

Klastering Wilayah Kota/Kabupaten Berdasarkan Data Persebaran Covid-19 di Propinsi Jawa Timur dengan Metode K-Means

Wiyli Yustanti¹, Naim Rahmawati², Yuni Yamasari³

^{1,2,3} Jurusan Teknik Informatika, Fakultas Teknik, Universitas Negeri Surabaya
wiyliyustanti@unesa.ac.id

Abstrak— Klastering persebaran pandemi COVID-19 di Jawa Timur dapat dilakukan dengan melihat dinamika data harian yang disajikan dalam situs resmi pemerintah Jawa Timur terkait informasi COVID-19. Sumber data untuk proses klastering adalah data jumlah orang dalam resiko (ODR), orang tanpa gejala (OTG), orang dalam pemantauan (ODP), pasien dalam pengawasan (PDP) dan jumlah pasien positif terpapar virus COVID-19 (*Confirm*) untuk seluruh kabupaten/kota di Propinsi Jawa Timur. Melalui algoritma klastering Non-Hirarki K-Means didapatkan bahwa jumlah klaster optimum adalah 5 klaster. Untuk membuktikan bahwa kelima klaster yang dibentuk ini berbeda secara signifikan maka dilakukan uji mean vektor dengan statistic *Wilks Lambda* dan dihasilkan perbedaan yang signifikan dengan tingkat kepercayaan 95%. Karakteristik klaster untuk setiap kelompok dapat disimpulkan secara umum adalah klaster dengan kasus PDP dan *Confirm* yang sangat tinggi (*red zone*), klaster dengan kasus jumlah *Confirm* Tinggi (*Orange Zone*), klaster dengan jumlah PDP Tinggi (*Yellow Zone*), klaster dengan jumlah kasus PDP dan *Confirmed* Sedang (*Green Zone*) dan klaster dengan PDP dan *Confirm* rendah (*Light Green Zone*).

Kata Kunci— *K-Means*, COVID-19, *Clustering*

I. PENDAHULUAN

Tidak bisa dipungkiri bahwa penyebaran wabah virus Corona atau COVID-19 terus bergerak signifikan di Indonesia. Sejak mulai terdeteksi pada awal Maret lalu hingga jelang pertengahan Mei, jumlah kasus positif COVID-19 sudah mencapai angka 15.438, angka tersebut dikutip dari laman resmi pemerintah untuk COVID-19 (<https://covid19.go.id/>) pada tanggal 13 Mei 2020, jumlah kasus sembuh mencapai 3.287 (21,29 persen) sedangkan angka pasien yang meninggal adalah 1.028 (6,67 persen). Khususnya di Propinsi Jawa Timur yang merupakan salah satu wilayah dimana penyebaran virus Corona juga terus menunjukkan pergerakan secara ekponensial dari hari ke hari dan belum menunjukkan pergerakan menurun sampai saat ini. Pada tanggal yang sama juga diinformasikan pada laman <http://infocovid19.jatimprov.go.id/> bahwa data kasus positif di propinsi Jawa Timur sudah mencapai 1766 dan jumlah kasus sembuh sebanyak 278 (15,74%) serta kasus meninggal sebanyak 166 (9,40%). Dengan menggunakan data-data yang disajikan dalam situs tersebut, maka dapat dianalisa tentang bagaimana sebenarnya peta penyebaran virus corona ini dari waktu ke waktu.

Klastering persebaran pandemi COVID-19 di Jawa Timur dapat dilakukan dengan melihat dinamika data harian yang disajikan dalam situs resmi pemerintah Jawa Timur terkait informasi COVID-19. Data harian tersebut akan menjadi inputan pada proses klastering, terdapat 5 variabel yaitu data jumlah orang dalam resiko (ODR), orang tanpa gejala (OTG), orang dalam pemantauan (ODP), pasien dalam pengawasan (PDP) dan jumlah pasien positif terpapar virus COVID-19 (*Confirm*) untuk seluruh kabupaten/kota di Propinsi Jawa Timur. Secara ilmiah, klastering dapat diartikan sebagai metode untuk melakukan pengelompokan dengan kriteria tertentu berdasarkan kemiripan karakteristik dari setiap objek yang akan dikelompokkan. Sehingga objek yang memiliki karakteristik yang hampir sama akan terkumpul dalam satu kelompok. Sementara itu, Sifat objek antar kelompok memiliki kesamaan yang rendah. Dengan kata lain, keragaman (variasi) dalam suatu kelompok kecil sedangkan keragaman antar kelompok tinggi. [7]. Untuk tujuan ini, maka digunakan sebuah metode klastering non-hirarki yaitu K-Means. Diharapkan bahwa hasil klastering dapat menunjukkan ada berapa jumlah kelompok wilayah yang berbeda secara signifikan dilihat dari data persebaran virus COVID-19. Pentingnya informasi tentang data klaster ini dapat membantu pemerintah untuk membuat strategi pencegahan penyebaran COVID-19 berdasarkan informasi hasil klastering wilayah kota/kab di Propinsi Jawa Timur. Tujuan dari penelitian ini adalah untuk mengetahui bagaimana hasil pengelompokan wilayah kabupaten/kota di Propinsi Jawa Timur berdasarkan data sebaran COVID-19 dengan menggunakan metode *K-Means*.

II. KAJIAN PUSTAKA

Dalam bagian ini akan dijelaskan mengenai beberapa definisi dan kajian teori yang terkait dengan tujuan penelitian ini.

A. Metode *K-Means*

Metode analisis klaster non hirarki yaitu *K-Means* bertujuan untuk mengelompokkan objek ke dalam klaster menurut kemiripan karakteristiknya, sehingga objek yang memiliki atribut-atribut yang mirip/sama dikelompokkan dalam satu klaster dan objek yang mempunyai atribut yang berbeda dikelompokkan ke dalam klaster yang berbeda. Secara umum metode *K-Means Cluster Analysis* [4] menggunakan algoritma sebagai berikut:

- 1) Tentukan K sebagai jumlah kluster yang terbentuk dengan mempertimbangkan teori atau konsep yang relevan sehingga dapat disepakati berapa banyak kluster yang ingin dibentuk.
- 2) Ambil sebanyak K titik pusat kluster. Penentuan titik pusat ini dapat dilakukan secara acak dari data untuk pertama kali, dan rumus berikut digunakan untuk menghitung titik pusat kluster berikutnya.:

$$y_i = \frac{\sum_{i=1}^n x_i}{n} \quad (1)$$

Keterangan:

y_i = centroid pada kluster

x_i = obyek pengamatan ke i

n = banyaknya obyek yang menjadi anggota kluster

- 3) Menghitung jarak menggunakan *Euclidian Distance* dengan persamaan sebagai berikut:

$$d = \frac{\sum_{i=1}^n (x_i - y_i)^2}{n} \quad (2)$$

Keterangan:

y_i = centroid pada kluster

x_i = obyek pengamatan ke i

n = banyaknya obyek yang menjadi anggota kluster

- 4) Setiap objek dinyatakan sebagai anggota kluster tersebut jika jaraknya terdekat dengan titik pusat kluster. Kemudian tentukan posisi *centroid* baru dengan menggunakan persamaan (1)
- 5) Kembali ke langkah 3 jika posisi *centroid* baru tidak sama.

Pemeriksaan konvergensi dilakukan dengan membandingkan matriks kelompok dalam iterasi sebelumnya dengan elemen matriks kelompok selama pengulangan. Jika hasilnya sama, algoritma K-mean selesai, tetapi jika berbeda, artinya belum konvergen, sehingga perlu dilakukan pengulangan sampai konvergen.

B. Pemilihan Jumlah Kluster Optimum

Pemilihan jumlah kelompok optimum menggunakan kriteria nilai *Pseudo F-statistics* dengan persamaan sebagai berikut:

$$Pseudo\ F - Statistics = \frac{R^2 / (k - 1)}{(1 - R^2) / (n - k)} \quad (3)$$

Dimana :

$$R^2 = \frac{(SST - SSW)}{SST}; \quad SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_j)^2; \quad SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_{jk})^2$$

SST (*Sum Square Total*) adalah total jumlah dari kuadrat jarak sampel terhadap rata-rata keseluruhan, sedangkan SSW (*Sum Square Within*) merupakan total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompoknya. Kemudian n adalah banyaknya sampel, c banyaknya variable, k adalah

banyaknya kelompok, x_{ijk} merupakan sampel ke- i pada variabel ke- j dan kelompok ke- k , \bar{x}_j adalah rata-rata seluruh sampel pada variabel ke- j dan \bar{x}_{jk} adalah rata-rata sampel pada variabel ke- j dan kelompok ke- k . Nilai *Pseudo F-statistics* tertinggi menunjukkan bahwa jumlah grup yang digunakan untuk mempartisi data paling tepat. Hal ini berarti bahwa variasi antar kelompok tinggi, dan didalam kelompok rendah [2].

C. Kriteria Kluster Terbaik

Untuk menentukan kriteria kluster terbaik dapat digunakan nilai *icdrate*. *Icdrate* (*internal cluster dispersion*) menggambarkan tingkat dispersi atau perbedaan dalam kluster. Nilai *icdrate* yang semakin kecil menunjukkan bahwa kluster tersebut semakin baik sebab antara anggota dalam satu kluster memiliki perbedaan yang rendah. Nilai *icdrate* dapat diperoleh dengan menggunakan persamaan [6]:

$$icdrate = 1 - \frac{SSB}{SST} \quad (4)$$

Dimana :

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (x_{ijk} - \bar{x}_j)^2; \quad SSB = \sum_{j=1}^c \sum_{k=1}^p (\bar{x}_{jk} - \bar{x}_j)^2$$

Dimana SST (*Sum Square Total*) merupakan total jumlah dari kuadrat jarak sampel terhadap rata-rata keseluruhan, dan SSB (*Sum Square Between*) adalah total jumlah dari kuadrat jarak rata-rata sampel pada variabel ke- j dan kelompok ke- k terhadap rata-rata seluruh sampel pada variabel ke- j , n adalah banyaknya sampel, c banyaknya variable, p banyaknya kelompok, x_{ijk} adalah sampel ke- i pada variabel ke- j dan kelompok ke- k , sedangkan \bar{x}_j adalah rata-rata seluruh sampel pada variabel ke- j serta \bar{x}_{jk} merupakan rata-rata sampel pada variabel ke- j dan kelompok ke- k .

D. One-way MANOVA

Pengujian MANOVA satu arah digunakan untuk membandingkan nilai vektor rata-rata antara perhitungan pada data multivariat. Dalam pengujian satu arah MANOVA, hanya satu faktor atau metode yang digunakan, dan tujuannya adalah untuk membandingkan perlakuan yang berbeda. Hipotesis yang digunakan dalam uji MANOVA satu arah ditunjukkan di bawah ini:

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_g$$

$$H_1 : \text{Minimal Ada Satu } \tau_l \neq 0, l = 1, 2, \dots, g$$

Adapun statistik uji yang digunakan adalah :

$$F_{hitung} = \frac{SS_{treatment} / (g - 1)}{SS_{residual} / (\sum_{i=1}^g n_i - g)} \quad (5)$$

Keterangan:

$SS_{treatment}$ = nilai *Sum of Square* perlakuan
 $SS_{residual}$ = nilai *Sum of Square* residual
 n_i = banyaknya data populasi ke 1
 g = banyaknya populasi

Tolak H_0 jika $F_{hitung} > F_{v_1, v_2, \alpha}$ dengan dan $v_1 = g - 1$ dan $v_2 = \sum_{i=1}^g n_i - g$. Pendekatan dengan statistik uji *Wilk's Lamda* juga dapat digunakan berdasarkan tabel uji *One-way MANOVA*

TABEL I
ONE-WAY MANOVA

Sumber Variasi	Matriks Jumlah Kuadrat (SS)	Derajat Bebas (db)
Treatment	$B = \sum_{i=1}^g n_i (\bar{x}_i - \bar{x})(\bar{x}_i - \bar{x})^T$	$g - 1$
Residual	$W = \sum_{i=1}^g \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)(x_{ij} - \bar{x}_i)^T$	$\sum_{i=1}^g n_i - g$
Total	$B + W$	$\sum_{i=1}^g n_i - 1$

Digunakan pendekatan *Wilk's Lamda* : $\Lambda = \frac{|W|}{|B + W|}$ dan

$$C = \left(\frac{\sum_{i=1}^g n_i - g - 1}{g - 1} \right) \left(\frac{1 - \sqrt{\Lambda}}{\sqrt{\Lambda}} \right)$$

Tolak H_0 jika $C > F_{v_1, v_2, \alpha}$ dengan dan $v_1 = g - 1$ dan $v_2 = v_2 = \sum_{i=1}^g n_i - g$. [4]

E. Epidemiologi

Menurut sejarah perkembangan, epidemiologi dibedakan atas epidemiologi klasik dan epidemiologi modern. Epidemiologi klasik merupakan ilmumempelajari tentang penyakit menular, wabah serta terjadinya penyakit. Sedangkan epidemiologi modern merupakan sekumpulan konsep yang digunakan dalam studi epidemiologi yang terutama bersifat analitik, selain untuk penyakit menular wabah dapat diterapkan juga untuk penyakit menular bukan wabah, penyakit tidak menular, serta masalah-masalah Kesehatan lainnya [3]. Dalam [3] juga dijelaskan bahwa istilah epidemi (wabah) di waktu lampau digunakan khusus untuk mendeskripsikan peristiwa berjangkitnya penyakit menular secara akut. Dalam UU Republik Indonesia No. 4 tahun 1984 tentang Wabah Penyakit Menular dan PP Republik Indonesia No. 40 tahun 1991 tentang Penanggulangan Wabah Penyakit Menular dinyatakan bahwa :

a Wabah (wabah penyakit menular) adalah kejadian berjangkitnya suatu penyakit menular dalam masyarakat yang jumlah penderitanya meningkat secara nyata

melebihi keadaan yang lazim pada wilayah dan periode tertentu serta dapat menimbulkan malapetaka.

b KLB (kejadian luar biasa) adalah timbulnya atau meningkatnya kejadian morbiditas/ mortalitas yang bermakna secara epidemiologis pada suatu wilayah dan periode tertentu, dan merupakan keadaan yang dapat menjurus pada terjadinya wabah.

Menurut teori dalam cabang epidemiologi deskriptif yang khusus mempelajari tentang kejadian dan distribusi penyakit. Umumnya distribusi penyakit dikelompokkan menurut faktor orang, tempat, dan waktu. Faktor orang dapat dilihat dari karakteristik orang antara lain usia, jenis kelamin, status pernikahan, suku/ras, status sosial ekonomi dan pekerjaan. Sedangkan faktor tempat dapat dilihat dari indikator wilayah pedesaan/perkotaan, faktor geografi, perbedaan standard layanan kesehatan dan mungkin tata kelola pelayanan kesehatan yang berbeda disetiap tempat. Sementara faktor waktu juga berkontribusi terhadap monitoring dan evaluasi penyebaran penyakit dengan melihat data jumlah kasus kematian, kesembuhan dan munculnya kasus baru dari waktu ke waktu. Dalam Penelitian ini, data yang digunakan adalah data yang diambil dalam waktu tertentu, sehingga hasil interpretasinya akan disandingkan dengan faktor kependudukan dan faktor wilayah yang secara teori epidemiologi merupakan variabel yang berpengaruh terhadap penyebaran penyakit (wabah).

F. Virus CORONA (COVID-19)

Corona virus termasuk dalam kelompok virus yang menyebabkan penyakit pada manusia dan hewan. Pada manusia biasanya menyebabkan penyakit infeksi saluran pernapasan, mulai flu biasa hingga penyakit yang serius seperti *Middle East Respiratory Syndrome* (MERS) dan Sindrom Pernafasan Akut Berat/ *Severe Acute Respiratory Syndrome* (SARS). *Corona virus* jenis baru yang ditemukan pada manusia sejak kejadian luar biasa muncul di Wuhan Cina, pada Desember 2019, kemudian diberi nama *Severe Acute Respiratory Syndrome Corona virus 2* (SARS-COV2), dan menyebabkan penyakit *Coronavirus Disease-2019* (COVID-19). COVID-19 disebabkan oleh SARS-COV2 yang termasuk dalam keluarga besar *corona virus* yang sama dengan penyebab SARS pada tahun 2003, hanya berbeda jenis virusnya. Gejalanya mirip dengan SARS, namun angka kematian SARS (9,6%) lebih tinggi dibanding COVID-19 (kurang dari 5%), walaupun jumlah kasus COVID-19 jauh lebih banyak dibanding SARS. COVID-19 juga memiliki penyebaran yang lebih luas dan cepat ke beberapa negara dibanding SARS.

Gejala umum yang terjadi akibat *Corona virus* antara lain demam $\geq 38^\circ\text{C}$, batuk kering, dan sesak napas. Jika ada orang yang dalam 14 hari sebelum muncul gejala tersebut pernah melakukan perjalanan ke negara terjangkit, atau pernah merawat/kontak erat dengan penderita COVID-19, maka terhadap orang tersebut akan dilakukan pemeriksaan laboratorium lebih lanjut untuk memastikan diagnosisnya. Seperti penyakit pernapasan lainnya, COVID-19 dapat

menyebabkan gejala ringan termasuk pilek, sakit tenggorokan, batuk, dan demam. Sekitar 80% kasus dapat pulih tanpa perlu perawatan khusus. Sekitar 1 dari setiap 6 orang mungkin akan menderita sakit yang parah, seperti disertai pneumonia atau kesulitan bernafas, yang biasanya muncul secara bertahap. Walaupun angka kematian penyakit ini masih rendah (sekitar 3%), namun bagi orang yang berusia lanjut, dan orang-orang dengan kondisi medis yang sudah ada sebelumnya (seperti diabetes, tekanan darah tinggi dan penyakit jantung), mereka biasanya lebih rentan untuk menjadi sakit parah.

Cara utama penularan penyakit ini adalah melalui tetes kecil (*droplet*) yang dilepaskan ketika seseorang batuk atau bersin. Saat ini, WHO menganggap bahwa risiko penularan oleh seseorang tanpa gejala COVID-19 kemungkinan kecil. Namun, banyak orang telah mengidentifikasi COVID-19 hanya memiliki gejala ringan, seperti batuk ringan, atau tidak mengeluh rasa sakit yang mungkin terjadi pada tahap awal penyakit. Sampai saat ini, para ahli melanjutkan penyelidikan untuk menentukan periode transmisi COVID-19 atau periode inkubasi. [5].

G. Penelitian Terkait Pemodelan Epidemiologi

Untuk menambah kedalaman interpretasi dari hasil klastering, perlu dikaji hasil penelitian sebelumnya terkait faktor-faktor yang berpengaruh terhadap penularan atau penyebaran suatu penyakit. Pada penelitian [8] menyebutkan bahwa faktor-faktor yang diduga mempengaruhi penyebaran penyakit Tuberkulosis (TB), yaitu persentase keluarga dengan kepemilikan sarana sanitasi dasar sehat, persentase penduduk miskin, persentase gizi buruk, persentase tenaga kesehatan terlatih dan jumlah rumah tangga ber-PHBS. Berdasarkan hasil pemodelan pada penelitian tersebut didapatkan model spline terbaik yaitu spline dengan kombinasi knot dengan nilai koefisien determinasi yang didapatkan adalah 83,42% dengan tiga variabel prediktor dalam model berpengaruh signifikan yaitu persentase gizi buruk masyarakat, persentase tenaga kesehatan terlatih TB, dan persentase rumah tangga ber-PHBS. Dalam penelitian yang lain [1] disimpulkan bahwa kelembaban udara mempunyai hubungan antara angka kejadian demam berdarah yang terjadi di Kabupaten Serang, dengan nilai $r=0,408$ dan $p\text{-value}= 0,007$.

Kemudian pada tahun 2020 beberapa peneliti juga sudah mempublikasikan kajian empirisnya terkait dengan faktor yang mempengaruhi penyebaran COVID-19. Dalam [9] menyimpulkan bahwa terdapat 11 faktor yang mempengaruhi epidemi COVID 19 di seluruh dunia melalui kajian literatur dan pendapat para ahli. Pendekatan TISM digunakan untuk membuat daftar faktor-faktor yang mempengaruhi epidemi COVID 19 dan untuk mengidentifikasi keterkaitan diantara mereka. Faktor-faktor ini diberi peringkat sesuai urutan kepentingan dan hierarki menggunakan analisis MICMAC. Hasil dari penelitian menunjukkan bahwa perubahan dalam setiap faktor individu yang disebutkan dalam penelitian ini dapat secara langsung atau tidak langsung mempengaruhi penyebaran virus. Kesenjangan sosial dan kesadaran masyarakat, usia, suhu udara, aliran udara dan ventilasi, kepadatan populasi dan kelembaban diidentifikasi sebagai

faktor pendorong atau kunci dalam model ini. Faktor-faktor ini secara independen dan saling tergantung ditemukan untuk memberikan dampak yang kuat pada epidemi dengan karakteristik SARS-CoV- 2. Perilaku host dan jumlah kontak serta praktik kebersihan pribadi juga ditemukan menjadi faktor keterkaitan antara hal-hal di atas. Jadi, manajemen hubungan yang efektif dapat secara efektif mengontrol penyebaran COVID-19 di seluruh dunia terlepas dari zona geografis atau iklim, usia, kepadatan penduduk, atau kekuatan sistem perawatan kesehatan. Dengan demikian, model ini adalah panduan untuk kebijakan pembuat layanan kesehatan dan pemerintah, yang secara tepat menunjukkan faktor epidemiologis COVID-19 yang berada di bawah kendali manusia dan membantu mereka memprioritaskan langkah darurat yang harus diambil untuk diterapkan aksi cepat tanggap.

Selain penelitian [9] ditemukan juga sebuah kajian analisis literature secara sistematis terkait dengan epidemi COVID-19 oleh Tan [10] menyimpulkan bahwa usia rata-rata pasien adalah 49,16 tahun. Sebanyak 51,46% pasien adalah laki-laki dan 52,32% bukan perokok. Hipertensi (50,82%) dan diabetes (20,89%) merupakan *komorbiditas* yang paling sering diamati. Gejala yang paling umum adalah demam (83,21%), batuk (61,74%), dan *mialgia* atau kelelahan (30,22%). Sebagian besar pasien (78,50%) pernah memiliki riwayat paru-paru bilateral, dan 5,86% menunjukkan tidak ada temuan CT yang mengindikasikan *pneumonia* virus. Selain itu terdapat sindrom gangguan pernapasan akut (28,36%), cedera jantung akut (7,89%), dan ginjal akut cedera (7,60%) merupakan komplikasi paling umum yang dicatat.

III. METODOLOGI

A. Sumber Data

Data yang digunakan pada penelitian ini merupakan data sekunder tentang variabel penyebaran COVID-19 di Propinsi Jawa Timur per tanggal 13 Mei 2020 yang diakses melalui laman resmi pemerintah Propinsi Jawa Timur dalam menyampaikan informasi kepada masyarakat terkait pergerakan harian jumlah kasus COVID-19 diberbagai kota di Jawa Timur (<http://infocovid19.jatimprov.go.id/>) selain itu juga data informasi tambahan terkait profil statistik Kesehatan dan kependudukan yang diakses dari publikasi Badan Pusat Statistik Provinsi Jawa Timur pada tahun 2020.

B. Variabel Penelitian

Variabel penelitian yang digunakan merupakan variabel pembentuk indikator pengelompokan wilayah terdampak pandemic COVID-19 meliputi variabel yang sudah di gunakan pada info grafis pada peta sebaran COVID-19 di Jawa Timur.

TABEL III
VARIABEL PENELITIAN

Variabel	Nama Variabel
X ₁	Jumlah Orang Dengan Resiko (ODR)
X ₂	Jumlah Orang Tanpa Gejala (OTG)
X ₃	Jumlah Orang Dalam Pemantauan (ODP)

X ₄	Jumlah Pasien Dalam Pengawasan (PDP)
X ₅	Jumlah Pasien Positive COVID-19 (CONFIRM)

Menurut definisi yang didapatkan pada laman resmi pemerintah terkait COVID-19, variabel-variabel diatas dapat dijelaskan sebagai berikut :

- a Variabel X₁ adalah jumlah ODR. ODR singkatan dari Orang Dengan Resiko, yaitu orang yang baru saja tiba dari daerah yang terinfeksi virus korona, yang merupakan *red zone* COVID-19. ODR belum tentu sakit dan masih terlihat seperti orang sehat.
- b Variabel X₂ adalah jumlah OTG. OTG singkatan dari Orang Tanpa Gejala adalah orang tanpa gejala tetapi yang memiliki kontak dekat dengan kasus positif COVID-19. Tujuan utamanya adalah untuk mengklasifikasikan OTG sehingga dapat dengan mudah mengetahui apa yang terjadi pada orang tersebut selama 14 hari.
- c Variabel X₃ yaitu Jumlah ODP. ODP adalah singkatan dari Orang Dalam Pemantauan. Orang dengan gejala demam ($\geq 38^{\circ}C$) atau dan dalam 14 hari sebelum gejala terakhir dan gejala penyakit pernapasan seperti pilek / sakit tenggorokan / batuk dan memiliki riwayat bepergian ke luar negeri atau tinggal di luar negeri.
- d Variabel X₄ yaitu jumlah PDP. PDP adalah singkatan dari Pasien Dalam Pengawasan. Orang dengan gejala demam ($\geq 38^{\circ}C$) atau dan dalam 14 hari sebelum gejala terakhir dan gejala penyakit pernapasan seperti pilek / sakit tenggorokan / batuk/sesak nafas ringan hingga berat serta memiliki riwayat bepergian ke luar negeri atau tinggal di luar negeri atau kontak dengan kasus konfirmasi atau *probabel* COVID-19.
- e Variabel X₅ yaitu jumlah kasus *confirmed* , merupakan status seseorang terinfeksi COVID-19 dengan hasil pemeriksaan laboratorium positif.

C. Teknik Analisa Data

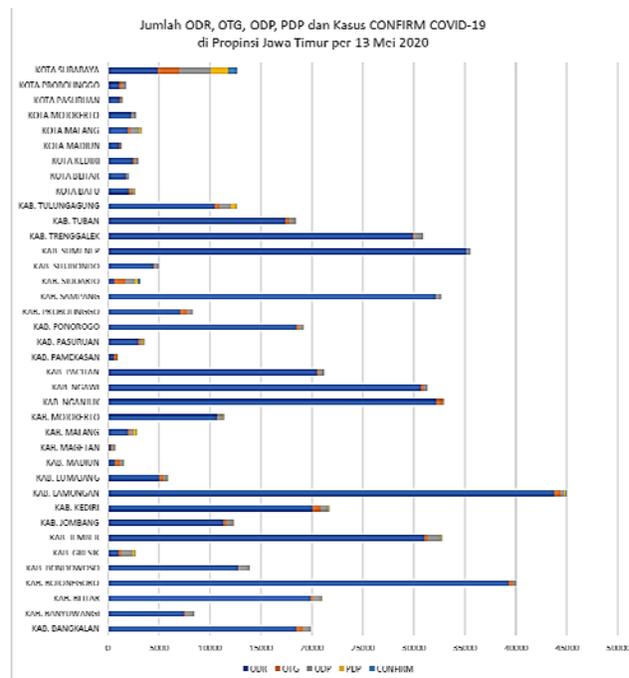
Analisis awal yang dilakukan adalah mendeskripsikan karakteristik masing-masing variabel berdasarkan kabupaten/kota di Jawa Timur dengan menggunakan nilai statistic deskriptif yaitu rata-rata, varian, nilai maksimum dan nilai minimum. Selanjutnya melakukan pengelompokan kabupaten/kota di Jawa Timur menggunakan algoritma *K-Means* sampai mendapatkan jumlah kelompok optimum berdasarkan nilai *Pseudo Fstatistics* pada persamaan (3). Serelah itu memilih hasil kelompok terbaik berdasarkan metode *K-Means* dengan menggunakan kriteria nilai *icdrate* pada persamaan (4) dan membandingkan nilai *icdrate* untuk memperoleh hasil pengelompokan terbaik. Setelah kelompok terbentuk kemudian dilakukan pengujian *One-way MANOVA* untuk membuktikan apakah masing-masing variabel penelitian yang digunakan untuk membentuk kelompok memiliki perbedaan di masing-masing klaster. Langkah terakhir adalah melakukan interpretasi hasil pengelompokan dan membuat kesimpulan.

IV. ANALISA DAN PEMBAHASAN

Pada bagian ini akan dilakukan Analisa dan pembahasan terhadap pengolahan data yang telah dilakukan dengan kerangka metode penelitian yang sudah dijelaskan pada bagian sebelumnya.

A. Eklporasi Data

Untuk mendapatkan petunjuk awal terkait dengan karakteristik sebaran COVID-19 di kota dan kabupaten yang ada di Propinsi Jawa Timur, gambar 1 merupakan histogram yang menampilkan pola sebaran setiap variabel yang digunakan sebagai dasar pengelompokan profil wilayah terjangkit COVID-19 di Jawa Timur.



Gbr. 1 Histogram jumlah data variabel COVID-19 di Jawa Timur.

a Variabel ODR (X₁)

Untuk kasus ODR angka tertinggi ditemukan sebanyak 43.772 orang dan terendah sebanyak 269 orang. Kasus ODR tertinggi terjadi pada Kabupaten Lamongan , secara visual juga dapat dilihat pada gambar 1. Sementara data terendah berada pada Kabupaten Magetan. Jika melihat data ODR Top 10 dari grafik, berturut-turut adalah Kabupaten Lamongan, Bojonegoro, Sumenep, Nganjuk, Sampang, Jember, Ngawi, Trenggalek, Pacitan dan Kediri. Menurut definisi, ODR adalah seseorang yang baru datang dari wilayah yang terjangkit corona virus yang merupakan *red zone* COVID-19. Sehingga faktor utama seseorang dikategorikan ODR adalah akibat mobilisasi atau bepergian dengan tujuan tertentu. Dalam publikasi (BPS, 2019) tentang data Statistik Mobilitas Penduduk dan Tenaga Kerja Tahun 2019 dijelaskan bahwa mobilitas tenaga kerja terbagi menjadi dua yaitu tipe

yaitu pekerja komuter dan pekerja sirkuler. Pekerja komuter adalah seseorang yang melakukan perjalanan rutin dengan tujuan bekerja, pergi dan pulang pada hari yang sama (setiap hari) antara tempat tinggal dan tempat Sedangkan definisi pekerja sirkuler menurut (BPS, 2019) adalah seseorang yang melakukan perjalanan rutin dengan tujuan bekerja, pergi dan pulang setiap minggu atau setiap bulan (< 6 bulan) antara tempat tinggal dan tempat bekerja yang berbeda kabupaten/kota. Pada laporan statistik mobilitas penduduk tercatat bahwa jumlah penduduk yang bekerja baik dengan sistem komuter dan sirkuler merupakan tertinggi kedua di Indonesia yaitu mencapai angka 20.099.220 hasil SAKERNAS 2017. Ini mengindikasikan bahwa terjadi *local transmission* yang cukup besar didaerah-daerah yang memiliki angka ODR tinggi.

Secara garis besar, ada tiga wilayah yang tergolong ODR tinggi yaitu untuk Kabupaten Lamongan, Bojonegoro (wilayah arah utara), Sumenep, Sampang, Jember (wilayah timur atau suku dominasi suku madura), Nganjuk, , Ngawi, Trenggalek, Pacitan dan Kediri (wilayah arah selatan). Sebagai sampel pengamatan, misalkan diambil wilayah *red zone* COVID 19 badalah Kota Surabaya dimana memiliki jumlah penduduk migran terbesar di Jawa Timur, dari bekerja yang berbeda kabupaten/kota. Dari data catatan kependudukan dan sipil Kota Surabaya (Humas, 2019) , tercatat ada 1232 per Oktober 2019 penduduk musiman yang ada di Surabaya dan data sebenarnya mungkin lebih dari itu karena tidak semua warga sementara mendaftarkan diri ke Dinas Kependudukan Kota Surabaya.

Jika diamati dari profesi pendatang tersebut, sebagian besar adalah pedagang baik pedagang yang menetap di suatu tempat secara permanen atau pedagang keliling adalah berasal dari wilayah Lamongan, Bojonegoro, Sampang, Sumenep dan Jember dengan usaha khas mereka yaitu kuliner dengan berbagai macam label seperti Soto Lamongan, Soto Madura, Sate Madura , Penyetan Lomongan, dan sebagainya. Sedangkan dari wilayah arah selatan seperti Nganjuk, , Ngawi, Trenggalek, Pacitan dan Kediri , umumnya pendatang berprofesi sebagai tukang bangunan atau pekerja kasar . Kesimpulan terkait profesi pendatang dari wilayah dominan ODR ini memang belum ditemukan secara ilmiah dari sebuah penelitian, akan tetapi masih berupa pengamatan subjektif penulis dengan melihat kebiasaan dan hubungan dengan pendatang dari wilayah-wilayah tersebut.

b Variabel OTG (X₂)

Statistik deskriptif dari data OTG menunjukkan nilai tertinggi 2086 kasus yang merupakan jumlah OTG di Kota Surabaya. Sedangkan data OTG terkecil adalah 17 yang merupakan data Kota Mojokerto. OTG merupakan seseorang yang tidak bergejala namun pernah memiliki kontak erat dengan kasus konfirmasi positif COVID-19. Sehingga secara intuitif variabel OTG akan berkorelasi positif dengan variabel CONFIRM. Dugaan ini dapat diuji dengan membuat matrik korelasi antar variabel menggunakan bantuan aplikasi SPSS. Dari hasil perhitungan korelasi antar variabel seperti pada table III dapat disimpulkan bahwa variabel OTG berkorelasi

positif dan signifikan dengan variabel ODP, PDP dan CONFIRM masing-masing sebesar 69,2 %, 81,7% dan 85,4%. Sehingga data OTG, ODP, PDP dan CONFIRM dapat digunakan untuk melihat wilayah *red zone* COVID-19. Dari grafik tampak bahwa Top 5 OTG secara berturut-turut adalah Kota Surabaya, Kabupaten Sidoarjo, Kediri, Probolinggo dan Nganjuk.

TABEL IIIII
KORELASI VARIABEL PENELITIAN

		Correlations				
		ODR	OTG	ODP	PDP	CONFIRM
ODR	Pearson Correlation	1	-.003	-.017	-.144	-.141
	Sig. (2-tailed)		.984	.921	.390	.399
	N	38	38	38	38	38
OTG	Pearson Correlation	-.003	1	.692**	.817**	.854**
	Sig. (2-tailed)	.984		.000	.000	.000
	N	38	38	38	38	38
ODP	Pearson Correlation	-.017	.692**	1	.851**	.807**
	Sig. (2-tailed)	.921	.000		.000	.000
	N	38	38	38	38	38
PDP	Pearson Correlation	-.144	.817**	.851**	1	.953**
	Sig. (2-tailed)	.390	.000	.000		.000
	N	38	38	38	38	38
CONFIRM	Pearson Correlation	-.141	.854**	.807**	.953**	1
	Sig. (2-tailed)	.399	.000	.000	.000	
	N	38	38	38	38	38

** Correlation is significant at the 0.01 level (2-tailed).

c Variabel ODP (X₃)

Data ODP menunjukkan jumlah terbanyak adalah sebesar 3090 terjadi di Kota Surabaya dan paling sedikit berjumlah 18 terjadi di Kabupaten Pamekasan, rata-rata jumlah ODP sebesar 572,05 dengan nilai standar deviasi 526,397. Nilai variasi ini dapat dikatakan cukup besar dalam data. Dari gambar 4.2 terlihat bahwa Top 5 untuk jumlah ODP terbanyak terjadi berturut-turut dari yang paling banyak ditemui adalah Kota Surabaya, Kabupaten Jember, Gresik, Tulungagung dan Bondowoso.

d Variabel PDP (X₄)

Untuk data PDP yaitu jumlah pasien dalam pengawasan, kasus terbanyak ditemui di Kota Surabaya sebesar 1711 kasus dan kasus paling sedikit terjadi di Kota Sampah dengan jumlah nihil (0 kasus). Rata-rata kasus yang terjadi sebesar 115,05 dengan standard deviasi 283,535. Data Top 5 dari PDP berturut-turut dari yang terbanyak adalah Kota Surabaya, Kabupaten Tulungagung, Sidoarjo, Kabupaten dan Kota Malang.

e Variabel Confirm (X₅)

Kasus positif (confirm) paling banyak terjadi di Kota Surabaya sebesar 870 kasus, dan kasus yang paling sedikit ditemui terjadi di Kabupaten Sampangan hanya 1 kasus. Nilai rata-rata kasus positif COVID-19 sebesar 46,47 dengan standard deviasi sebesar 142,136. Sama halnya dengan

variabel lainnya, nilai varians dari data cukup besar. Untuk kasus positif COVID-19, ranking 5 terbanyak dari atas adalah Kota Surabaya, Kabupaten Sidoarjo, Lamongan, Magetan dan Malang.

B. Hasil Klustering

Prosedur pengelompokan non-hirarki dengan *K-Means* dilakukan dengan dimulai jumlah klaster 2,3,4, 5 dan 6. Dari hasil percobaan tersebut dapat dilihat sebagai berikut :

TABEL IVV
REKAPITULASI ANGGOTA KLASTER

Jumlah Klaster (K)	Klaster 1 (N ₁)	Klaster 2 (N ₂)	Klaster 3 (N ₃)	Klaster 4 (N ₄)	Klaster 5 (N ₅)	Klaster 6 (N ₆)
2	8	30	-	-	-	-
3	10	20	8	-	-	-
4	6	20	2	10	-	-
5	6	6	6	2	18	-
6	6	5	6	4	15	2

a. Pemilihan Klaster Terbaik

Untuk mengetahui jumlah klaster optimum maka digunakan *Pseudo F-Statistics* Semakin tinggi nilai *Pseudo F-Statistics* maka jumlah klaster semakin optimum. Berikut akan dibandingkan nilai *Pseudo F-Statistics* untuk hasil klustering dengan k=2,3 dan 4.

TABEL V
PSEUDO F-STATISTICS

Jumlah Klaster (K)	<i>Pseudo F-Statistics</i>	<i>icdrate</i>
2	102,937	0,9217
3	239,156	0,9087
4	236,405	0,7932
5	448,327	0,7854
6	5,924	0,7766

Dari tabel V dapat disimpulkan bahwa nilai *Pseudo F-Statistics* tertinggi terjadi ketika dilakukan klaster sebanyak 5 kelompok, sehingga ditetapkan bahwa jumlah klaster optimum adalah k=5. Selanjutnya jika dilihat ukuran internal dari pengelompokan dengan menghitung *icdrate* (*internal cluster dispersion rate*) dimana semakin kecil nilainya menunjukkan bahwa perbedaan dalam kelompok tersebut semakin kecil atau dengan kata lain hasil pengelompokan semakin homogen. Pada tabel V juga dapat dilihat bahwa nilai dispersi dalam kelompok mulai konvergen disekitar 0,78. Dengan demikian dapat disimpulkan titik optimum jumlah klaster terbaik adalah sebanyak 5 kelompok.

b. Pengujian One Way MANOVA

Untuk membuktikan bahwa ke 5 kelompok yang dihasilkan berbeda secara signifikan, dapat dilakukan pengujian MANOVA (*Multivariate Analysis of Variance*) untuk mengetahui ada atau tidaknya perbedaan pada setiap klaster yang terbentuk. Pengujian ini untuk menguji hipotesis apakah ada perbedaan pada kelompok yang terbentuk. Dimana satu faktor tersebut terdiri dari 5 kategori, yaitu kelompok 1, kelompok 2, kelompok 3, kelompok 4 dan

kelompok 5. Variabel respon pada pengujian *One-way MANOVA* adalah variabel ODR, OTG, ODP, PDP dan CONFIRM. Dengan menggunakan hipotesis :

$$H_0 : \tau_1 = \tau_2 = \dots = \tau_g$$

$$H_1 : \text{Minimal Ada Satu } \tau_g \neq 0, l = 1, 2, \dots, g$$

Dilihat dari nilai *p-value* dari *one way-Manova*, didapatkan nilai *p-value* < 5% pada statistik Wilks Lambda sehingga dapat dikatakan Tolak H₀, artinya bahwa secara multivariat kelompok memberikan pengaruh terhadap variabel ODR, OTG, ODP, PDP dan CONFIRM dengan tingkat kepercayaan 95%.

C. Interpretasi Hasil Klustering

Berdasarkan kajian literatur yang sudah dilakukan sebelumnya, menyebutkan bahwa terdapat dua faktor yang mempengaruhi penyebaran sebuah wabah yaitu faktor orang (usia, kebiasaan merokok, penyakit penyerta, jenis kelamin) dan faktor tempat (kepadatan penduduk, iklim, kelembapan, kesenjangan sosial, layanan kesehatan, ketersediaan tenaga kesehatan dan rumah bersih). Untuk memperdalam analisis karakteristik hasil klustering, digunakan informasi tambahan dari data hasil Susenas BPS 2019.

TABEL VI
NILAI RATA-RATA VARIABEL SETIAP KLASTER
(FINAL CENTROID)

KLASTER	ODR (X ₁)	OTG (X ₂)	ODP (X ₃)	PDP (X ₄)	CONFIRM (X ₅)
1	19140,83	376,00	643,50	44,83	15,67
2	9968,67	287,00	756,83	112,17	18,00
3	31836,00	234,00	531,50	35,50	8,50
4	41550,50	475,00	308,50	77,00	39,00
5	2038,06	346,78	529,44	170,17	79,72

Untuk memudahkan interpretasi maka urutan nilai setiap variabel diberikan label dengan 5 tingkatan bertipe kategorik mulai dari nilai terendah sampai tertinggi yaitu Rendah (R), Sedang (S), Cukup Tinggi (CT), Tinggi (T), Sangat Tinggi (ST). Dengan demikian jika tabel VI dikonversi kedalam label kategori akan menjadi seperti tabel VII.

TABEL VII
TINGKATAN VARIABEL SETIAP KLASTER
(Final Centroid)

KLASTER	ODR (X ₁)	OTG (X ₂)	ODP (X ₃)	PDP (X ₄)	CONFIRM (X ₅)	Zone
1	CT	T	T	S	S	Green
2	S	S	ST	T	CT	Yellow
3	T	R	CT	R	R	Light Green
4	ST	ST	R	CT	T	Orange
5	R	CT	S	ST	ST	Red

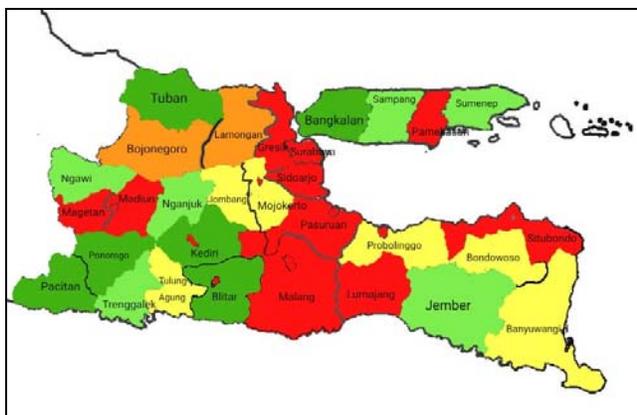
Wilayah kabupaten/kota yang masuk dalam klaster 1 adalah Kab. Pacitan, Kab. Ponorogo, Kab. Blitar, Kab. Kediri, Kab. Tuban Dan Kab. Bangkalan merupakan klaster dengan kasus PDP dan CONFIRM tingkatan Sedang dengan kasus ODR nya cukup tinggi sedangkan OTG dan ODP nya

kategori tinggi. Karakteristik dari klaster ini adalah data yang cukup tinggi untuk jumlah ODR, OTG dan ODP mengindikasikan bahwa ada mobilitas yang cukup besar dalam aktivitas masyarakat di kota-kota tersebut.

Selanjutnya wilayah kabupaten/kota yang masuk dalam klaster 2 adalah Kab. Tulungagung, Kab. Banyuwangi, Kab. Bondowoso, Kab. Probolinggo, Kab. Mojokerto, Kab. Jombang merupakan klaster dengan kasus PDP dan CONFIRM tingkatan Tinggi dan Cukup Tinggi dan dengan kasus ODR, OTG termasuk kategori Sedang dan kasus ODP nya sangat tinggi. Karakteristik dominan yang dapat dilihat dari dilihat dari kota-kota yang masuk dalam klaster ini memiliki iklim dengan suhu rata-rata 26 derajat Celsius dengan kelembapan udara diatas 84%.

Sedangkan wilayah kabupaten/kota yang masuk dalam klaster 3 adalah Kab. Trenggalek, Kab. Jember, Kab. Nganjuk, Kab. Ngawi, Kab. Sampang dan Kab. Sumenep, merupakan klaster dengan kasus PDP dan CONFIRM serta OTG rendah, meskipun data ODR dan ODP nya termasuk kategori tinggi dan cukup tinggi. Jika dikonfirmasi dengan data tambahan dari BPS, rata-rata kab/kota di klaster ini memiliki jumlah layanan kesehatan untuk RS Swasta atau RSUD sekitar 4, puskesmas terdapat 27 dan poliklinik sekitar 24 layanan.

Kemudian wilayah kabupaten/kota yang masuk dalam klaster 4 adalah Kab. Bojonegoro dan Kab. Lamongan yang merupakan wilayah dengan jumlah kasus PDP dan CONFIRM kategori Cukup Tinggi dan Tinggi, meskipun data ODP Rendah akan tetapi data ODR dan OTG sangat tinggi. Situasi yang terjadi pada klaster ini hampir mirip dengan kondisi pada klaster 2 (*yellow zone*), jika pada wilayah klaster 2 diduga penyebaran virus COVID-19 dipengaruhi suhu dan kelembapan udara di wilayah tersebut, akan tetapi di klaster ini diduga akibat mobilitas penduduk yang masih tinggi atau kedisiplinan masyarakat terkait social dan physical distancing yang mungkin belum dipatuhi secara menyeluruh, hal ini dapat dilihat dari angka ODR dan OTG yang sangat tinggi.



Gbr. 2 Pemetaan Hasil Klastering Penyebaran COVID-19

Terakhir, wilayah kabupaten/kota yang masuk dalam klaster 5 adalah Kab. Malang, Kab. Lumajang, Kab. Situbondo, Kab. Pasuruan, Kab. Sidoarjo, Kab. Madiun, Kab. Magetan, Kab. Gresik, Kab. Pamekasan, Kota Kediri, Kota Blitar, Kota Malang, Kota Probolinggo, Kota Pasuruan, Kota Mojokerto, Kota Madiun dan Kota Surabaya serta Kota Batu, merupakan wilayah dengan kasus PDP dan CONFIRM sangat tinggi. Untuk memahami karakteristik dari kabupaten atau kota di klaster ini perlu dilihat data hasil Susenas (BPS J., 2020). Dalam data tersebut, ciri dominan dari klaster ini adalah angka kepadatan penduduk yang paling tinggi diantara klaster yang lain. Dengan jumlah penduduk yang padat dan dibandingkan jumlah fasilitas kesehatan yang ada, didapatkan sekitar 51% masyarakat menyatakan bahwa waktu tunggu untuk mendapatkan pelayanan kesehatan lama.

V. PENUTUP

Kesimpulan dari penelitian berdasarkan data jumlah kasus COVID-19 menurut variable ODR, OTG, ODP, PDP dan CONFIRM didapatkan 5 klaster dengan karakteristik singkat sebagai berikut :

- Zona Hijau Muda (*light green zone*) yang memiliki ciri khas memiliki kasus PDP dan CONFIRM rendah dan terjadi di Kota/Kab dengan jumlah faskes terbatas, persentase jumlah penduduk miskin dan perokok aktif diatas usia 15 tahun yang tinggi.
- Zona Hijau (*green zone*) yang memiliki ciri khas memiliki kasus PDP dan CONFIRM sedang dan terjadi di Kota/Kab dengan jumlah angka penderita DBD paling tinggi dan rata-rata tenaga medis yang dimiliki paling sedikit dibandingkan dengan klaster lain.
- Zona Kuning (*yellow zone*) yang memiliki ciri khas memiliki kasus PDP tinggi dan kasus CONFIRM cukup tinggi terjadi di Kota/Kab yang memiliki iklim dengan suhu rata-rata 26 derajat Celsius dengan kelembapan udara diatas 84% dan lebih dari 40% masyarakat menyatakan waktu tunggu mendapatkan layanan kesehatan lama.
- Zona Oranye (*orange green zone*) yang memiliki ciri khas memiliki kasus PDP cukup tinggi dan kasus confirm tinggi dan terjadi di Kota/Kab dengan jumlah kasus ODR dan OTG sangat tinggi yang mengindikasikan mobilitas yang tinggi yang didukung dengan rata-rata jumlah pekerja merupakan angka tertinggi dari klaster lain yaitu mencapai 650.000 orang (70% dari Angkatan kerja).
- Zona Merah (*red zone*) yang memiliki ciri khas memiliki kasus PDP dan CONFIRM sangat tinggi dan terjadi di kota/kab dengan padat penduduk dan lebih dari 50% penduduknya menyatakan waktu tunggu mendapatkan layanan Kesehatan lama.

REFERENSI

- [1] Alizkan, U. (2017). Analisis Korelasi Kelembaban Udara Terhadap Epidemi Demam Berdarah Yang Terjadi Di Kabupaten Dan Kota Serang. *GRAVITY :Jurnal Ilmiah Penelitian dan Pembelajaran Fisika*, 3(1), 23-29.
- [2] Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. *Communications in Statistics - Theory and Methods*, 3(1), 1-27.
- [3] Harlan, J. (2008). *Epidemiologi Kebidanan*. Jakarta: Universitas Gunadarma Press.
- [4] Johnson, R., & Wichern, D. (2007). Clustering, Distance Methods, And Ordination. In R. Johnson, & D. Wichern, *Applied Multivariate Statistical Analysis* (pp. 696-702). Upper Saddle River: Pearson Prentice Hall.
- [5] Kementerian Kesehatan, R. (2020). *Frequently Asked Questions (FAQ) COVID-19*. Jakarta.
- [6] Lima, J., & Mingoti, S. (2006). Comparing SOM neural network with Fuzzy c-means, K-means and traditional hierarchical clustering algorithms. *European Journal of Operational Research*, 174(3), 1742-1759.
- [7] Matjijk , A., & Sumertajaya, I. (2011). *Sidik Peubah Ganda dengan menggunakan SAS*. Bogor: IPB Press.
- [8] Nisa', F. F. (2016). *Pemodelan Faktor-Faktor Yang Mempengaruhi Jumlah Kasus Tuberkulosis Di Jawa Timur Menggunakan Regresi Nonparametrik Spline*. Surabaya.
- [9] Priyadarsini, S., & Suresh, M. (2020). Factors influencing the epidemiological characteristics of pandemic COVID 19:A TISM approach. *International Journal Of Healthcare Management*, 1-10.
- [10] Tan, S. C. (2020). *Clinical and epidemiological characteristics of Coronavirus Disease 2019 (COVID-19) Patient*. USA: MedRxiv : The Preprint Server for Health Science.