

IMPLEMENTASI ALGORITMA *PARTITIONING AROUND MEDOIDS* (PAM) DALAM PENGELOMPOKAN PENYEBARAN COVID-19 DI INDONESIA

Oleh: Supriatun
Univeristas Muhammadiyah Semarang

Article history	Abstract
Submission : Revised : Accepted :	<i>This study wants to classify cases of Covid-19 to find out which areas are most affected by Covid-19 and areas that are not much exposed to Covid-19. The covid-19 distribution here uses the Partitioning Around Medoids algorithm. The data grouping technique that is often used is the partitioning technique, one of the partitioning algorithms is partitioning around medoids (PAM) or better known as k-Medoids. This research is directed at the use of the k-Medoids algorithm to determine the spread of covid-19 in Indonesia based on the criteria for the level of covid-19 distribution. Measuring the quality of the cluster using the Silhouette obtained the optimum cluster with $k = 3$. $k = 3$ where cluster 1 has 268 data, cluster 2 has 62 data and cluster 3 has 12 data. However, in this clustering, if using $k = 3$ the clusters are not good, then re-testing is done using $k = 2$ and the result is that for cluster 1 there are 325 data and for cluster 2 there are 17 data.</i>
Keyword: <i>Covid-19, Medoids, Clustering, PAM, Silhouette.</i>	

PENDAHULUAN

Indonesia mulai di landa pandemi *Covid-19* pada awal Maret 2020, virus ini pertama kali muncul di wilayah Wuhan-Tiongkok ditemukan pada akhir Desember 2019. *Coronavirus Disease 2019* (COVID-19) adalah penyakit jenis baru yang belum pernah diidentifikasi sebelumnya pada manusia. Virus penyebab *Covid-19* ini dinamakan *Sars-CoV-2*. Virus *corona* adalah *zoonosis* (ditularkan antara hewan dan manusia). Adapun hewan yang menjadi sumber penularan *Covid-19* ini masih belum diketahui. Berdasarkan bukti ilmiah, *Covid-19* dapat menular dari manusia ke manusia melalui percikan batuk/bersin (droplet), Orang yang paling berisiko tertular penyakit ini adalah orang yang kontak erat dengan pasien *Covid-19* termasuk yang merawat pasien *Covid-19* (Kemenkes RI, 2020). Tanda dan gejala umum

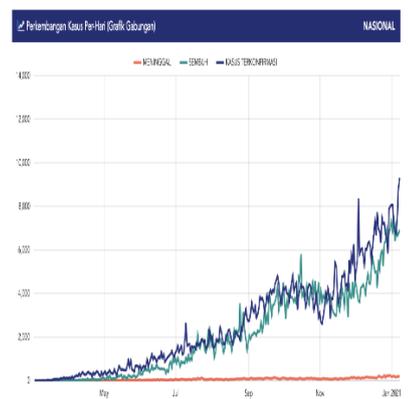
infeksi *covid-19* termasuk gejala gangguan pernapasan akut seperti demam, batuk, dan sesak napas dengan masa *inkubasi* rata-rata adalah 5 – 6 hari. Pada kasus yang parah, *covid-19* dapat menyebabkan *pneumonia*, sindrom pernapasan akut, gagal ginjal, dan bahkan kematian (Tosepu et al., 2020).

Wabah ini telah ditetapkan sebagai darurat kesehatan global. Virus ini sempat membuat semua kegiatan sehari-hari manusia terhambat. Karantina saja mungkin tidak cukup untuk mencegah penyebaran virus *Covid-19* ini, dan dampak global dari infeksi virus ini adalah salah satu yang semakin memprihatinkan (Sohrabi et al., 2020). Pemerintah Indonesia telah melakukan banyak langkah-langkah dan kebijakan untuk mengatasi permasalahan *pandemic* ini. Salah satu langkah awal yang dilakukan oleh pemerintah yaitu

mensosialisasikan gerakan *Social Distancing* untuk masyarakat.

Langkah ini bertujuan untuk memutus mata rantai penularan pandemi *COVID-19* ini karena langkah tersebut mengharuskan masyarakat menjaga jarak aman dengan manusia lainnya minimal 1 meter, tidak melakukan kontak langsung dengan orang lain serta menghindari pertemuan massal (Buana D.R, 2020). Namun, pada kenyataannya langkah-langkah tersebut tidak disikapi dengan baik oleh masyarakat, sehingga jumlah kasus terus meningkat. Di samping itu, pelayanan kesehatan di Indonesia dan SDM kesehatan yang ada dalam menangani kasus pandemic *COVID-19* ini juga belum memadai sedangkan kasus terus melonjak naik.

Pasien yang terkonfirmasi *COVID-19* di Indonesia berawal dari suatu acara di Jakarta dimana penderita kontak dengan seseorang warga Negara asing (WNA) asal Jepang yang tinggal di Malaysia. Setelah pertemuan tersebut penderita mengeluh demam, batuk dan sesak nafas (WHO, 2020). WHO mengumumkan *Covid-19* pada 12 Maret 2020 sebagai *pandemic*. Jumlah kasus di Indonesia terus meningkat dengan pesat. Total keseluruhan penyebaran *COVID-19* di Indonesia yang terkonfirmasi saat ini mencapai 516.753 ribu jiwa, untuk kasus penyebaran sembuh *COVID-19* di Indonesia mencapai 433.649 ribu jiwa dan angka kematian mencapai 16.352 jiwa. Data ini di ambil pada tanggal 26 November 2020 jam 21.53 WIB. (*Satuan Tugas Penanganan Covid-19*).



Sumber: Satuan tugas penanganan *COVID-19*

Gambar 1.1 Grafik kasus Gabungan

Dari grafik di atas dapat dilihat bahwa dari bulan Maret 2020 – Januari 2021 kasus yang terkena *COVID-19* terus melonjak namun pada bulan November 2020 sempat terjadi penurunan. Dan pada bulan Desember terjadi peningkatan kembali (*Satuan Tugas Penanganan Covid-19*).

Penambangan data (*data mining*) adalah proses yang memperkerjakan satu atau lebih teknik pembelajaran komputer (*machine learning*) untuk menganalisis dan mengekstraksi pengetahuan (*knowledge*) secara otomatis. *Data mining* berisi pencarian trend atau pola yang diinginkan dalam database besar untuk membantu pengambilan keputusan di waktu yang akan datang. Pola – pola ini dikenali oleh perangkat tertentu yang dapat memberikan suatu analisa data yang berguna dan berwawasan yang kemudian dapat dipelajari dengan lebih teliti, yang mungkin saja menggunakan perangkat pendukung keputusan yang lainnya (Hermawati, 2013). *Clustering* merupakan salah satu teknik dalam penambangan data.

Clustering merupakan salah satu teknik dalam penambangan data. *Clustering* bertujuan untuk mengelompokkan sejumlah data/obyek ke dalam kluster (*group*) sehingga dalam setiap kluster akan berisi data yang semirip mungkin. Salah satu *Algoritma clustering* adalah *Partitioning Around Medoids* (PAM). *Algoritma Partitioning Around Medoids* (PAM) lebih dikenal dengan *Algoritma K-Medoids*. *Algoritma* ini merupakan teknik partisi klasik untuk *clustering* yang melakukan klusterisasi data dari n obyek ke dalam k *cluster* yang dikenal dengan *a priori*. Perbedaan antara *K-Means* dan *K-Medoids* terletak pada cara kedua metode tersebut menghentikan perhitungan dan melakukan perulangan. Jika pada *K-Means* berpatokan pada kondisi konvergen, pada *K-Medoids* tergantung pada kualitas kluster yang didapat pada perulangan terakhir (Tiwari, 2012). *Algoritma K-Medoids* lebih kuat daripada *Algoritma K-Means* bila dihadapkan dengan *noise* dan *outliers*, karena *medoids* kurang dipengaruhi oleh *outliers* atau nilai - nilai ekstrim yang lain daripada sebuah mean (Yusupa, 2015). Tujuan dari penelitian tersebut adalah untuk mengimplementasikan *algoritma clustering Partitioning Around Medoid* (PAM) dalam mengelompokkan penyebaran *covid-19* di Indonesia. Diharapkan penelitian ini dapat memberikan informasi kesamaan karakteristik virus *Covid-19* yang terjangkit di Dunia. Sehingga dapat memberikan informasi dalam pembuatan antivirus atau vaksin.

Penelitian yang dilakukan oleh Dedy Hartama dkk (2020) yang berjudul “ Analisis Algoritma K-Medoids Clustering dalam Pengelompokan Penyebaran *Covid-19* di Indonesia” yang bertujuan untuk mengetahui seberapa banyak pengelompokan pada masalah kasus *covid-19* di Indonesia menggunakan metode *K-Medoids Clustering*. Penelitian berikutnya dari Achmad Solichin dan Khansa Khairunnisa dengan judul “*Klusterisasi Persebaran Virus Corona (Covid-19) Di DKI Jakarta Menggunakan Metode K-Means*” Tujuan dilakukannya penelitian ini untuk mengkluster persebaran Virus Corona di provinsi DKI Jakarta berdasarkan parameter jumlah ODP, PDP, kasus Positif, pasien sembuh dan pasien meninggal. *Klusterisasi* dilakukan menggunakan metode *K-means*, yang mengelompokkan data ke dalam beberapa kluster berdasarkan kemiripan data.

Serta penelitian lainnya yaitu dari Setiyawati (2017) yang berjudul “ Implementasi *Algoritma Partitioning Around Medoids* (PAM) untuk Pengelompokkan Sekolah Menengah Atas di DIY berdasarkan Nilai Daya Serap Ujian Nasional” yang bertujuan untuk membantu dalam proses pengelompokkan data Sekolah Menengah Atas dengan *Algoritma Partitioning Around Medoids* (PAM). Alasan dalam penelitian tersebut menggunakan *Partitioning Around Medoid* (PAM) dalam menganalisis kluster adalah karena *algoritma* ini memiliki kelebihan yaitu hasil proses clustering tidak bergantung pada urutan masuk dataset.

Disamping itu, algoritma ini mengatasi sensitif terhadap noise dan outlier, dimana objek dengan nilai yang besar yang memungkinkan menyimpang dari distribusi data. Dan penelitian serupa oleh Heryn Februariyanti dan Dwi Budi Santoso dengan judul “Algoritma *Partitioning Around Medoids* (PAM) *Clustering* untuk Melihat Gambaran Umum Skripsi Mahasiswa” Tujuan yang ingin dicapai dalam penelitian tersebut adalah merancang dan melakukan implementasi algoritma *Partitioning Around Medoids* (PAM) *Clustering* untuk Melihat Gambaran Umum Kemampuan Akademik Mahasiswa.

Berdasarkan latar belakang di atas penulis ingin mengetahui pengelompokan penyebaran *Covid-19*. Penulis menggunakan metode Penerapan *Algoritma Partitioning Around Medoids* (PAM) agar dapat diketahui pola pemilihan penentuan pengelompokan penyebaran *covid-19* di berbagai wilayah di Indonesia. *Partitioning Around Medoids* (PAM) merupakan metode *partitional clustering* dimana bertujuan untuk menemukan satu set *k-cluster* di antara data yang paling mencirikan objek dalam kumpulan suatu data.

LANDASAN TEORI

ANALISIS DESKRIPTIF

Analisis Deskriptif Menurut Hasan (2001), statistik deskriptif atau statistik deduktif adalah bagian dari statistik mempelajari cara pengumpulan data dan penyajian data sehingga mudah dipahami. Statistik deskriptif hanya berhubungan dengan hal menguraikan atau memberikan keterangan-keterangan mengenai suatu data atau keadaan atau fenomena.

Knowledge Discovery in Database (KDD)

Knowledge Discovery in Database (KDD) adalah kegiatan yang meliputi pengumpulan, pemakaian data historis untuk menemukan keteraturan, pola, atau hubungan dalam set data berukuran besar.

Penambangan Data (Data Mining)

Data mining adalah suatu proses penambangan informasi penting dari suatu data. Informasi penting ini didapat dari suatu proses yang amat rumit seperti menggunakan artificial intelligence, teknik statistik, ilmu matematika, machine learning, dan lain sebagainya. Teknik-teknik rumit tersebut nantinya akan mengidentifikasi dan mengekstraksi informasi yang bermanfaat dari suatu database besar. (Efraim Turban, dkk 2005).

Standarisasi Data / Pembakuan Data

Menurut Hair, et al. (2016) pembakuan data adalah proses mengkonversi nilai masing – masing variabel awal menjadi nilai standar dengan rata – rata 0 dan standar deviasi 1 untuk menghilangkan bias yang disebabkan karena perbedaan skala dari beberapa variabel yang digunakan dalam analisis.

Clustering

Clustering yaitu menemukan kumpulan obyek hingga obyek-obyek dalam satu kelompok sama (atau punya hubungan) dengan yang lain dan berbeda (atau tidak berhubungan) dengan obyek – obyek dalam kelompok lain. Tujuan dari *clustering* adalah untuk meminimalkan jarak di dalam *cluster* dan memaksimalkan jarak antar *cluster*. Dalam mengukur jarak dalam *clustering* dapat dilakukan dengan menggunakan *Euclidean Distance*. *Euclidean Distance* merupakan pengukuran jarak obyek dan pusat *cluster* yang

banyak digunakan secara luas dalam berbagai kasus *pattern matching*, termasuk *clustering*. (Astri Widiastuti Setiyawati, 2017)

Partitioning Around Medoids (PAM)

Partitioning Around Medoids (PAM) atau di kenal dengan K – Medoids adalah algoritma pengelompokan yang berkaitan dengan algoritma K – Means dan algoritma K – Medoids. *Algoritma Partitioning Around Medoids* (PAM) dikembangkan oleh Leonard Kuaufman dan Peter J. Rousseuw. Algoritma ini sangat mirip dengan algoritma K – Means, terutama karena kedua algoritma ini partitional. Dengan kata lain, kedua algoritma ini memecah dataset menjadi kelompok –kelompok dan kedua algoritma ini berusaha untuk meminimalkan kesalahan. Tetapi algoritma *Partitioning Around Medoids* (PAM) bekerja dengan menggunakan Medoids

Algoritma Partitioning Around Medoids (PAM)

Algoritma Partitioning Around Medoids (PAM) atau K – Medoids adalah sebagai berikut (Han & Kamber, 2006):

1. Secara acak pilih k obyek pada sekumpulan n obyek sebagai *medoids*.
2. Ulangi langkah 3 hingga langkah 6.
3. Tempatkan obyek *non-medoids* ke dalam *cluster* yang paling dekat dengan *medoids*.
4. Secara acak pilih 0_{random} sebuah obyek *non-medoids*.

5. Hitung total biaya, S , dari pertukaran *medoids* 0_j dengan 0_{random} .
6. Jika $S < 0$ maka tukar 0_j dengan 0_{random} untuk membentuk sekumpulan k obyek baru sebagai *medoids*.
7. Hingga tidak ada perubahan. Nilai total biaya/*cost* dinyatakan dengan persamaan:

Total *cost* =

$$\sum Dist \sqrt{\sum_{k=1}^n (p_k - q_k)^2}$$

Nilai S dinyatakan dengan persamaan:

$$S = \text{Total } cost \text{ baru} - \text{Total } cost \text{ lama}$$

Dimana :

Total *cost* baru = jumlah biaya/*cost* non-medoids.

Total *cost* lama = jumlah biaya/*cost* medoids.

n = jumlah fitur dalam suatu data

k = indeks data

P_k = nilai atribut (fitur) ke- k dari p

q_k = nilai atribut (fitur) ke- k dari q

K – medoids sangat mirip dengan K – means, perbedaan utama diantara dua algoritma tersebut adalah jika pada K – means *cluster* diwakili dengan pusat dari *cluster*, sedangkan pada K – medoids *cluster* diwakili oleh obyek terdekat dari pusat *cluster*.

Silhotte Index (SI)

Jika DBI digunakan untuk mengukur validasi seluruh *cluster* dalam set data, maka *Silhouette Index* (SI) dapat digunakan untuk memvalidasi baik sebuah data, *cluster* tunggal (satu *cluster* dari sejumlah *cluster*), atau bahkan keseluruhan *cluster*. Metode ini yang paling banyak digunakan untuk memvalidasi *cluster* yang menggabungkan nilai kohesi dan separasi. Untuk menghitung nilai SI dari sebuah data ke-i, ada 2 komponen yaitu a_i dan b_i . a_i adalah rata – rata jarak ke-i terhadap semua data lainnya dalam satu *cluster* sedangkan b_i didapatkan dengan menghitung rata – rata jarak data ke-i terhadap semua data dari *cluster* yang lain tidak dalam satu *cluster* dengan data ke-i, kemudian diambil yang terkecil (Tan et al, 2006 & Petrovic,2003).

$$\frac{1}{m_j - 1} \sum_{r=1, r \neq i}^{m_j} d(x_i^j, x_r^j), i = 1, 2, \dots, m_j \quad (2.3)$$

Dimana :

$j = cluster$

$i = index\ data$

$a_i^j =$ rata – rata jarak data ke-i terhadap semua data lainnya dalam satu *cluster*.

$m_j =$ jumlah data dalam *cluster* ke-j.

$d(x_i^j, x_r^j)$ adalah jarak data ke-i dengan data ke-r dalam satu *cluster* j.

Berikut formula untuk menghitung b_i^j :

$$\min_{n=1, \dots, k, n \neq j} \left\{ \frac{1}{m_n} \sum_{r=1, r \neq i}^{m_n} d(x_i^j, x_r^n) \right\} i = 1, 2, \dots, m_j \quad (2.4)$$

Dimana :

$j = cluster$

$n = cluster$

$i = index\ data$

$m_n =$ banyak data dalam satu *cluster*

$b_i^j =$ nilai terkecil dari rata – rata jarak data ke-i terhadap semua data dari *cluster* yang lain tidak dalam satu *cluster* dengan data ke-i

$d(x_i^j, x_r^n)$ adalah jarak data ke-i dalam satu *cluster* j dengan data ke-r dalam suatu *cluster* n

Untuk mendapatkan *Silhouette Index* (SI) data ke-i menggunakan persamaan berikut :

$$SI_i^j = \frac{b_i^j - a_i^j}{\max(b_i^j, a_i^j)} \quad (2.5)$$

Dimana :

$SI_i^j =$ *Silhouette Index* data ke-i dalam satu *cluster*

$b_i^j =$ nilai terkecil dari data rata – rata jarak data ke-i terhadap semua data dari *cluster* yang lain tidak dalam satu *cluster* dengan data ke-i

a_i^j = rata – rata jarak data ke-
i terhadap semua data lainnya dalam
satu *cluster*

Nilai a_i mengukur seberapa
tidak mirip sebuah data dengan
cluster yang diikutinya, nilai yang
semakin kecil menandakan semakin
tepatnya data tersebut berada dalam
cluster tersebut. Nilai b_i yang besar
menandakan seberapa jeleknya data
terhadap *cluster* yang lain. Nilai SI
yang didapat dalam rentang [-1,+1].
Nilai SI yang mendekati 1
menandakan bahwa data tersebut
semakin tepat berada dalam *cluster*
tersebut. Nilai SI negatif ($a_i > b_i$)
menandakan bahwa data tersebut
tidak tepat berada di dalam *cluster*
tersebut (karena lebih dekat ke
cluster yang lain). SI bernilai 0 (atau
mendekati 0) berarti data tersebut
posisinya berada di perbatasan di
antara dua *cluster*.

Untuk nilai SI dari sebuah
cluster didapatkan dengan
menghitung rata – rata nilai SI semua
data yang bergabung dalam *cluster*
tersebut, seperti pada persamaan
berikut :

$$SI_j = \frac{1}{m_j} \sum_{i=1}^{m_j} SI_i^j \quad (2.6)$$

Dimana :

SI_j = rata – rata *Silhouette Index*
cluster j

m_j = jumlah data dalam *cluster* ke-j

SI_i^j = *Silhouette Index* data ke-i dalam
satu *cluster*

i = *index*

Sementara nilai SI global
didapatkan dengan menghitung rata –
rata nilai SI dari semua *cluster* seperti
pada persamaan berikut :

$$SI = \frac{1}{k} \sum_{j=1}^k SI_j \quad (2.7)$$

Dimana :

SI = rata – rata *Silhouette Index* dari
database

k = jumlah *cluster*

SI_j = rata – rata *Silhouette Index*
cluster j

METODE PENELITIAN

Sumber Data

Data yang digunakan pada
penelitian ini adalah data
sekunder. Data yang digunakan
adalah data bulanan kasus *Covid-19*
yaitu kasus positif, kasus
sembuh, dan kasus meninggal
yang tercatat dan dilaporkan di
Kementrian Kesehatan Republik
Indonesia dan variabel yang saya
gunakan berasal dari satuan tugas
penanganan *covid-19* mulai dari 12
Maret 2020 sampai 31 Desember
2020.

Variabel Penelitian dan Struktur Data

Variabel yang digunakan
terdiri dari kasus positif, sembuh,
dan meninggal. Dan data yang
saya gunakan sebanyak 342 data.

Variabel Penelitian

Variabel penelitian yang
digunakan pada penelitian ini akan

diuraikan pada tabel 3.2.1 dibawah ini.

Tabel 3.2.1 Variabel Penelitian

Variabel	Keterangan
X1	Kasus Positif
X2	Kasus Sembuh
X3	Kasus Meninggal

Struktur Data

Struktur data yang digunakan pada penelitian ini ditunjukkan dibawah ini.

Tabel 3.2.2 Struktur Data

Provinsi	Bulan	X1	X2	X3
Aceh	Januari	X _{1.1}	X _{1.2}	X _{1.3}
	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮
	Desember	X _{12.1}	X _{12.2}	X _{12.3}
Bali	Januari	X _{1.1}	X _{1.2}	X _{1.3}
	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮
	Desember	X _{12.1}	X _{12.2}	X _{12.3}
Sumatera Utara	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮
	⋮	⋮	⋮	⋮
	Januari	X _{1.1}	X _{1.2}	X _{1.3}
⋮	⋮	⋮	⋮	
⋮	⋮	⋮	⋮	

⋮
Desember X_{342.1} X_{342.2} X_{342.3}

Langkah Penelitian

Dalam Proses Perhitungan pada metode *Partitioning Around Method* (PAM) berikut langkah-langkahnya sebagai berikut :

1. Analisis Deskriptif Statistik
2. *Missing value* dan *Outlier*
3. Standarisasi data

4. Uji Asumsi Multikolinearitas jika terjadi multikolinearitas maka dilakukan perbaikan dengan reduksi variabel melalui PCA

5. Analisis komponen utama atau (*Principal Component Analysis*) PCA

6. Menentukan K (jumlah cluster) dengan metode *silhoutte*

7. Analisis *partitioning around medoids* (PAM)

8. Menarik kesimpulan

HASIL PENELITIAN & PEMBAHASAN

Analisis Deskriptif

Total kasus positif terendah berada pada bulan maret (0 jiwa) dan Total kasus positif tertinggi berada pada bulan Desember (23.745 jiwa) dengan nilai rata-rata sebesar 194,099 dan nilai standar deviasi 2477,588. Total kasus

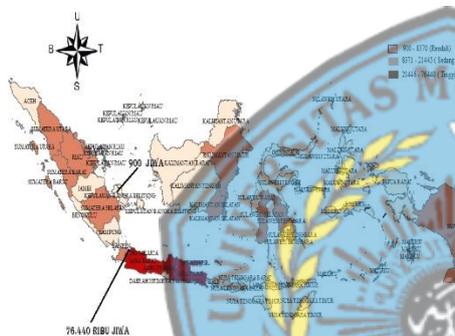
sembuh terendah berada pada maret (0 jiwa) dan nilai tertinggi berada pada bulan Desember (164.787 Jiwa) dengan nilai rata-rata sebesar 113,915 nilai standar deviasi 15316,77 . Hasil total kasus meninggal terendah berada pada bulan maret (0 jiwa) dan Total kasus meninggal tertinggi berada pada bulan Desember (5.827 jiwa) dengan nilai rata-rata sebesar 114,3 nilai standar deviasi 627,9637.

Total Positif	Logical	342
Total Sembuh	Logical	342
Total Meninggal	Logical	342

Dari 342 data yang terdiri dari kasus Positif, sembuh, dan meninggal tidak ada yang memiliki *missing value* dilihat dari Data False.

Mengecek Data *Outlier*

Peta Penyebaran Kasus *Covid-19* di Indonesia



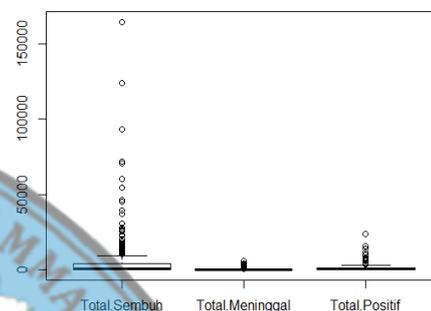
Gambar 4.2.1 Peta Kasus Positif *Covid-19* di Indonesia

Penyebaran kasus *Covid-19* di Indonesia terlihat pada Gambar 4.2.1 untuk DKI Jakarta adalah provinsi yang tinggi terkena kasus *covid-19* di Indonesia sebanyak 76.440 ribu jiwa dan untuk provinsi yang paling rendah adalah provinsi Kepulauan Bangka Belitung yaitu sebanyak 900 jiwa selama tahun 2020.

Missing Value dan *Outlier* Mengecek *Missing Value*

Tabel 4.3.1 *Missing Value*

Variabel	Mode	Data False
----------	------	------------



Gambar 4.3.2 Deteksi *Outlier*

Masing-masing variabel memiliki *outlier*, dapat dilihat pada *boxplot* di atas bahkan dari variabel pasien Sembuh banyak yang *outlier* bahkan titik outliernya pada kasus sembuh sangat jauh. Oleh karena itu dapat dilanjutkan ke pengujian selanjutnya karena metode *Partitioning Around Medoids* ini mengatasi data yang *outlier*.

Standarisasi Data

Dari 3 variabel yang digunakan memiliki ukuran yang berbeda – beda maka perlu dilakukan penyeragaman dengan cara standarisasi atau pembakuan nilai menggunakan Z score. Hasil standarisasi data dapat dilihat pada tabel 4.4 dibawah ini.

Tabel 4.4 Hasil Standarisasi Data

No	Total Sembuh	Total Meninggal	Total Positif
1	-0.351246427	-0.380222768	-0.513290394
2	-0.351246427	-0.380222768	-0.513290394
3	-0.351246427	-0.378630319	-0.513290394
4	-0.351246427	-0.380222768	-0.513290394
5	-0.351246427	-0.380222768	-0.513290394
...
...
340	4270356036	1486127261	4176350933
341	4358298887	8898976567	1985107042
342	1040735679	4827084907	5814637717

Uji Multikolinearitas

Pada analisis kluster uji asumsi yang diperlukan yaitu tidak terjadi multikolinearitas antara variabel. Uji asumsi multikolinearitas dilakukan dengan tujuan untuk mengetahui apakah terjadi multikolinearitas atau tidak pada data yang dianalisis. Jika tidak terjadi maka data dapat langsung dianalisis, namun jika terjadi multikolinearitas maka harus dilakukan penanganan data terlebih dahulu sebelum data dianalisis menggunakan kluster. Pengujian untuk memastikan adanya multikolinearitas atau tidak dalam data adalah dengan menggunakan kriteria statistik diantaranya dengan menggunakan nilai $VIF > 10$. Selain itu, bisa dengan melihat nilai dari matrik korelasi antar variabel. Menurut Yamin dan Kurniawan, 2016 dikatakan terjadi multikolinearitas apabila nilai korelasi antar variabel

lebih dari 0,70. Nilai korelasi dapat juga dilihat pada tabel 4.5

Tabel 4.5 Matrix Uji Multikolinearitas

Variabel	Matrix		
Total Sembuh	1	0.800281	0.748124
Total Meninggal	0.800281	1	0.6848668
Total Positif	0.7481247	0.6848668	1

Dari *output* matriks korelasi 3 x 3 pada gambar di atas, menunjukkan adanya multikolinearitas sebab ada nilai korelasi antar variabel yang lebih besar dari 0,70 yakni korelasi antar variabel X_2 dengan X_1 sebesar 0,8002813, X_3 dengan X_1 0,7481247. Terjadinya multikolinearitas ini perlu dilakukan upaya perbaikan dengan reduksi variabel melalui *Principial Component Analysis* (PCA) atau Analisis Komponen Utama (AKU) untuk memperoleh variabel baru yang akan digunakan dalam analisis kluster.

Principal Component Analysis (PCA)

Salah satu solusi untuk menangani data yang multikolinearitas adalah dengan *Principal Component Analysis* (PCA) atau disebut juga dengan Analisis Komponen Utama (AKU). Berikut didapatkan nilai *eigen value*, proporsi varian, dan proporsi kumulatif, dan dapat juga dilihat pada tabel 4.6 dibawah ini.

Tabel 4.6 Nilai Eigen, Proporsi Varian, Proporsi Kumulatif

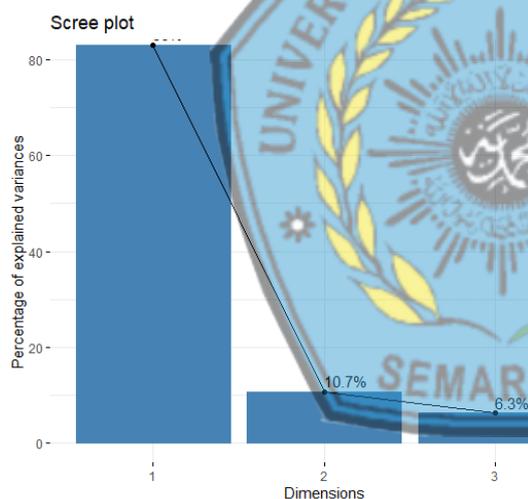
Variabel	Nilai Eigen	Proporsi Varian	Proporsi Kumulatif
Total Sembuh	2.5	0.6	-0.2
Total Meninggal	0.3	0.6	-0.6
Total Positif	0.2	0.6	0.8

Tabel 4.6.2 Penentuan Faktor

	PC1	PC2
Total.Sembuh	0.6	-0.2
Total.Meninggal	0.6	-0.6
Total.Positif	0.6	0.8

Berdasarkan tabel 4.6 diatas, dapat diketahui nilai *eigen value* masing – masing variabel. Nilai *eigen value* tersebut digunakan untuk menentukan jumlah komponen yang akan dipilih. Didapatkan hasilnya dari nilai eigen value yang nilainya lebih dari 1 ada 1 faktor yaitu 2.5. Maka jumlah faktor yang akan digunakan ada 1, selanjutnya untuk membuktikannya lagi menggunakan sebuah histrogram.

Diketahui bahwa yang termasuk ke dalam salah faktor dari 1 faktor bila nilainya berada di bawah - 0.5 dan diatas 0.5. Bila ada yang tidak berada dalam batas maka dilihat dari nilai korelasi terbesar. Hasilnya dapat dilihat pada tabel diatas ini dimana setiap variabel akan dominan pada salah satu komponen. Maka didapatkan hasil pengelompokkan adalah Faktor 1 tersebut terdiri dari total sembuh, total meninggal dan total positif.



Gambar 4.6 Presentase Faktor yang digunakan

Berdasarkan hasil grafik dengan melihat kelandaian bentuk grafik maka didapatkan bahwa faktor yang akan terbentuk adalah 1. Selanjutnya dilakukan penentuan faktor.

Tabel 4.6.3 Hasil PCA

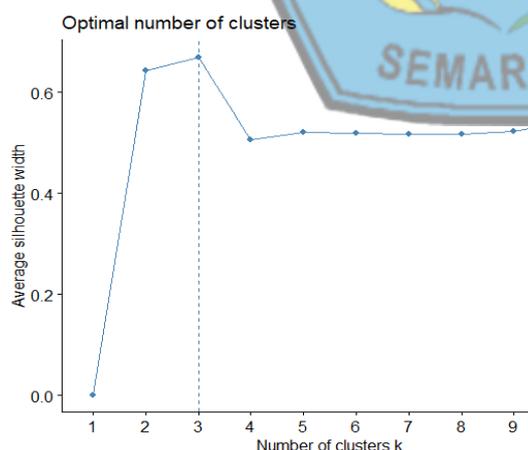
No	PC1	PC2
1	-0.716116327	-0.1194597578
2	-0.716116327	-0.1194597578
3	-0.715197837	-0.1203947198
4	-0.716116327	-0.1194597578
5	-0.716116327	-0.1194597578
...
...
339	10.051.247.732	34.847.283.388
340	5.736.070.027	16.640.119.624
341	8.830.994.835	-
342	12.219.766.872	-0.0978544410

Diatas merupakan hasil dari PCA, maka data sudah siap untuk

dilakukan analisis. Untuk membuktikan apakah pada data sudah tidak terdapat multikolinearitas maka bisa dilakukan ulang uji *bartlett*. Dengan menggunakan tingkat kepercayaan 95%, maka data yang ada menjelaskan bahwa tidak terdapat multikolinearitas pada data, karena nilai p-value (2,26) > 0.05. Maka asumsi pada data sudah terpenuhi dengan bantuan PCA. Maka data sudah siap untuk dilakukan analisis.

Penentuan Jumlah K

Pada penelitian ini dilakukan pengelompokan kasus *covid-19* di Indonesia berdasarkan indikator kasus positif, sembuh dan meninggal menggunakan analisis *Partitioning Around Medoids* (PAM). Metode *Partitioning Around Medoids* (PAM) merupakan metode pengklasteran untuk mengumpulkan sekumpulan n objek ke k klaster, pengelompokan ini menggunakan *medoids* sebagai pusat klasternya. Penelitian ini menggunakan metode *Silhouette*.



Gambar 4.7 Penentuan Jumlah K

Penentuan jumlah k , penulis menggunakan metode *Silhouette* selain itu Metode *Silhouette* menggunakan pendekatan nilai rata-rata untuk menduga kualitas dari *cluster* yang terbentuk. Semakin tinggi nilai rata-ratanya maka akan semakin baik. Berdasarkan saran dari metode *Silhouette*, diperoleh jumlah *cluster* optimum yang terbentuk adalah $k = 3$, sedangkan opsi kedua adalah $k = 2$. Hal ini dikarenakan nilai rata-rata *Silhouette* pada $k = 3$ dan $k = 2$ merupakan yang tertinggi dari yang lain. Berdasarkan k optimal yang di dapatkan, maka dapat dilakukan pengklasteran dengan $k=3$. Namun $k=2$ juga mendekati k optimal jadi dilakukan pengklasteran dengan ke k optimal yang didapatkan.

Clustering

Tabel 4.8.1 Hasil Medoids

ID	Provin si	Total Sembuh	Total Mening gal	Total Positif
209	NTT	274	6	144
203	Sulawesi Selatan	9247	360	2371
240	DKI Jakarta	60231	1729	11776

Dari *output* di atas dapat diketahui bahwa *medoid* atau objek yang mewakili untuk menjadi titik tengah pada *cluster-1* adalah **obyek ke-209 Provinsi Nusa Tenggara Timur**, *cluster-2* adalah **obyek ke-203 Provinsi Sulawesi Selatan**, sedangkan *medoid* yang digunakan *cluster-3* adalah **obyek ke-240**

Sumber : Olah data R tahun 2021

Provinsi DKI Jakarta. Sehingga, objek yang memiliki jarak terdekat dengan *medoid* suatu *cluster* akan masuk ke *cluster* tersebut. Jarak yang digunakan pada metode *K-Medoids* adalah jarak *Euclidean*.

Tabel 4.8.2 Output Menurut Bulan

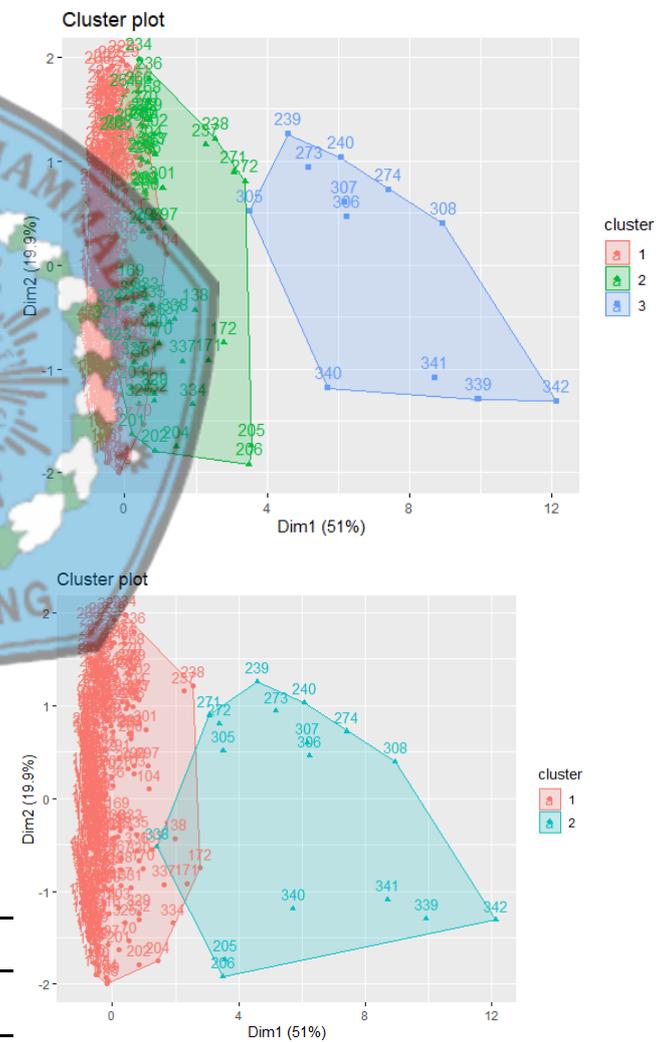
Cl u	Bulan						
	M ar	A pr	M ei	O kt	No v	De s
1	35	35	...	34	19	17	13
2	0	0	...	0	13	13	17
3	0	0	...	0	2	4	4

Dari hasil uji menggunakan metode silhoutte di dapatkan $k = 3$, dan dapat kita lihat bahwa cluster 1 beranggotakan 35 provinsi (bulan Maret), 35 provinsi (bulan April), 34 provinsi (bulan Mei), sampai dengan bulan Desember terdapat 13 provinsi. Cluster 2 beranggotakan 2 provinsi (bulan Juni), 4 provinsi (bulan juli), sampai dengan bulan Desember terdapat 17 provinsi. Dan Cluster 3 beranggotakan 2 provinsi (bulan September), 2 provinsi (bulan Oktober), sampai dengan bulan Desember terdapat 4 provinsi.

Tabel 4.8.3 Numerical information per cluster

Numerical information per cluster:			
Urutan data	max_dis	av_diss	diameter
268	498.47	1,046,13	5,282,206
62	22130.7	5,495,45	26,551,37

Numerical information per cluster menunjukkan informasi jumlah data, jarak maksimum, dan nilai rata-rata jarak dari masing masing cluster. Untuk *cluster* pertama terdapat sebanyak 268 data, pada *cluster* kedua terdapat sebanyak 62 data. Dan *cluster* ketiga sebanyak 12 data.



Gambar 4.8 visualisasi K-Medoids

Dari metode *silhouette* diperoleh k optimal adalah $k=3$ dan dapat dilihat pada gambar di atas bahwa *cluster* 1 berwarna merah yang terdapat 268 data, *cluster* 2 berwarna hijau yang terdapat 62 data dan *cluster* 3 terdapat 12 data. Namun dalam pengklasteran ini jika menggunakan $k=3$ kurang bagus clusternya, karena tidak terpisah secara maksimal maka dari itu dilakukan pengujian kembali menggunakan $k=2$ dan dihasilkan untuk *cluster* 1 berwarna merah yang terdapat 325 data dan untuk *cluster* 2 terdapat 17 data. Berdasarkan gambar 4.8 *cluster* 2 tidak terpisah namun lebih baik dibandingkan *cluster* 3.

Rata – rata Cluster

Tabel 4.9 Rata – rata Cluster

Cluster	Total. Sembuh	Total. Meninggal	Kasus. Aktif
1	878	2586	123
2	8807	58786	2455

Cara untuk melihat karakteristik datanya yaitu dengan menghitung nilai rata-rata variabel pada masing-masing *cluster*. Dari nilai rata-rata di atas, kita dapat mengetahui bahwa karakteristik dari *cluster*-1 yang memiliki jumlah kasus yang tergolong sangat tinggi, dan *cluster*-2 memiliki kasus tergolong rendah.

Kesimpulan

Berdasarkan hasil dari pembahasan dapat ditarik kesimpulan sebagai berikut:

1. Kasus *covid-19* yang paling tinggi berada pada provinsi

DKI Jakarta selama tahun 2020 terdapat 76.440 ribu jiwa yang terinfeksi virus tersebut dan provinsi yang paling rendah terinfeksi kasus *covid-19* yaitu Kepulauan Bangka Belitung sebanyak 900 jiwa

2. Pengelompokan 34 Provinsi di Indonesia dari bulan Maret - Desember berdasarkan kasus *covid-19* menggunakan metode *Partitioning Around Medoids* diperoleh bahwa dengan nilai *Silhouette* kita dapat menghasilkan 3 *cluster*. Namun $k=2$ juga mendekati k optimal jadi dilakukan pengklasteran dengan ke k optimal yang didapatkan.

Dari metode *silhouette* diperoleh k optimal adalah $k=3$ dan dapat dilihat pada gambar di atas bahwa *cluster* 1 berwarna merah yang terdapat 268 data, *cluster* 2 berwarna hijau yang terdapat 62 data dan *cluster* 3 terdapat 12 data. Namun dalam pengklasteran ini jika menggunakan $k=3$ kurang bagus clusternya, karena tidak terpisah secara maksimal maka dari itu dilakukan pengujian kembali menggunakan $k=2$ dan dihasilkan untuk *cluster* 1 berwarna merah yang terdapat 325 data dan untuk *cluster* 2 terdapat 17 data. Berdasarkan gambar 4.8 *cluster* 2 tidak terpisah namun lebih baik dibandingkan *cluster* 3.

Saran

Pada skripsi ini penulis hanya mengkaji tentang pengklasteran

Partitioning Around Medoids. Diharapkan untuk peneliti selanjutnya dapat mengembangkan dengan

DAFTAR PUSTAKA

- Adisaputra, D. J., Nugroho, S., & Novianti, P. (2013). Pengelompokan Provinsi Di Indonesia Berdasarkan Rata-Rata Produksi Tanaman Pangan Menggunakan Metode Klaster K-Menas. *Journal Of Chemical Information And Modeling*, 53(9), 1689–1699.
- Amanah Nur, Wahyuningsih Sri, dan Amijaya F.D.T. (2017). Analisis Cluster Non-Hirarki Dengan Menggunakan Metode K-Modes pada Mahasiswa Program Studi Statistika Angkatan 2015 FMIPA Universitas Mulawarman. *Jurnal EKSPONENSIAL*, Vol 8 No 1.
- Anjarsari A.Y, Nilogiri Agung, Faruq H.A.A. (2020). *Alogaritma Partitioning Around Medoids (PAM) untuk mengelompokkan kabupaten/kota di Jawa Timur berdasarkan komponen pembentuk indeks pembangunan manusia*. Program Studi Teknik Informatika, Universitas Muhammadiyah Jember
- Archika Nazwa Dwi. 2020. Makalah Corona Virus Disease 2019. SMA Negeri 3 Medan. 4 Mei. Medan.
- Arora, P., Deepali, & Varshney, S. (2016). Algoritma K-Medoid dan Contoh Perhitungan Manual. <http://studyshut.blogspot.com/2018/12/algoritma-k-medoid-dan-contoh.html>. 28 Maret 2021 (09.00)
- menggunakan analisis cluster lainnya mengingat cakupan analisis cluster yang cukup banyak.
- Budiman M Puja Alif. (2020). Analysis of K-Means and KMedoids Algorithm for Big Data. *Procedia Computer Science* 78, 507-512.
- Chrisnanto Y Herry dan Abdillah Gunawan. (2015). Penerapan Alogaritma Partitioning Around Medoids (PAM) Clustering untuk Melihat Gambaran Umum Kemampuan Akademik Mahasiswa. *Simposium Nasional Teknologi Informasi dan Komunikasi*.
- Februariyanti Hery dan Santoso D Budi. (2016). Alogaritma Partitioning Around Method (PAM) Clustering untuk Melihat Gambaran Umum Skripsi Mahasiswa. *Jurnal Teknologi Informasi*, Vol 21 No 1.
- G. D. Rembulan, T. Wijaya, D. Palullungan, K. N. Alfina, dan M. Qurthuby, (2020) “Kebijakan Pemerintah Mengenai Coronavirus Disease (COVID-19) di Setiap Provinsi di Indonesia Berdasarkan Analisis Klaster Government Policy Regarding Coronavirus Disease (COVID-19) in Each Province in Indonesia Based on Clustering Analysis,” *J. Ind. Eng. Manag. Syst.*, vol. 13, no. 2, hal. 74–86.
- Gio Prana Ugiana dan Irawan Dasapta Erwin. (2016). *Belajar Statistik dengan R*. Pusat Sistem Informasi (PSI) Kampus USU Jl. Universitas No. 9 Medan 20155, Indonesia.

- Gugus Tugas Percepatan Penanganan COVID19, "Peta Sebaran Covid-19," covid19.go.id, 2020. [Daring]. Tersedia pada: <https://covid19.go.id/petasebaran>. [Diakses: 28-Jul-2020].
- Hrisnanto, Yulison Herry, 2013, *Algoritma Partitioning Around Medoids (PAM) Sebagai Teknik Clustering Pada Data Mining*, Majalah Ilmiah Aristoteles FMIPA Unjani, Volume 11 Nomor 2.
- <https://www.statistikian.com/2014/03/analisis-cluster.html>
- Informasi Coronavirus (COVID-19) Indonesia.* (2020). Retrieved from [s https://covid19.go.id/peta-sebaran](https://covid19.go.id/peta-sebaran)
- Iski Zaliman, Tri Basuki Kurniawan, Darius Antoni. (2020). Sistem Penentuan Lokasi Pusat Layanan Terpadu Bagi Penderita Penyakit Demam Berdarah Dengan Menggunakan K-Means Clustering. *Jurnal Ilmiah Informatika Global*. Vol. 11, No 2.
- Jakarta Open Data, 2020. "Data Kasus Positif Covid19 Provinsi DKI Jakarta Tahun 2020," data.jakarta.go.id, [Daring]. Tersedia pada: <https://data.jakarta.go.id/dataset/dat akasus-positif-covid-19-dki-jakarta>. [Diakses].
- Muhammad Syarifuddin (2020). Analisis Sentimen Opini Publik Mengenai Covid-19 pada Twitter Menggunakan Metode Naive Bayes dan KNN. *Jurnal Inti Nusa Mandiri*. Vol. 15. No 1.
- Mujab Akhmal Fajrul (2020). Pengelompokan Penderita Penyakit Covid-19 Dengan Metode K-Means (Studi Kasus Jawa Timur). *Skripsi*. Program Studi Teknik Informatika-S1 Fakultas Teknologi Industri Institut Teknologi Nasional Malang.
- Riadi Rini Aprianty. (2020). Analisis PCA Menggunakan Rstudio <https://medium.com/@17611063/analisis-pca-menggunakan-rstudio-3201c252badb>. 28 Maret 2021 (09.00)
- Rika Elizabet Sihombing, Dewi Rachmatin, Jarnawi Afgani Dahlan, (2019), Program Aplikasi Bahasa R Untuk Pengelompokan Objek Menggunakan Metode K-Medoids Clustering, *Jurnal Eurekamatika*, Vol. 7, No. 1
- Sangga, V. A. P. (2018b). Perbandingan K-Means Dengan K-Medoids Dalam Pengelompokan Komoditas Peternakan Di Jawa Tengah Pada 2015. *Journal Of Visual Languages & Computing* (Vol. 11).
- Setiawati A Widiastuti. (2017). *Implementasi Algoritma Partitioning Around Method (PAM) untuk Pengelompokan Sekolah Menengah Atas berdasarkan Nilai Daya Seray Ujian Nasional*. Program Studi Teknik Informatika, Universitas Sanata Dharma, Yogyakarta.
- Sindi, S. Ningse, W. R. O. Sihombing, I. A. P. P. P. A. N. W. Zer, F. I. R. R. Hartama, D. (2020). Analisis Algoritma K-

- Medoids Clustering dalam Pengelompokan Penyebaran Covid-19 di Indonesia. *Jurnal Teknologi Informasi, Vol 4* No 1.
- Sitepu, R., Irmeilyana, I., & Gultom, B. (2011). Analisis Cluster Terhadap Tingkat Pencemaran Udara Pada Sektor Industri Di Sumatera Selatan. *Jurnal 53 Penelitian Sains*, 14(3), 168311.
- Solichin Achmad , Khairunnisa Khansa. (2020). Klasterisasi Persebaran Virus Corona (Covid-19) di DKI Jakarta Menggunakan Metode K-Means. *Fountain of Informatics Journal*. Vol 5, No 2.
- Wibisono, Yudi, 2011, Perbandingan Partitioning Around Medoids (PAM) dan k-Means Clustering Untuk Tweets, KNSI2011.
- WHO, “WHO Coronavirus Disease (COVID19) Dashboard,” WHO, 2020. [Daring]. Tersedia pada: <https://covid19.who.int/>. [Diakses: 17-Sep-2020].
- Wulandari S dan Dwitiyanti N. (2020). Implementasi Alogaritma Clustering Partitioning Around Medoids (PAM) dalam clustering virus Mers-CoV. *Satuan Tulisan Riset dan Inovasi Teknologi*, Vol. 5 No 1
- Yulia Darmi , Agus Setiawan. (2016). Penerapan Metode Clustering K-Means Dalam Pengelompokan Penjualan Produk. *Jurnal Media Infotama*. Vol. 12 No. 2.
- Yang, F., Sun, T., & Zhang, C. (2009). An Efficient Hybrid Data Clustering Method Based On K-Harmonic Means And Particle Swarm Optimization. *Expert Systems With Applications*, 36(6), 9847–9852. <https://doi.org/10.1016/j.eswa.2009.02.003>
- Yeh, W., Jiang, Y., Chen, Y.-F., & Chen, Z. (2016). A New Soft Computing Method For K-Harmonic Means Clustering. *Plos One*, 11(11), E0164754. <https://doi.org/10.1371/journal.pone.0164754>
- Zhang, B., Hsu, M., & Dayal, U. (1999). Clustering Algorithm K-Harmonic Means -A Data Clustering Algorithm. *Hewlett-Packard Labs Technic*