



ANALISA PERBANDINGAN K-MEANS DAN FUZZY C-MEANS DALAM PENGELOMPOKAN DAERAH PENYEBARAN COVID-19 INDONESIA

Aina Latifa Riyana Putri ✉, Nurkaromah Dwidayati

Jurusan Matematika, FMIPA, Universitas Negeri Semarang, Indonesia

Gedung D7 Lt.1, Kampus Sekaran Gunungpati, Semarang, 50229

Info Artikel

Sejarah Artikel:

Diterima September 2021

Disetujui November 2021

Dipublikasikan November 2021

Keywords:

k-Means Algorithm, Fuzzy c-Mean

Davies Bouldin-Index

Abstrak

Penelitian ini bertujuan untuk mengetahui algoritma terbaik dalam pengelompokan daerah penyebaran Covid-19 per provinsi di Indonesia yang mana dilakukan berdasarkan fakta dimana saat ini Indonesia diguncangkan oleh mewabahnya Covid-19. Metode analisis data melakukan perbandingan antara algoritma k-Means dan Fuzzy c-Means dengan uji validitas cluster menggunakan Dunn-Index dan Davies Bouldin-Index untuk memperoleh hasil cluster optimal berbantuan Rstudio. Hasil penelitian menunjukkan bahwa pengujian clustering k-Means menghasilkan nilai akurasi yang lebih besar sebesar 1,165219 dibandingkan Fuzzy c-Means. Sehingga clustering k-Means diambil untuk menentukan pengelompokan daerah penyebaran Covid-19 per provinsi Indonesia. Diperoleh 4 cluster optimal berdasarkan beberapa data kasus Covid-19 dari 15 Maret 2020 - 30 Juli 2021 menggunakan metode Elbow yang terbentuk dengan algoritma k-Means yaitu cluster yang berpotensi sangat tinggi dalam penyebaran kasus Covid-19 berisi 2 provinsi, cluster yang berpotensi tinggi dalam penyebaran kasus Covid-19 berisi 22 provinsi, cluster yang berpotensi sedang dalam penyebaran kasus Covid-19 berisi 8 provinsi, dan cluster yang berpotensi rendah dalam penyebaran kasus Covid-19 berisi 2 provinsi. Saran yang diberikan sebaiknya pemerintah lebih menertibkan lockdown hingga giat edukasi perihal vaksin sebagai alternatif cara untuk menekan kasus Covid-19.

Abstract

This study aims to find out the best algorithm for grouping Covid-19 distribution areas per province in Indonesia, which is based on the fact that Indonesia is currently being shaken by the Covid-19 outbreak. The data analysis method compares the k-Means and Fuzzy c-Means algorithms with cluster validity tests using Dunn-Index dan Davies Bouldin-Index to obtain optimal cluster results assisted by Rstudio. The results showed that the k-Means clustering test resulted in a greater accuracy value of 1,165219 than the Fuzzy c-Means. So, the k-Means clustering was taken to determine the grouping of the Covid-19 distribution areas per Indonesian province. Four optimal clusters were obtained using the Elbow method which was formed with the k-Means algorithm, namely a cluster with a very high potential in spreading Covid-19 cases containing 2 provinces, a cluster with a high potential for spreading Covid-19 cases containing 22 provinces, a cluster with a potential to be spread Covid-19 cases contain 8 provinces, and clusters with low potential in spreading Covid-19 cases contain 22 provinces. The advice given is that the government should put more control over the lockdown to be active in educating about vaccines as an alternative way to suppress Covid-19 cases.

How to cite:

Putri, A. L. R., Dwidayati, N. K., (2021). Analisa Perbandingan K-Means dan Fuzzy C-Means dalam Pengelompokan Daerah Penyebaran Covid-19 Indonesia. Unnes Journal of Mathematics, 7(2), 50-55.

© 2021 Universitas Negeri Semarang

✉Alamat korespondensi:

E-mail: ainalatif47@gmail.com

p-ISSN 2252-6943

e-ISSN 2460-5859

PENDAHULUAN

Sepanjang peradaban manusia, penyakit menular telah menjadi penyebab utama kematian jutaan manusia. Seperti pada tahun 1320 terjadi wabah PES di China, wabah flu Spanyol pada tahun 1918-1920 dengan menewaskan 50 juta hingga 100 juta orang di seluruh dunia. Seperti saat ini, dunia tengah diguncangkan oleh mewabahnya Covid-19 yang disebabkan oleh SARS-CoV-2 yang merupakan satu kelompok virus yang sama dengan virus yang menyebabkan penyakit SARS. Keluhan yang diderita dapat mengalami demam, batuk kering, kesulitan bernapas bahkan dapat berkembang menjadi pneumonia berat, kegagalan multiorgan bahkan kematian. Berdasarkan informasi yang diperoleh dari <https://covid19.go.id/> hingga 30 Juli 2021 telah terjadi 3.372.374 kasus terkonfirmasi penyakit Covid-19 di Indonesia dengan menewaskan 92.311 korban jiwa.

Walaupun program vaksinasi Covid-19 telah dilakukan di Indonesia, berbagai program tetap dicanangkan pemerintah baik pusat maupun daerah di Indonesia untuk membantu mengurangi penyebaran penyakit Covid-19 seperti meminta warga untuk mempraktikkan aktivitas jaga jarak hingga berdiam diri di rumah. Banyaknya program yang dicanangkan tentu saja perlu didukung oleh strategi perencanaan yang matang. Strategi perencanaan program bisa saja berbeda untuk tiap provinsi (Anuraga,2015), biasanya pemerintah provinsi menyesuaikan dengan pola pandemi yang dihadapi. Melihat pandemi Covid-19 yang sampai saat ini belum menunjukkan tanda-tanda penurunan di beberapa provinsi, menumbuhkan keinginan untuk memiliki gagasan menentukan pengelompokan daerah penyebaran Covid-19 pada 34 provinsi Indonesia dengan membandingkan clustering menggunakan algoritma k-Means dan Fuzzy c-Means berbantuan software Rstudio. Diharapkan hasil yang diperoleh ini nantinya dapat menjadi masukan kepada pemerintah dalam menangani penyebaran Covid-19 dengan lebih optimal.

Adapun variabel yang digunakan pada penelitian ini adalah 6 indikator kesehatan masyarakat seperti rata-rata persentase penurunan kasus positif, kasus meninggal, kasus sembuh, laju insidensi, angka kematian, hingga cakupan vaksin yang telah dilakukan masyarakat. K-Means Cluster merupakan salah satu metode cluster analysis non hierarki yang mengelompokkan n buah objek ke dalam k kelas berdasarkan jaraknya dengan pusat

kelas. Dalam melakukan clustering selain menggunakan algoritma k-Means, dapat juga menggunakan algoritma Fuzzy c-Means. Algoritma Fuzzy c-Means merupakan suatu teknik clustering yang banyak digunakan dalam aplikasi clustering (Simbolon,2013) dengan pendekatan fuzzy dan setiap data memiliki derajat keanggotaan untuk tiap cluster. Adapun parameter pembanding algoritma adalah dengan melakukan uji validitas cluster menggunakan metode Dunn-Index dengan menghitung nilai minimum dari perbandingan antara nilai fungsi dissimilaritas antara dua kluster sebagai separation dan nilai maksimum dari diameter kluster sebagai compactness untuk mencari kombinasi algoritma Fuzzy c-Means terbaik dan metode Davies Bouldin Index dengan menghitung rata-rata nilai setiap titik pada himpunan data untuk membandingkan algoritma k-Means dan Fuzzy c-Means yang diperoleh.

METODE

Data yang digunakan pada penelitian ini adalah data sekunder yaitu Data Kasus Positif Covid-19 Harian per provinsi di Indonesia 15 Maret 2020-30 Juli 2021, Data Kasus Meninggal Akibat Covid-19 Harian per provinsi di Indonesia dari 21 Maret 2020-30 Juli 2021, Data Sembuh Harian per provinsi di Indonesia 21 Maret 2020-30 Juli 2021 diperoleh dari website KawalCOVID19 (<https://kawalcovid19.id/>), Data Jumlah Penduduk per provinsi di Indonesia diperoleh dari website resmi Kementerian Dalam Negeri Indonesia (<https://www.kemendagri.go.id/>), Data cakupan vaksinasi Covid-19 hingga dosis kedua per provinsi diperoleh dari website resmi Kementerian Kesehatan Indonesia (<https://www.kemkes.go.id/>).

Penelitian ini akan dilaksanakan berdasarkan rancangan penelitian sebagai berikut.



Gambar 1. Alur Penelitian

Metode analisis data melakukan perbandingan antara algoritma k-Means dan Fuzzy c-Means dengan uji validitas cluster menggunakan Dunn-Index dan Davies

Bouldin-Index untuk memperoleh hasil cluster optimal memanfaatkan software Rstudio dengan 6 variabel indikator kesehatan masyarakat.

Analisis cluster adalah teknik multivariat untuk mengelompokkan individu atau objek menjadi beberapa kelompok tertentu dimana setiap objek yang berada dalam cluster yang sama mempunyai kemiripan satu dengan yang lain dibandingkan dengan anggota cluster yang lain (Pitaloka, 2018). Clustering merupakan metode segmentasi data yang sangat berguna dalam prediksi dan analisa masalah bisnis tertentu. Misalnya Segmentasi pasar, marketing dan pemetaan zonasi wilayah (Irwansyah,2015). K-Means Clustering adalah suatu metode penganalisaan data atau metode data mining yang melakukan proses pemodelan tanpa supervise (Nasari,2013). Sedangkan Algoritma Fuzzy C-Means merupakan suatu teknik clustering yang menerapkan pengelompokan fuzzy dimana setiap data dapat menjadi anggota dari beberapa cluster dengan derajat keanggotaan yang berbeda tiap clusternya.

HASIL DAN PEMBAHASAN

Lima data Covid-19 per provinsi yang telah disebutkan pada bagian sebelumnya, akan dihitung kemudian dikelompokkan menjadi data 6 indikator epidemiologi yang telah ditetapkan pemerintah sebagai indikator zona Covid-19 pada 34 provinsi di Indonesia.

Tabel 1. Indikator

x_i = Indikator	
x_1	= Rata-rata persentase penurunan jumlah kasus positif selama 2 minggu terakhir dari puncak kasus positif terjadi di setiap provinsi
x_2	= Rata-rata persentase penurunan jumlah meninggal dari kasus positif selama 2 minggu terakhir dari puncak kasus meninggal akibat Covid-19 terjadi di setiap provinsi
x_3	= Rata-rata persentase kenaikan jumlah sembuh dari kasus positif
x_4	= Rata-rata persentase penurunan laju insidensi kasus positif.
x_5	= Rata-rata persentase penurunan angka kematian
x_6	= Cakupan vaksinasi dosis kedua

Pada Tabel 1 dapat terlihat indikator yang digunakan pada penelitian ini, Dengan $x_1, x_2, x_3, x_4, x_5, x_6$ merupakan variabel penelitian yang akan dijelaskan lebih lanjut pada Tabel 3.

Tabel 2. Objek Pengamatan

i = Objek Pengamatan					
i	Provinsi	i	Provinsi	i	Provinsi
1	Aceh	13	Kaltim	25	Sulteng
2	Bali	14	Kalteng	26	Lampung
3	Banten	15	Kalsel	27	Riau
4	Babel	16	Kalut	28	Malut
5	Bengkulu	17	Kep. Riau	29	Maluku
6	Jogjakarta	18	NTB	30	Papbar
7	Jakarta	19	Sumsel	31	Papua
8	Jambi	20	Sumbar	32	Sulbar
9	Jabar	21	Sulut	33	NTT
10	Jateng	22	Kaltim	34	Gorontalo
11	Jatim	23	Sultra		
12	Kalbar	24	Sulsel		

Pada Tabel 2 dapat terlihat Objek Pengamatan yaitu 34 Provinsi di Indonesia yang digunakan pada penelitian ini untuk dilakukan pengelompokan wilayah Covid-19.

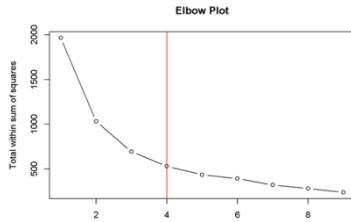
Tabel 3. Data Indikator

i	Variabel Penelitian					
	X_1	X_2	X_3	X_4	X_5	X_6
1	-0,014	0,033	0,530	-0,405	0,002	0,066
2	-0,053	-0,054	0,290	-0,205	0,003	0,242
3	0,005	-0,107	0,870	-0,533	-0,0009	0,084
....
33	-0,227	-0,370	1,788	-0,870	0,0001	0,073
34	-0,627	0,196	0,595	-0,732	0,0012	0,1007

Pada Tabel 3 merupakan nilai indikator pada masing-masing provinsi. Sebelumnya dilakukan Standarisasi data terlebih dahulu agar data memiliki rentang nilai (scale) yang sama.

1. Algoritma k-Means

Langkah pertama untuk melakukan clustering k-Means pada Rstudio adalah dengan membentuk matriks jarak dimana dipilih metode Euclidean Distance. Setelah itu, akan dicoba k optimal untuk digunakan clustering menggunakan metode Elbow.



Gambar 2. Jumlah Cluster Optimal

Berdasarkan output pada Gambar 2, metode Elbow dengan melihat pergerakan grafik yang landai setelah grafik yang curam yaitu terdapat pada $k = 4$. Berdasarkan metode tersebut, didukung oleh teori dari (Aditya,2020) banyak cluster optimal yang terbentuk adalah 4. Diperoleh pengelompokan dengan k optimal 4 sebagai berikut.

```
## Clustering vector:
## 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
## 4 2 4 4 4 4 2 1 4 4 4 4 4 4 4 4 3 3 4 4 3 4 3 4 4 1
## 27 28 29 30 31 32 33 34
## 4 4 4 3 4 3 3 3
## ..
```

Gambar 3. Clustering k-Means

Berdasarkan output pada Gambar 3 dengan pemberian nama “klaster1”, diperoleh cluster 1 memiliki anggota dengan jumlah anggota terkecil sama dengan cluster 2 sebanyak 2 provinsi, cluster 3 memiliki anggota yaitu sebanyak 8 provinsi, cluster 4 memiliki anggota dengan jumlah anggota terbanyak sebanyak 22 provinsi.

2. Algoritma Fuzzy c-Means

Sedangkan clustering menggunakan data Indikator menggunakan pengujian algoritma Fuzzy c-Means diperoleh dua model Fuzzy c-means yaitu Model “klaster2” ($k=10$ dengan metrik jarak menggunakan Euclidean Distance) dan Model “klaster3” ($k=4$ dengan metrik jarak menggunakan Square Euclidean Distance). Setelah itu untuk memilih model Fuzzy c-Means mana yang terbaik adalah dengan melakukan uji validitas cluster menggunakan Dunn-Index. Diperoleh Model “klaster3” sebagai model Fuzzy c-Means yang lebih baik untuk digunakan karena memiliki nilai Dunn-Index sebesar 0,388 lebih besar dari Model “klaster2” sebesar 0,147.

```
## [1] 1 2 3 2 1 2 2 3 2 1 3 2 3 3 3 1 4 4 2 1 4 2 4 1 1 3 2 1 1 4 1 4 4 4
```

Gambar 4. Clustering Fuzzy c-Means

Berdasarkan output pada Gambar 4 dengan pemberian nama “klaster3”, diperoleh cluster 1 memiliki anggota dengan jumlah

anggota terkecil sama dengan cluster 2 sebanyak 10 provinsi, cluster 2 memiliki anggota yaitu sebanyak 9 provinsi, cluster 3 memiliki anggota dengan jumlah anggota terbanyak sebanyak 7 provinsi, cluster 4 memiliki anggota dengan jumlah anggota terbanyak sebanyak 8 provinsi.

3. Uji Validitas Cluster

Perbandingan clustering menggunakan algoritma k-Means dan Fuzzy c-Means diperoleh berdasarkan nilai Davies-Bouldin sebagai berikut.

```
print(index.DB(data=indikatorscale, klaster=k$cluster, centroidypes="centroids"))
## $DB
## [1] 1.165219

print(index.DB(data=indikatorscale, klaster=k$clustering, centroidypes="centroids"))
## $DB
## [1] 1.74298
```

Gambar 5. Uji Davies Bouldin-Index

Tabel 4. Nilai Akurasi Perbandingan Algoritma

Algoritma	Davies Bouldin Index
Fuzzy C-Means	1,74298
k-Means	1,165219

Nilai Davies Bouldin-Index menggambarkan seberapa baik cluster yang dibentuk. Semakin kecil nilai Davies-Bouldin Index atau semakin mendekati nilai 0 menunjukkan seberapa baik klaster yang diperoleh (Alith,2015). Oleh karena itu, berdasarkan Tabel 4 dan Gambar 5 terlihat bahwa dengan menggunakan algoritma k-Means (klaster1) memberikan hasil clustering yang lebih baik jika dibandingkan menggunakan algoritma fuzzy c-Means (klaster3). Ini berarti bahwa pengelompokan wilayah penyebaran Covid-19 di Indonesia lebih baik menggunakan algoritma k-Means.

4. Implementasi Algoritma Clustering

Setelah memilih algoritma k-Means untuk dilakukan clustering, Berdasarkan Gambar 3, dapat dilihat untuk cluster 1 memiliki 2 anggota yaitu provinsi 8 dan 26; untuk cluster 2 memiliki 2 anggota yaitu provinsi 7 dan 2; untuk cluster 3 memiliki 8 anggota yaitu provinsi 17, 18, 21, 23, 30, 32, 33, 34; untuk cluster 4 memiliki 22 anggota yaitu provinsi 1, 3, 4, 5, 6, 9, 10, 11, 12, 13, 14, 15, 16, 19, 20, 22, 24, 25, 27, 28, 29, dan 31.

```
## Within cluster sum of squares by cluster:
## [1] 71.92961 38.89286 106.96687 312.06803
## (between_SS / total_SS = 73.0 %)
```

Gambar 6. Akurasi Clustering k-Means

Selanjutnya diperoleh nilai within-cluster sum of squares dari cluster k-Means dengan k = 4 yaitu untuk cluster 1 sebesar 71,92961; cluster 2 sebesar 38,89286; cluster 3 sebesar 106,96687; dan cluster 4 sebesar 312,06803. Dengan nilai between_SS (between sum of squares) dibagi nilai total_SS (total sum squares) atau Akurasi sebesar 73,0% yang mempresentasikan seberapa berbeda nilai antar cluster dengan cluster lainnya. Akurasi menggambarkan seberapa baik cluster yang dibentuk bisa merepresentasikan matriks observasi. Semakin tinggi nilai akurasi maka keragaman antar cluster dengan cluster juga meningkat yang mana akan mengakibatkan semakin berkualitasnya cluster yang dilakukan (Kencana,2020). Within-cluster adalah nilai antar anggota dalam satu cluster, sedangkan between-cluster adalah nilai antar cluster dengan cluster lainnya. Ciri-ciri cluster yang baik adalah memiliki homogenitas (kesamaan) yang tinggi pada within-cluster dan memiliki heterogenitas (perbedaan) yang tinggi antar between-cluster (Goreti,2017).

Setelah cluster telah terbentuk seperti yang terlihat pada Gambar 3, selanjutnya adalah memberi ciri spesifik untuk menggambarkan keanggotaan masing-masing cluster yang telah terbentuk dengan melihat rata-rata masing-masing variabel (indikator kesehatan masyarakat) pada tiap masing provinsi.

Tabel 5. Nilai Rata-Rata Indikator

Provinsi	Rata-Rata	Cluster
Jambi	-0,263962862	Cluster 1
Lampung	-0,470574219	
Bali	0,037169528	
Jakarta	0,114400048	
Kep. Riau	0,225982132	Cluster 2
NTB	0,160170979	
.....	
NTT	0,065903585	
Gorontalo	-0,077426256	Cluster 3
Aceh	0,035275948	
.....	
Maluku	-0,036722717	
Papua	-0,031225928	Cluster 4

Berdasarkan tabel 5 akan diperoleh sebagai berikut.

1. Cluster pertama

$$= \frac{-0,263962862 + -0,470574219}{2}$$

$$= -0,36726854$$

Pada cluster pertama terlihat bahwa rata-rata masing-masing variabel sebesar -0,36726854. Oleh karena itu cluster pertama termasuk daerah dengan potensi kasus penyebaran Covid-19 sangat tinggi dari empat cluster yang terbentuk.

2. Cluster kedua

$$= \frac{0,037169528 + 0,114400048}{2}$$

$$= 0,075784788$$

Pada cluster kedua terlihat bahwa rata-rata masing-masing variabel sebesar 0,0757847. Oleh karena itu cluster kedua termasuk daerah dengan potensi kasus penyebaran Covid-19 rendah dari empat cluster yang terbentuk.

3. Cluster ketiga

$$= \frac{0,225982132 + \dots + (-0,077426256)}{8}$$

$$= 0,043557313$$

Pada cluster ketiga terlihat bahwa rata-rata masing-masing variabel sebesar 0,043557313. Oleh karena itu cluster pertama termasuk daerah dengan potensi kasus penyebaran Covid-19 sedang dari empat cluster yang terbentuk.

4. Cluster keempat

$$= \frac{0,035275948 + \dots + (-0,031225928)}{22}$$

$$= -0,009478537$$

Pada cluster keempat terlihat bahwa rata-rata masing-masing variabel sebesar -0,009478537. Oleh karena itu cluster keempat termasuk daerah dengan potensi kasus penyebaran Covid-19 tinggi dari ke empat cluster yang terbentuk.

PENUTUP

Berdasarkan hasil penelitian dan pembahasan dapat disimpulkan bahwa algoritma clustering dapat diterapkan untuk pengelompokan daerah penyebaran Covid-19 pada 34 provinsi di Indonesia berbantuan software Rstudio dimana pada penelitian ini

melakukan perbandingan dua algoritma yaitu algoritma k-Means dan Fuzzy c-Means. Diperoleh pengelompokan daerah penyebaran Covid-19 pada 34 provinsi Indonesia lebih baik dilakukan menggunakan algoritma k-Means. Pada penelitian ini menggunakan algoritma k-Means terbentuk 4 cluster pengelompokan daerah penyebaran Covid-19 pada 34 provinsi Indonesia. Dengan Provinsi Jambi dan Lampung termasuk daerah dengan potensi kasus penyebaran Covid-19 sangat tinggi di Indonesia; Provinsi Aceh, Banten, Bangka Belitung, Bengkulu, DIY Jogjakarta, Jawa Barat, Jawa Tengah, Jawa Timur, Kalimantan Barat, Kalimantan Timur, Kalimantan Tengah, Kalimantan Selatan, Kalimantan Utara, Sumatera Selatan, Sumbar, Kalimantan Timur, Sulawesi Selatan, Sulawesi Tengah, Riau, Maluku Utara, Maluku, dan Papua termasuk daerah dengan potensi kasus penyebaran Covid-19 tinggi di Indonesia; Provinsi Kepulauan Riau, Nusa Tenggara Barat, Kalimantan Barat, Sulawesi Tenggara, Papua Barat, Sulawesi Barat, Nusa Tenggara Timur, dan Gorontalo termasuk daerah dengan potensi kasus penyebaran Covid-19 sedang di Indonesia; dan Provinsi Bali dan Jakarta termasuk daerah dengan potensi kasus penyebaran Covid-19 rendah di Indonesia. Terdapat beberapa masukan untuk penulisan selanjutnya yaitu dapat menggunakan indikator kesehatan masyarakat pemerintah yang lain seperti penurunan jumlah kasus probable Covid-19, penurunan jumlah meninggal probable Covid-19, dan lain-lain.

DAFTAR PUSTAKA

- Anuraga, G. (2015). Hierarchical Clustering Multiscale Bootstrap Untuk Pengelompokan Kemiskinan Di Jawa Timur. *Jurnal Statistika Universitas Muhammadiyah Semarang*, 3(1).
- Simbolon, C. L., Kusumastuti, N., & Irawan, B. (2013). Clustering lulusan mahasiswa matematika fmipa untan pontianak menggunakan algoritma fuzzy c-means. *BIMASTER*, 2(1).
- Pitaloka, G. F., Dwidayati, N. K., & Mulyono, M. (2019). Perbandingan Metode Dalam Analisis Cluster Untuk Mengelompokkan Kabupaten/Kota Di Jawa Tengah. *Edusaintek*, 3.
- Irwansyah, E., & Faisal, M. (2015). *Advanced Clustering: Teori dan Aplikasi*. Deepublish.
- Nasari, F., & Darma, S. (2013). Penerapan K-Means Clustering Pada Data Penerimaan Mahasiswa Baru (Studi Kasus: Universitas Potensi Utama). *Semnasteknomedia Online*, 3(1), 2-1.
- Aditya, A., Jovian, I., & Sari, B. N. (2020). Implementasi K-Means Clustering Ujian Nasional Sekolah Menengah Pertama di Indonesia Tahun 2018/2019. *Jurnal Media Informatika Budidarma*, 4(1), 51-58.
- Alith, M. F. (2015). Klasterisasi Proses Seleksi Pemain Menggunakan Algoritma K-Means (Studi Kasus: Tim Hockey Kabupaten Kendal). *Skripsi*. Semarang: Fakultas Ilmu Komputer Universitas Dian Nuswantoro.
- Kencana, E.N. (2020). *Sains Data dengan R: Clusterisasi Menggunakan K-Means Clustering*. Diunduh di https://www.researchgate.net/publication/341787776_Sains_Data_dengan_R_Klasterisasi_Menggunakan_K-Means_Clustering tanggal 26 Juli 2021
- Goreti, M., Nasution, Y. N., & Wahyuningsih, S. (2017). Perbandingan Hasil Analisis Cluster dengan Menggunakan Metode Single Linkage dan Metode C-Means. *Jurnal Eksponensial*, 7(1), 9-16.