau objek-objek pada penelitian ini yaitu dasar capaian kompetensi Laboratorium belum berada pada kelompok yang tepat, hal ini mengindikasikan bahwa Metode K-Means kurang tepat diterapkan pada penelitian ini. Adapun plot dari *silhouette* untuk K-Means adalah dibawah ini :

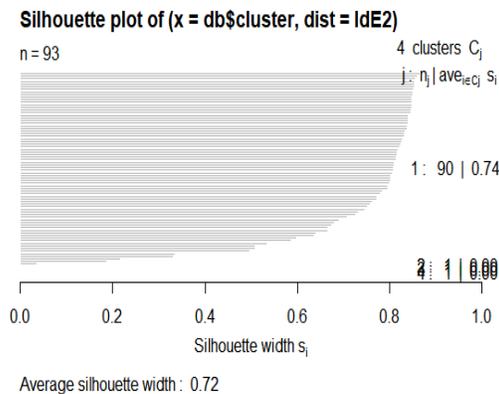


nilai *coefficient silhouette* terbesar dan nilai Indeks Dunn yang mana lebih besar daripada nilai koefisien *Silhouette* dan Indeks Dunn Metode K-Means. Hal ini mengidentifikasi bahwa kelompok berdasarkan kepadatan objek yang diamati.

Gambar 2. Plot *Silhouette* K-Means

3.4 Hasil Pengelompokan DBSCAN

Berdasarkan hasil analisis dari metode *Elbow* diperoleh bahwa jumlah kelompok yang terbentuk adalah 4 dengan nilai *Average silhouette* sebesar 0,72 yang berarti termasuk struktur kuat, menurut (Kaufman dan Rousseeuw, 1990) nilai *Average silhouette* yang mendekati 1 menunjukkan hasil pengelompokan yang baik. Atau objek-objek yang pada penelitian ini adalah dasar capaian kompetensi Laboratorium telah berada pada kelompok yang tepat. Sehingga metode DBSCAN lebih tepat diterapkan dibandingkan metode K-Means, adapun plot dari *silhouette* DBSCAN adalah dibawah ini :



Gambar 3. Plot *Silhouette* DBSCAN

Berdasarkan Tabel 3 dapat diketahui Metode kelompok terbaik pada data Dasar Capaian Laboratorium Semester I Institut Teknologi Sepuluh Nopember Surabaya Tahun 2017 menggunakan Metode DBSCAN berdasarkan

Tabel 3. Perbandingan Indeks *Silhouette* dan Indeks Dunn

Metode	Indeks <i>Silhouette</i>	Indeks Dunn
K-Means	0,40	0,12
DBSCAN	0.72	0,57

IV. KESIMPULAN

Berdasarkan analisis data dan pembahasan tentang Dasar Capaian Laboratorium Semester I Institut Teknologi Sepuluh Nopember Surabaya Tahun 2017, dapat diambil kesimpulan sebagai berikut:

1. Hasil analisis Metode *Elbow* untuk pengelompokan K-Means dan DBSCAN menghasilkan jumlah kelompok yang optimal adalah 4 kelompok
2. Metode Pengelompokan terbaik adalah Metode DBSCAN dengan nilai *Silhouette Coefficient* sebesar 0,72 dengan $\epsilon = 5.5$ dan $Minpts = 1$.

DAFTAR PUSTAKA

Ashari, S. B., Otniel, C. S., & Rianto. 2019. *Perbandingan Kinerja K-Means dengan DBSCAN untuk Metode Clustering Data Penjualan Online Retail*. Jurnal Siliwangi Vol.5. No.2.

Dewi A.I.C.D., & Pramita A.K.D. 2019. *Analisis Perbandingan Metode Elbow dan Silhouette pada Algoritma Clustering K-Medoids dalam Pengelompokan Produksi Kerajinan Bali (Jurnal Matrix)*, Vol. 9, No. 3, November 2019.

- Halim N. N., & Widodo, E. 2017. *Clustering Dampak Gempa Bumi di Indonesia Menggunakan Kohonen Self Organizing Maps*, Prosiding SI MaNis (Seminar Nasional Integrasi Matematika dan Nilai Islami) Vol.1, No.1, pp 188-194.
- Hartanti, T. N. 2020. *Metode Elbow dan K-Means Guna Mengukur Kesiapan Siswa SMK Dalam Ujian*. Jurnal Nasional Teknologi dan Sistem Informasi. Vol. 06 NO. 02. pp 82-89.
- Hermanto, P.M., Elvira, 2019. *Pengelompokan Kabupaten Kota Di Provinsi Jawa Timur Berdasarkan Indikator Kesehatan Mortalitas Dengan Algoritma Variable Weighting K-Means*. J Statistika, Vol 12, No. 1, pp 20-25.
- Jain, & Dubes. 1988. *Algorithm for Clustering Data*. https://homepages.inf.ed.ac.uk/rbf/BOOKS/JAIN/Clustering_Jain_Dubes.pdf, diakses tanggal 27 November 2020.
- Kaufman, L., & Rousseuw, P.J. 1990. *Finding Groups in Data: An Introduction to Cluster Analysis*. New York. Wiley.
- Musfiani. 2019. *Analisis Cluster Dengan Menggunakan Metode Partisi Pada Pengguna Alat Kontrasepsi Di Kalimantan Barat*. Buletin Ilmiah Math. Stat. dan Terapannya (Bimaster). Vol 8. No. 4. pp 893-902.
- Safitri, D., Wuryandari, T., & Rahmawati, R. 2017. *Metode DBSCAN untuk Pengelompokan Kabupaten/Kota Di Provinsi Jawa Tengah Berdasarkan Produksi Padi Sawah Dan Padi Ladang*. Statistika, Vol. 5, No. 1.
- Suyanto. 2017. *Data Mining Untuk Klasifikasi dan Klasterisasi Data*. Bandung: Informatika Bandung
- T. M. Kodinariya, & P. R. Makwana, 2013. *Review on determining number of cluster in K- Means Clustering*, Int. J. Adv. Res. Comput. Sci. Manag. Stud. Vol. 1. No. 6.
- Yohannes. 2016. *Analisis Perbandingan Algoritma Fuzzy C-Means dan K-Means*. Prosiding Annual Research. Vol 2. No. 1.
- Qadrini, L. (2020). Ensemble Fuzzy, Ensemble Rock Pada Pengelompokan Pelamar Bidikmisi Sejava Timur Tahun 2017. *Jurnal MSA (Matematika Dan Statistika Serta Aplikasinya)*, 8(1), 46–50. <http://journal.uin-alauddin.ac.id/index.php/msa/article/view/JMSA.VOL8N1046/pdf>