

IMPLEMENTASI METODE *K-MEANS* UNTUK PENGELOMPOKAN KASUS COVID-19 TINGKAT PROVINSI DI INDONESIA

Andreas Nugroho Sihananto¹⁾, Anggraini Puspita Sari²⁾, Heri Khariono³⁾, Rifky Akhmad
Fernanda⁴⁾, ⁵⁾Devan Cakra Mudra Wijaya¹⁾

Email: ¹⁾andreas.nugroho.jarkom@upnjatim.ac.id, ²⁾anggraini.puspita.if@upnjatim.ac.id,
³⁾herikhariono7@gmail.com, ⁴⁾rifyakhmad911@gmail.com, ⁵⁾mudrawijaya@gmail.com

^{1,2,3,4,5} Informatika, Fakultas Ilmu Komputer, Universitas Pembangunan Nasional “Veteran” Jawa
Timur

Abstrak

Pada tanggal 30 Januari 2020, WHO secara global menerangkan bahwa status wabah COVID-19 menjadi kesehatan darurat global terpusat (PHEIC). COVID-19 merupakan mikroorganisme RNA (120-160 nm) yang dapat menginfeksi manusia dan hewan melalui berbagai macam penularan yang mungkin terjadi jika berada di ruang lingkup yang tidak ideal. Pada tahun 2019 sampai dengan saat ini, terdapat berbagai macam kasus dan mutasi virus COVID-19. Adanya kenaikan dan penurunan kasus COVID-19 di seluruh dunia terutama Indonesia perlu diwaspadai karena dapat berdampak pada kesejahteraan masyarakat dan sangat mempengaruhi kinerja dalam suatu negara. Oleh karena itu, pada penelitian ini dilakukan penggolongan terhadap perkembangan kasus COVID-19 yang ada di seluruh provinsi Indonesia sehingga penanganan COVID-19 lebih tepat sasaran. Metode yang digunakan dalam penelitian ini ialah *k-means clustering* dan *silhouette index* (SI). Metode *k-means clustering* dipakai untuk klusterisasi data COVID-19 dan metode SI dipakai untuk uji keakurasian klusterisasi data COVID-19. Tujuan dari penelitian ini untuk mengetahui dan mengidentifikasi provinsi mana di Indonesia yang memiliki tingkat kasus penyebaran COVID-19 sedang dan tinggi sehingga dapat dilakukan tindak lanjut dan penanganan yang lebih intensif untuk mengatasi lonjakan kasus COVID-19 kedepannya. *Dataset* yang digunakan dalam penelitian ini didapatkan dari sumber Kaggle. Berdasarkan hasil dari pengujian algoritma *k-means clustering* dinyatakan bahwa uji validasi menunjukkan nilai SI dengan tingkat akurasi sebesar 0,857 atau 85,7%, artinya bahwa tingkat keakurasian data mendekati nilai sebenarnya.

Kata Kunci: *k-means, clustering, COVID-19, machine learning*

1. PENDAHULUAN

Pertama kali kemunculan COVID-19 terjadi pada bulan Desember 2019 di Wuhan, provinsi Hubei yang pada akhirnya menjangkit secara cepat ke seluruh China bahkan ke seluruh dunia. WHO secara global menerangkan bahwa status wabah COVID-19 menjadi kesehatan darurat global terpusat (PHEIC) pada tanggal 30 Januari 2020[1]. COVID-19 merupakan mikroorganisme RNA (120-160 nm) yang sangat menular. Penularan COVID-19 dapat terjadi jika berada di ruang lingkup yang tidak ideal [2]. Penularan COVID-19 biasanya melalui percikan bersin atau batuk. Selain itu, COVID-19 dapat tertular melalui benda-benda yang terpapar *droplet* penderita COVID-19 yang tidak sengaja terkena mata, hidung, atau mulut. Virus COVID-19 dapat bertahan di permukaan benda selama beberapa jam, sedangkan di udara sekitar satu jam [3]. Virus ini tergolong ganas hingga dapat menyebabkan kematian dan hanya menyerang hewan serta manusia [4].

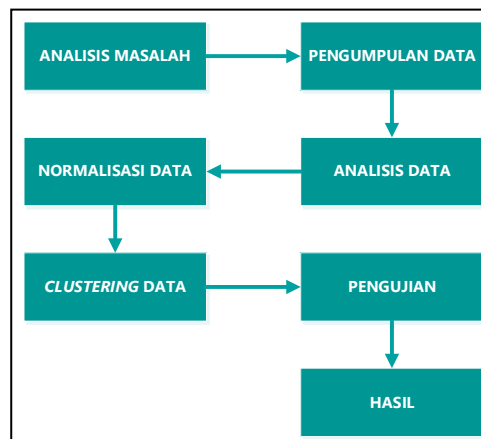
Pada akhir tahun 2019 secara internasional kasus sembuh dari COVID-19 kurang lebih sebesar 84 ribu dan kematian dari COVID-19 kurang lebih sebesar 4,6 ribu, sedangkan di Indonesia belum ada kasus COVID-19. Pada tahun 2020 secara internasional kasus sembuh dari COVID-19 sekitar 57 juta dan kematian dari COVID-19 sekitar 1,8 juta. Sedangkan di Indonesia kasus sembuh dari COVID-19 kurang lebih 604 ribu dan kematian dari COVID-19 kurang lebih 22 ribu. Pada tahun 2021 secara global kasus sembuh dari COVID-19 sebesar 266 juta dan kematian dari COVID-19 sebesar

5,26 juta. Sedangkan di Indonesia kasus sembuh dan kematian dari COVID-19 masih terus meningkat [5][8].

Adanya kenaikan kasus COVID-19 yang drastis di dunia terutama Indonesia perlu diwaspadai karena dapat berdampak pada kesejahteraan masyarakat dan sangat mempengaruhi kinerja dalam suatu negara. Apabila terjadi sedikit kelalaian dalam pengambilan keputusan bisa berakibat fatal dan dapat menimbulkan banyak korban baik merugikan secara finansial maupun jiwa. Oleh karena itu, pada penelitian ini dilakukan pengelompokan data terhadap perkembangan kasus COVID-19 yang terbatas pada lingkup seluruh provinsi di Indonesia sehingga diharapkan penanganan COVID-19 di Indonesia diharapkan dapat tepat sasaran. Tujuan utama dari penelitian ini yaitu untuk mengetahui provinsi mana yang memiliki tingkat kasus penyebaran COVID-19 tinggi dan sedang di Indonesia sehingga dapat dilakukan tindakan preventif untuk mengatasi lonjakan kasus COVID-19.

2. METODOLOGI

Metode pada penelitian ini mencakup beberapa tahapan yang dikerjakan secara berurutan yang dapat dilihat pada Gambar 1. Adapun penjelasan Gambar 1 dapat diketahui sub-bab sebagai berikut.



Gambar 1. Tahapan Penelitian

2.1 Analisis Masalah

Analisis masalah merupakan proses atau tahapan dalam kegiatan analisis yang berfungsi untuk mengetahui penyebab dari suatu permasalahan yang ada pada penelitian.

2.2 Pengumpulan Data

Pengumpulan data merupakan proses atau tahapan yang bertujuan untuk mendapatkan informasi berdasarkan sumber yang valid, misal: jurnal ilmiah, repositori, *website*, buku, *dataset*, dan sebagainya. Hal tersebut berguna sebagai pendukung dalam suatu penelitian [9].

2.3 Analisis Data

Analisis data merupakan proses atau tahapan dalam mengamati suatu kumpulan data lalu mencari data tertentu yang sesuai untuk digunakan dalam suatu penelitian, misal: melakukan pengamatan pada parameter tertentu.

2.4 Normalisasi Data

Normalisasi data merupakan proses atau tahapan dalam mengurangi *redundancy data* atau meminimalkan anomali yang ada pada data utama sehingga didapatkan data yang lebih sederhana dan lebih akurat untuk digunakan dalam berbagai keperluan yang

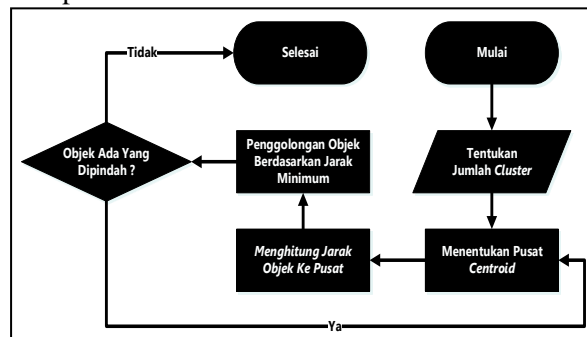
spesifik, misal: dipakai pada implementasi *clustering* data [10]. Pada penelitian ini menggunakan metode normalisasi data *min-max normalizer*. *Min-max normalizer* merupakan metode dalam normalisasi data yang umum dipakai untuk mendapatkan kestabilan nilai *compare* antardata sebelum dan sesudah proses. Hal tersebut dilakukan dengan menerapkan fungsi transformasi linier terhadap data utama [11]. Selain itu metode ini dapat melakukan skala data ulang dari *range* satu ke *range* baru lainnya dalam rentang 0 dan 1 [12]. Rumus *min-max normalizer* dapat dilihat pada persamaan 1.

$$X_{scaled} = \frac{(X - X_{min}) \cdot (NX_{max} - NX_{min})}{(X_{max} - X_{min}) + NX_{min}} \quad (1)$$

Pada persamaan (1) dapat diketahui bahwa X_{scaled} merupakan data dari hasil normalisasi *min-max normalizer*. Dimana X merupakan data utama, X_{min} merupakan data minimum dari data per kolom, X_{max} merupakan data maksimum dari data per kolom, NX_{min} merupakan batas minimum yang ditetapkan, dan NX_{max} merupakan batas maksimum yang ditetapkan.

2.5 Clustering Data

Clustering data merupakan proses atau tahapan dalam mengelompokkan suatu data menjadi beberapa bagian tertentu untuk memperoleh hasil yang mempunyai tipe atau jenis yang sama [13][14]. Pada tahapan ini menggunakan metode *k-means clustering* yang merupakan salah satu teknik yang dapat mengakomodasi partisi data berdasarkan titik *centroid* terdekat [15]. Tahapan-tahapan dari metode *k-means clustering* yang digunakan dapat dilihat pada Gambar 2.



Gambar 2 Tahapan K-Means Clustering

Adapun rumus Euclidean pada *k-means clustering* digunakan untuk menghitung jarak ke- k (X_i) pada pusat *cluster* ke- k (C_k) yang diberi nama (d_{ij}) yang dapat dilihat pada persamaan 2.

$$d_{ij} = \sqrt{\sum_{j=1}^m (X_{ij} - C_{kj})^2} \quad (2)$$

Kelebihan dari metode *k-means clustering* yaitu sebagai berikut [16][18]:

1. Mudah dalam pengaplikasiannya.
2. Waktu yang diperlukan dalam menjalankan *learning rate* tergolong cepat.
3. Kemudahan dalam beradaptasi.
4. Metode yang digunakan tergolong umum sehingga banyak referensi.
5. Menggunakan prinsip yang sederhana sehingga dapat dijelaskan dalam non-statistik.

Sedangkan kekurangan dari metode *k-means clustering* adalah sebagai berikut [18]:

1. Inisialisasi dilakukan secara random sehingga penggolongan data yang didapatkan ada kemungkinan dapat berubah-ubah sehingga hasilnya tidak optimal.
2. Ditemui masalah *curse of dimensionality* jika data pelatihan mempunyai dimensi yang sangat tinggi.
3. Perhitungan dan pencarian titik terdekat memerlukan waktu yang tidak sedikit. Hal tersebut terjadi jika terdapat banyak titik data.

2.6 Pengujian

Pengujian merupakan proses dalam menentukan tingkat validasi suatu sistem berfungsi dengan baik atau tidak. Pada tahapan ini menggunakan metode *Silhouette Index* (SI). SI adalah metode validitas klasterisasi berbasis kriteria internal dengan penilaian susunan objek pada setiap *cluster* dilakukan dengan membandingkan jarak rata-rata objek dalam *cluster* yang sama atau sebaliknya. Rumus untuk mendapatkan nilai SI ditunjukkan pada persamaan 3 berikut ini [19].

$$SI = \frac{1}{n} \sum_{i=1}^n \left(\frac{b(i) - a(i)}{\max\{a(i), b(i)\}} \right) \quad (3)$$

Dengan:

1. $a(i)$ = jarak rata-rata sampel i ke sampel lain dalam *cluster*.
2. $b(i)$ = jarak minimum sampel dari sampel i ke *cluster* lain.

3. HASIL DAN PEMBAHASAN

3.1. Analisis Masalah

Masalah yang ada pada penelitian ini berupa kasus COVID-19 yang hingga saat ini masih terjadi di Indonesia. Untuk mengantisipasi beberapa hal yang berada diluar dugaan, maka diperlukan proses *cluster*. Hal tersebut dilakukan untuk mengetahui provinsi mana yang memiliki tingkat kasus penyebaran COVID-19 sedang dan tinggi.

3.2. Pengumpulan Data

Sumber data utama (*dataset*) yang dipakai dalam penelitian ini ialah data COVID-19 di Indonesia yang sumbernya didapatkan dari *platform* kaggle. *Dataset* tersebut memiliki data sebanyak 20.817 yang terdiri dari 34 provinsi di Indonesia. Data tersebut terkumpul antara akhir Oktober hingga awal November 2021 yang ditunjukkan dalam Tabel 1.

3.3. Analisis Data

Data sampel awal yang dipakai merupakan hasil dari pengelompokan kasus COVID-19 di Indonesia yang terjadi akhir Oktober sampai awal November 2021. Hal tersebut didapatkan 2 macam jenis data yang digunakan dalam pengembangan penelitian yang terdiri dari total kasus dan total kematian yang ditunjukkan dalam Tabel 1.

Tabel 1. Dataset

Tanggal	Provinsi	Total Kasus	Total Kematian
10/29/2021	Gorontalo	11.830	460
11/1/2021	Maluku	14.571	260
11/2/2021	Bengkulu	23.099	472
11/3/2021	Jambi	29.745	789
11/4/2021	Sulawesi Tenggara	20.136	527
11/5/2021	Aceh	38.347	2.056
11/5/2021	Bali	113.959	4.028
11/5/2021	Banten	132.435	2.686
11/5/2021	DKI Jakarta	862.063	13.582
11/5/2021	Daerah Istimewa Yogyakarta	156.067	5.250
11/5/2021	Jawa Barat	706.376	14.700
11/5/2021	Jawa Tengah	485.429	30.147
11/5/2021	Jawa Timur	398.543	29.631
11/5/2021	Kalimantan Barat	41.180	1.056
11/5/2021	Kalimantan Selatan	69.853	2.388
11/5/2021	Kalimantan Tengah	45.553	1.423
11/5/2021	Kalimantan Timur	158.022	5.447
11/5/2021	Kalimantan Utara	35.823	802
11/5/2021	Kepulauan Bangka Belitung	52.069	1.448
11/5/2021	Kepulauan Riau	53.856	1.745
11/5/2021	Lampung	49.588	3.822
11/5/2021	Maluku Utara	12.086	303
11/5/2021	Nusa Tenggara Barat	27.726	812
11/5/2021	Nusa Tenggara Timur	63.707	1.331
11/5/2021	Papua	34.224	556
11/5/2021	Papua Barat	23.170	355
11/5/2021	Riau	128.499	4.103
11/5/2021	Sulawesi Barat	12.337	345
11/5/2021	Sulawesi Selatan	109.750	2.234
11/5/2021	Sulawesi Tengah	47.056	1.595
11/5/2021	Sulawesi Utara	34.636	1.039
11/5/2021	Sumatera Barat	89.786	2.147
11/5/2021	Sumatera Selatan	59.878	3.068
11/5/2021	Sumatera Utara	105.888	2.888

3.4. Normalisasi Data

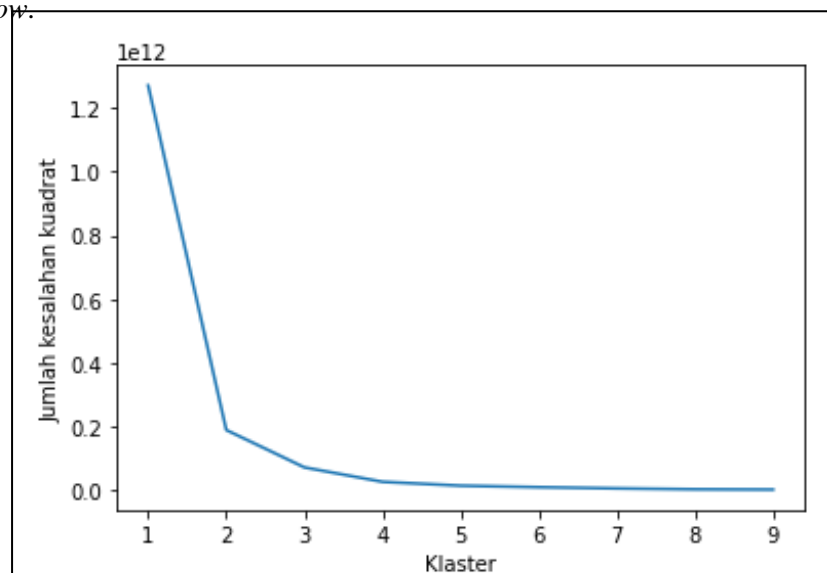
Berdasarkan Tabel 1 selanjutnya dilakukan normalisasi data dengan menggunakan *min-max normalizer* (lihat persamaan (1)) agar didapatkan *range* nilai antara 0 - 1 sehingga pengelompokan data dapat dilakukan lebih baik. Hasil normalisasi data ditunjukkan dalam Tabel 2.

Tabel 2. Hasil Normalisasi Data	
<i>X normalization</i>	<i>Y normalization</i>
1,00000000	0,98661625
0,99355236	1,00000000
0,97349197	0,98581323
0,95785861	0,96459999
0,98046183	0,98213270
0,93762416	0,87981397
0,75976232	0,74785024

0,71630130	0,83765517
1,00000000	0,10850872
0,66071183	0,66607555
0,63377803	0,03369358
0,11404521	1,00000000
0,09033641	0,96546994
0,93096010	0,94673269
0,86351271	0,85759695
0,92067351	0,92217352
0,65611309	0,65289256
0,94356135	0,96373005
0,90534595	0,92050055
0,90114239	0,90062569
0,91118199	0,76163549
0,99939781	0,99712249
0,96260790	0,96306086
0,87796992	0,92833004
0,94732268	0,98019206
0,97332496	0,99364272
0,72555993	0,74283133
0,99880739	0,99431191
0,76966314	0,86790243
0,91713801	0,91066350
0,94635353	0,94787031
0,81662438	0,87372436
0,88697686	0,81209221
0,77874771	0,82413758

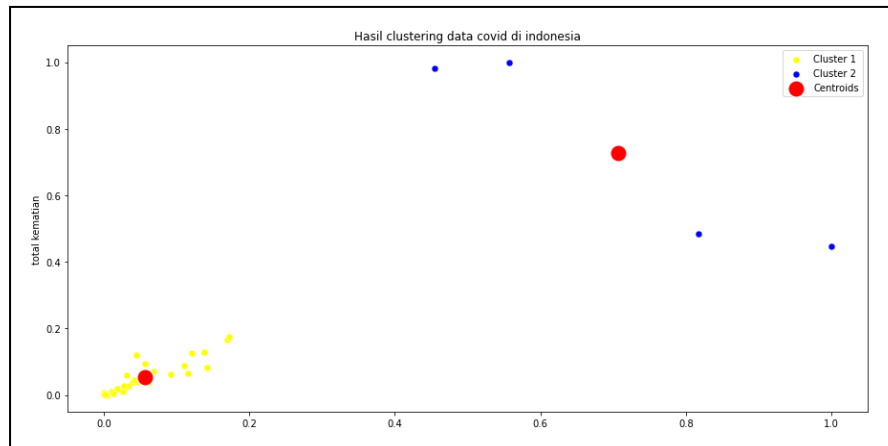
3.5. K-means Clustering

Untuk menentukan jumlah k (*cluster*) optimal dari data yang akan dilakukan *clustering* dengan *k-means* menggunakan teknik *elbow*. Gambar 3 menunjukkan hasil dari teknik *elbow*.



Gambar 3. Hasil metode *elbow*

Nilai *cluster* yang digunakan sebagai *cluster* optimal dalam metode ini adalah titik yang membentuk siku. Berdasarkan Gambar 3, *cluster* yang digunakan yaitu 2 dengan titik tersebut mengalami penurunan yang signifikan. Dari nilai *k* tersebut, dapat dilakukan klusterisasi data dengan *k-means clustering* yang dapat dilihat pada Gambar 4.



Gambar 4. Visualisasi hasil *clustering*

Tabel 3. Titik *Centroid*

<i>Cluster</i>	X	Y
C0	0,05645417	0,05312789
C1	0,70718585	0,72790846

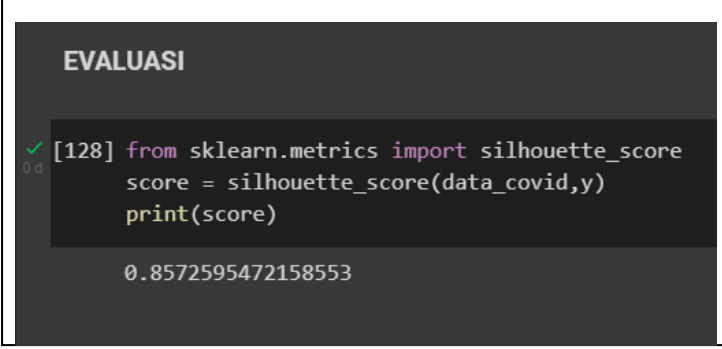
Pada Tabel 3, *cluster* C0 memiliki titik *centroid* $x=0,05645417$ dan $y=0,05312789$. Sedangkan *cluster* C1 memiliki titik *centroid* $x=0,70718585$ dan $y=0,72790846$. Dari Tabel 3, data dikelompokkan menjadi 2 *cluster* yaitu meliputi 4 provinsi dengan kasus penyebaran COVID-19 tinggi dan 30 provinsi dengan kasus penyebaran COVID-19 sedang yang dapat dilihat pada Tabel 4. C0 menunjukkan kasus penyebaran COVID-19 sedang dan C1 menunjukkan kasus penyebaran COVID-19 tinggi.

Tabel 4. Hasil Pengelompokkan Data

Provinsi	Cluster
Gorontalo	C0
Maluku	C0
Bengkulu	C0
Jambi	C0
Sulawesi Tenggara	C0
Aceh	C0
Bali	C0
Banten	C0
DKI Jakarta	C1
Daerah Istimewa Yogyakarta	C0
Jawa Barat	C1
Jawa Tengah	C1
Jawa Timur	C1
Kalimantan Barat	C0
Kalimantan Selatan	C0
Kalimantan Tengah	C0
Kalimantan Timur	C0
Kalimantan Utara	C0
Kepulauan Bangka Belitung	C0
Kepulauan Riau	C0
Lampung	C0
Maluku Utara	C0
Nusa Tenggara Barat	C0
Nusa Tenggara Timur	C0
Papua	C0
Papua Barat	C0
Riau	C0
Sulawesi Barat	C0
Sulawesi Selatan	C0
Sulawesi Tengah	C0
Sulawesi Utara	C0
Sumatera Barat	C0
Sumatera Selatan	C0
Sumatera Utara	C0

3.6. Pengujian Data

Untuk menghitung tingkat keakurasian *k-means clustering* digunakan metode SI yang ditunjukkan dalam Gambar 5.



```
EVALUASI
[128] from sklearn.metrics import silhouette_score
      score = silhouette_score(data_covid,y)
      print(score)
0.8572595472158553
```

Gambar 5 Pengujian SI

Berdasarkan Gambar 5 dapat diketahui bahwa pengujian dengan metode *k-means clustering* didapatkan hasil akurasi sebesar 0.857 atau 85.7 %.

4. KESIMPULAN DAN SARAN

Berdasarkan hasil penelitian tersebut dapat ditarik kesimpulan bahwa masalah yang diteliti berupa pengelompokan kasus COVID-19 di tingkat provinsi di Indonesia berdasarkan tingkat penyebaran sedang dan tinggi menggunakan metode *k-means clustering* didapatkan 2 *cluster* yaitu 30 provinsi dengan kasus penyebaran COVID-19 sedang (C0) dan 4 provinsi dengan kasus penyebaran COVID-19 tinggi (C1). Setelah dilakukan evaluasi atau pengujian didapatkan tingkat keakurasian metode *k-means clustering* sebesar 0.857 atau 85.7%. Saran untuk penelitian kedepannya yaitu dapat dikembangkan dengan perpaduan dua algoritma, misalnya *k-means clustering* dengan KNN (k-nearest neighbor) sehingga diharapkan dapat meningkatkan tingkat akurasi data.

5. DAFTAR RUJUKAN

- [1] R. F. Nursofwa, M. H. Sukur, B. K. Kurniadi, dan Haris, 2020. Penanganan Pelayanan Kesehatan Di Masa Pandemi Covid-19 Dalam Perspektif Hukum Kesehatan. *Jurnal Hukum Inicio Legis*, 1 (1), pp. 1-17.
- [2] I. G. S. M. Diyasa, D. C. M. Wijaya, R. A. Ramadani, S. Handayani, dan M. Ikhlasurramadhan, 2021. Manajemen Informasi dan Konten Berbasis Digital Dengan Analisis SWOT: KKN Tlogopatut Gresik. *Seminar Nasional Sains Data 2021 (SENADA 2021)*, 1 (1), pp. 43-51.
- [3] M. K. Sari, 2020. Sosialisasi tentang Pencegahan Covid-19 di Kalangan Siswa Sekolah Dasar di SD Minggiran 2 Kecamatan Papar Kabupaten Kediri. *Jurnal Karya Abdi*, 4 (1), pp. 80-83.
- [4] I. Wahidah dkk, 2020. Pandemi Covid-19: Analisis Perencanaan Pemerintah dan Masyarakat dalam Berbagai Upaya Pencegahan COVID-19 Pandemic: Analysis of Government and Community Planning in Various Prevention Measures. *Jurnal Manajemen. dan Organisasi (JMO)*, 11 (3), pp. 179-188.
- [5] Coronavirus Pandemic (COVID-19) – the data, *Our World in Data (2019-2021)*. [Online]. Available: <https://ourworldindata.org/coronavirus-data>. [Accessed: 06-Dec-2021].
- [6] COVID-19 Data Repository by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University, *Johns Hopkins University (2019-2021)*. [Online]. Available: <https://github.com/CSSEGISandData/COVID-19>. [Accessed: 06-Dec-2021].
- [7] Coronavirus in the U.S.: Latest Map and Case Count, *The New York Times (2019-2021)*. [Online]. Available: <https://www.nytimes.com/interactive/2021/us/covid-cases.html>. [Accessed: 06-Dec-2021].
- [8] COVID-19 pandemic by country and territory, *Wikipedia (2019-2021)* [Online]. Available: https://en.wikipedia.org/wiki/COVID-19_pandemic_by_country_and_territory. [Accessed: 06-Dec-2021].
- [9] K. A. Marpaung, D. C. M. Wijaya, and A. Agussalim, 13 November 2021. Desain dan Manajemen Jaringan (Studi Kasus: Fakultas Ilmu Komputer UPN Veteran Jatim). *Prosiding Seminar. SITASI*, 1 (1), pp. 330-340.
- [10] Z. Efendy, 2018. Normalisasi Dalam Desain Database. *Jurnal Corel IT*, 4 (1), pp. 34-43.
- [11] D. A. Nasution, H. H. Khotimah, and N. Chamidah, 2019. Perbandingan Normalisasi Data untuk Klasifikasi Wine Menggunakan Algoritma K-NN. *Journal of Computer Engineering, Science, and System*, 4 (1), pp. 78-82.
- [12] N. Chamidah, . W., dan U. Salamah, 2012. Pengaruh Normalisasi Data pada Jaringan Syaraf Tiruan Backpropagasi Gradient Descent Adaptive Gain (BPGDAG)

- untuk Klasifikasi. *IT Smart : Jurnal Teknologi Informasi*, 1 (1), pp. 28-33.
- [13] E. P. Lestari, A. Mahmudi, and S. Achmadi, 2020. Penerapan Metode K-Means Untuk Proses Penentuan Golongan Penerima Zakat (Mustahiq) di Yayasan Dana Sosial Al-Falah Malang,” *JATI (Jurnal Mahasiswa Teknologi Informasi)*, 4 (1), pp. 389-396.
- [14] M. A. Rheza dan F. Metandi, 2020. Implementasi Metode K-Means Clustering Untuk Penentuan Jenis Komentar Pada Tweet PSSI, *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, 12 (2), pp. 73-78.
- [15] T. R. Tulili, K. B. Utomo, dan D. Lestari, 2019. Pengelompokan Sekolah Dasar Di Kota Samarinda Berdasarkan Indikator Standar Pelayanan Minimal Pendidikan Menggunakan Metode K-Means. *Just TI (Jurnal Sains Terapan Teknologi Informasi)*, 11 (2), pp. 24-28.
- [16] B. Harahap, 2019. Penerapan Algoritma K-Means Untuk Menentukan Bahan Bangunan Laris (Studi Kasus Pada UD. Toko Bangunan YD Indarung). *Jurnal READY STAR-2*, 1 (1), pp. 394-403.
- [17] D. H. Dailamy, P. D. Kusuma, dan A. L. Prasasti, 2018. Clustering Produksi Barang Pada Lalu Lintas Angkutan Udara Menggunakan Algoritma K-Means. *e-Proceeding Engineering*, 6 (1), pp. 1424-1431.
- [18] Y. Kusumawardani, A. Hamzah, dan Suraya, 2018. Perbandingan Metode Clustering Menggunakan Hierarchical Clustering dan Partitional Clustering untuk Mengelompokkan Dokumen Berita. *Jurnal Script*, 5 (2), pp. 8-13.
- [19] R. Adha, N. Nurhaliza, U. Soleha, dan Mustakim, 2021. Perbandingan Algoritma DBSCAN dan K-Means Clustering untuk Pengelompokan Kasus Covid-19 di Dunia. *Jurnal Sains, Teknologi dan Industri*, 18 (2), pp. 206–211.