

Analisis *Cluster* dan Korespondensi terhadap Indikator Pertumbuhan Penduduk Kota Surabaya Tahun 2020

Noviana Maulidia dan Sri Pingit Wulandari
Departemen Statistika Bisnis, Institut Teknologi Sepuluh Nopember (ITS)
e-mail: sripingitwulandari@gmail.com

Abstrak—Surabaya sebagai Ibu Kota Jawa Timur selain pertumbuhan penduduknya sangat cepat juga memiliki kepadatan penduduk yang tertinggi yaitu 8.707 jiwa/km². Berdasarkan data Kota Surabaya Dalam Angka tahun 2020 menunjukkan bahwa setiap Kecamatan memiliki laju pertumbuhan penduduk yang beragam. Permasalahan dari penelitian ini adalah besarnya keragaman dari indikator pertumbuhan penduduk, yang diukur menggunakan variabel jumlah kelahiran, jumlah kematian, jumlah pindah keluar, dan jumlah pindah datang pada 31 Kecamatan di Surabaya. Keragaman antar Kecamatan dikelompokkan dengan menggunakan analisis *cluster*. Analisis korespondensi untuk mengetahui wilayah Kecamatan yang cenderung secara visualisasi mempunyai jarak terdekat dengan indikator pertumbuhan penduduk. Data penelitian menggunakan data yang diperoleh dari Dinas Kependudukan dan Pencatatan Sipil Surabaya. Hasil analisis dengan Metode *Ward's*, menggunakan *Pseudo F* didapatkan paling optimum 3 *cluster*, yaitu *cluster 1* mempunyai pertumbuhan penduduk terklasifikasi sedang dengan jumlah anggota 13 Kecamatan, *cluster 2* mempunyai pertumbuhan penduduk terklasifikasi cepat dengan jumlah anggota 5 Kecamatan, dan *cluster 3* mempunyai pertumbuhan penduduk terklasifikasi lambat dengan jumlah anggota 13 Kecamatan. Hasil analisis korespondensi *cluster 1* dan *cluster 3* cenderung memiliki kedekatan jarak dengan indikator jumlah kelahiran, *cluster 2* cenderung memiliki kedekatan jarak dengan indikator jumlah pindah keluar.

Kata Kunci—Analisis *Cluster*, Analisis Korespondensi, Metode *Ward's*, dan Pertumbuhan Penduduk.

I. PENDAHULUAN

SURABAYA adalah Ibu Kota dari Provinsi Jawa Timur dengan jumlah penduduk terbanyak nomor dua setelah Jakarta. Jumlah penduduk Surabaya mengalami pertumbuhan sangat pesat dari tahun ke tahun. Indikator pertumbuhan penduduk, yaitu jumlah kelahiran, jumlah kematian, jumlah pindah keluar, dan jumlah pindah datang sangat menentukan jumlah penduduk di Kota Surabaya. Kota Surabaya terdiri dari 31 Kecamatan dan 154 Kelurahan. Surabaya memiliki jumlah penduduk yang tercatat sebesar 2.904.751 jiwa. Luas Kota Surabaya 326,81 km² menjadikan Kota Surabaya padat penduduk sebesar 8.707 jiwa/km² [1].

Pertumbuhan penduduk merupakan suatu peristiwa bertambah atau berkurangnya jumlah penduduk dari waktu ke waktu dalam suatu wilayah tertentu yang dapat disebut dengan dinamika penduduk. Gejala dinamika penduduk dipengaruhi oleh tiga indikator utama, yaitu kelahiran, kematian, dan perpindahan penduduk [2].

Setiap Kecamatan memiliki laju pertumbuhan penduduk yang beragam, terdapat wilayah Kecamatan yang memiliki laju pertumbuhan penduduk tertinggi yaitu Kecamatan Kenjeran sebesar 4,08% tahun 2020 dan laju pertumbuhan

penduduk terendah pada wilayah Kecamatan Dukuh Pakis sebesar 0,43% tahun 2020 [3]. Setiap Kecamatan memiliki jumlah kelahiran, jumlah kematian, jumlah pindah keluar, dan jumlah pindah datang yang sangat beragam. Analisis *cluster* dan analisis korespondensi dapat membantu untuk mengelompokkan Kecamatan sehingga varian didalam kelompok minimal dan varian antar kelompok maksimal.

Dinas kependudukan dan Pencatatan Sipil Kota Surabaya menjelaskan bahwa rata-rata pertumbuhan penduduk Surabaya naik sebanyak 3% atau mencapai 80 ribu jiwa pertahun dan penduduk pindah keluar ataupun pindah datang lebih banyak dibandingkan dengan kelahiran dan kematian. Hal tersebut menjadi dasar dirumuskannya permasalahan untuk mengetahui Kecamatan apa saja yang memiliki kecenderungan indikator pertumbuhan Kota Surabaya tahun 2020.

II. TINJAUAN PUSTAKA

A. Analisis Cluster

Analisis *cluster* merupakan salah satu teknik analisa yang bertujuan untuk menggabungkan pengamatan kedalam kelompok-kelompok yang homogen dalam kelompok yang sama dan heterogen antar kelompok masing-masing [4].

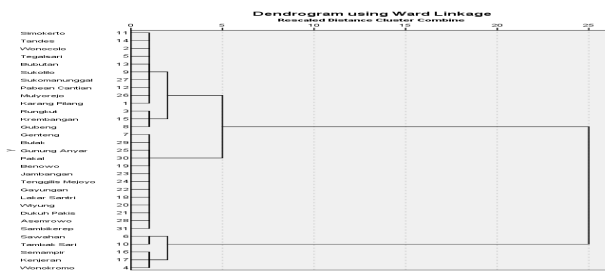
B. Metode Hierarki

Metode *hierarki* merupakan metode pengelompokan yang hasilnya disajikan secara bertingkat atau berjenjang dari n , $n-1$, ..., 1 kelompok. Jarak *Euclidean* merupakan fungsi jarak yang sering didefinisikan sebagai jarak antara observasi ke- i dan ke- j dan k merupakan kelompok ke- k [5]. Rumus jarak *Euclidean* dari objek ke- i menuju objek ke- j , yang dirumuskan pada persamaan (1).

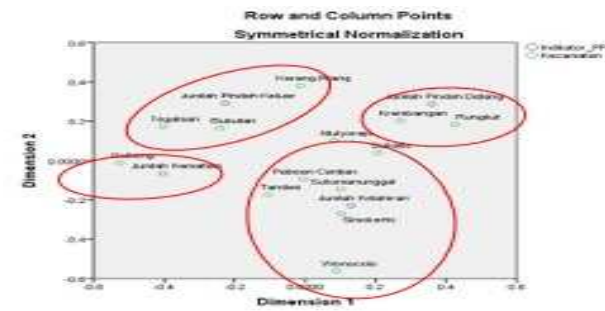
$$d(i,j) = \sqrt{\sum_{k=1}^p (X_{ik} - X_{jk})^2}; \quad i \neq j \quad (1)$$

C. Ward's Method

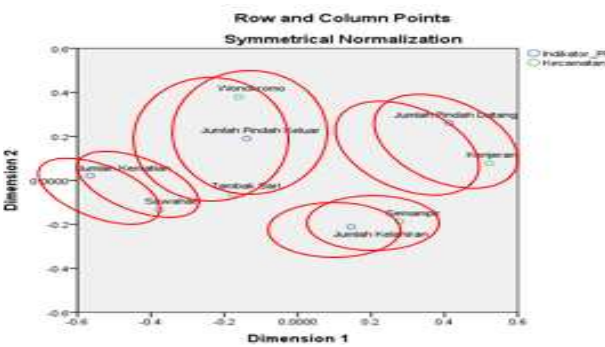
Metode *Ward's* merupakan metode *clustering hierarki clustering hierarki* yang bersifat *agglomerative* untuk memperoleh *cluster* yang memiliki varian internal sekecil mungkin. Metode *Ward's* dikenal sebagai metode terbaik dalam analisis *cluster hierarki*. Berdasarkan identifikasi hasil dendrogram, metode *Ward's* merupakan metode yang memberikan hasil yang lebih mudah dalam mengelompokkan pada metode *hierarki* lainnya yang dapat meminimumkan *Error Sum of Squared* (ESS). *Agglomerative* merupakan prosedur pengelompokkan *hierarki* di mana setiap objek berawal dari *cluster* yang terpisah. *Cluster-cluster* dibentuk dengan mengelompokkan objek ke dalam *cluster* yang



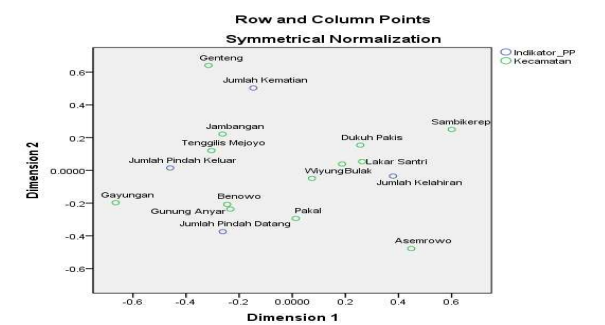
Gambar 1. Dendrogram indikator pertumbuhan penduduk.



Gambar 2. Plot korespondensi kecenderungan cluster sedang.



Gambar 3. Plot kecenderungan cluster cepat.



Gambar 4. Plot korespondensi kecenderungan cluster lambat.

semakin banyak objek menjadi anggotanya. Proses ini dilanjutkan sampai semua objek menjadi anggota dari cluster tunggal [6]. Jika ada k kelompok maka ESS merupakan penjumlahan dari ESS_k dan dirumuskan pada persamaan (2).

$$ESS = ESS_1 + ESS_2 + ESS_3 + \dots + ESS_k \quad (2)$$

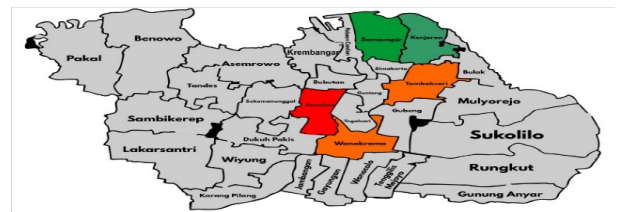
Apabila semua kelompok bergabung menjadi satu kelompok dari n objek, maka untuk menghitung jarak antara dua kelompok dapat dirumuskan pada persamaan (3) [4].

$$ESS = \sum_{j=1}^n (X_j - \bar{X})' (X_j - \bar{X}) \quad (3)$$

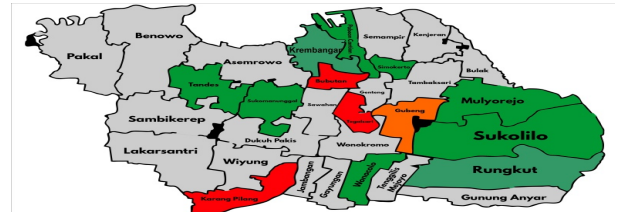
Persamaan (3) X_j merupakan ukuran asosiasi multivariate dengan j item dan \bar{X} merupakan rata-rata dari seluruh item.

D. Calinski – Harabasz Pseudo F-statistic

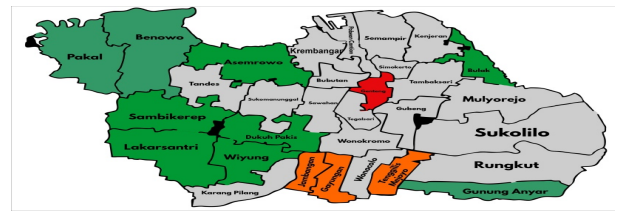
Pseudo F-statistic merupakan metode yang digunakan



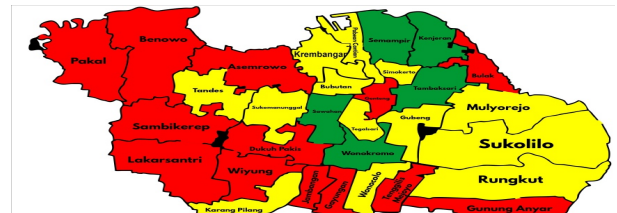
Gambar 5. Kecenderungan cluster cepat.



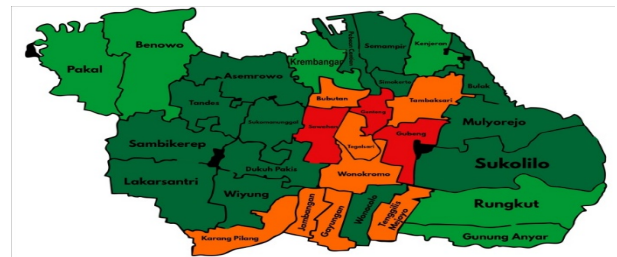
Gambar 6. Kecenderungan cluster sedang.



Gambar 7. Kecenderungan cluster lambat.



Gambar 8. Pengelompokkan dengan analisis cluster.



Gambar 9. Pemetaan hasil analisis korespondensi.

untuk menentukan banyaknya kelompok yang optimum. *Pseudo F* tertinggi menunjukkan bahwa kelompok yang memiliki hasil optimal dengan keragaman dalam kelompok sangat homogeny sedangkan antar kelompok sangat heterogen. Metode *Pseudo F-statistic* dapat dirumuskan dalam persamaan (4) [7].

$$Pseudo - F = \frac{\left(\frac{R^2}{k-1}\right)}{\left(\frac{1-R^2}{n-k}\right)} \quad (4)$$

Dimana,

$$R^2 = \frac{(SST-SSW)}{SST} \quad (5)$$

$$SST = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (X_{ijk} - \bar{X}_j)^2 \quad (6)$$

$$SSW = \sum_{i=1}^n \sum_{j=1}^c \sum_{k=1}^p (X_{ijk} - \bar{X}_{jk})^2 \quad (7)$$

Persamaan (6) SST merupakan total jumlah dari kuadrat

Tabel 1.
Struktur Data Tabel Kontingensi

| Variabel X | Variabel Y | | | | | Total Baris |
|-------------|------------|----------|----------|-----|----------|-------------|
| | 1 | 2 | 3 | ... | J | |
| 1 | X_{11} | X_{12} | X_{13} | ... | X_{1j} | $X_{1.}$ |
| 2 | X_{21} | X_{22} | X_{23} | ... | X_{2j} | $X_{2.}$ |
| 3 | X_{31} | X_{32} | X_{33} | ... | X_{3j} | $X_{3.}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| I | X_{i1} | X_{i2} | X_{i3} | ... | X_{ij} | $X_{i.}$ |
| Total Kolom | $X_{.1}$ | $X_{.2}$ | $X_{.3}$ | ... | $X_{.j}$ | $X_{..}$ |

Tabel 2.
Variabel Penelitian

| Variabel | Ket | Definisi Operasional | Satuan | Skala |
|----------|----------------------|--|--------|-------|
| X_1 | Jumlah Kelahiran | Banyaknya bayi yang dilahirkan oleh seorang wanita | Jiwa | Rasio |
| X_2 | Jumlah Kematian | Berakhirnya fungsi biologis dikarenakan terlepasnya ruh dari jasad manusia | Jiwa | Rasio |
| X_3 | Jumlah Pindah Keluar | Perpindahan penduduk keluar dari suatu daerah asal | Jiwa | Rasio |
| X_4 | Jumlah Pindah Datang | Masuknya penduduk ke suatu daerah tempat tujuan | Jiwa | Rasio |

Tabel 3.
Struktur Data

| | A ₁ | A ₂ | A ₃ | A ₄ | Total |
|-----------------|----------------|----------------|----------------|----------------|-----------|
| B ₁ | X_{11} | X_{12} | X_{13} | X_{14} | $X_{1.}$ |
| B ₂ | X_{21} | X_{22} | X_{23} | X_{24} | $X_{2.}$ |
| B ₃ | X_{31} | X_{32} | X_{33} | X_{34} | $X_{3.}$ |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | ⋮ |
| B ₃₁ | X_{311} | X_{312} | X_{313} | X_{314} | $X_{31.}$ |
| Total | $X_{.1}$ | $X_{.2}$ | $X_{.3}$ | $X_{.4}$ | $X_{..}$ |

jarak sampel terhadap rata-rata keseluruhan, n merupakan banyaknya sampel. c merupakan banyaknya variabel, p merupakan banyaknya kelompok, X_{ijk} merupakan sampel ke- I pada variabel ke- j kelompok ke- k , dan \bar{X}_j merupakan rata-rata seluruh sampel pada variabel ke- j . persamaan (7) SSW merupakan total jumlah dari kuadrat jarak sampel terhadap rata-rata kelompok, \bar{X}_{jk} merupakan rata-rata sampel pada variabel ke- j dan kelompok ke- k .

E. Tabel Kontingensi

Tabel kontingensi merupakan tabel yang berisi data dengan ukuran baris I dan kolom J yang diperoleh dalam sampel. Tabel kontingensi menghubungkan dua variabel, yaitu variabel pada baris I dan variabel pada kolom J . frekuensi-frekuensi sel merupakan frekuensi-frekuensi yang menempati beberapa sel dalam tabel kontingensi. Sedangkan frekuensi merjinal merupakan frekuensi total dalam setiap baris atau kolom [8]. I merupakan banyaknya kategori pada variabel X, J merupakan banyaknya kategori pada variabel Y, $X_{.}$ merupakan jumlah observasi keseluruhan, $X_{i.}$ merupakan jumlah observasi pada variabel X dengan kategori I , dan $X_{.j}$ merupakan jumlah observasi pada variabel Y dengan kategori J .

F. Uji Independensi

Uji independensi digunakan untuk mengetahui apakah dua variabel memiliki hubungan yang signifikan [9]. Uji

Tabel 4.
Nilai Pseudo F-statistic Setiap Kelompok

| Jumlah | SST | SSW | R ² | Pseudo F |
|--------|-------------|----------|----------------|----------|
| 2 | 40375946,45 | 10705625 | 0,73485 | 80,3726 |
| 3 | 40375946,45 | 5185078 | 0,87158 | 95,0173 |
| 4 | 40375946,45 | 3764687 | 0,90676 | 87,5242 |

Tabel 5.
Anggota Cluster

| Kelompok | Anggota Cluster | Total |
|----------|--|-------|
| 1 | Simokerto, Tandes, Wonocolo, Tegalsari, Bubutan, Sukolilo, Sukomanunggal, Pabean Cantian, Mulyorejo, Karang Pilang, Rungkut, Krembangan, Gubeng. | 13 |
| 2 | Sawahan, Tambak Sari, Semampir, Kenjeran, Wonokromo. | 5 |
| 3 | Genteng, Bulak, Gunung Anyar, Pakal, Benowo, Jambangan, Tenggilis Mejoyo, Gayungan, Lakar Santri, Wiyung, Dukuh Pakis, Asemrowo, Sambikerep. | 13 |

Tabel 6.
Deskripsi Cluster Baru yang Terbentuk (Jiwa)

| Indikator Penduduk | Pertumbuhan | Cluster 1 | Cluster 2 | Cluster 3 |
|----------------------------------|-------------|-----------|-----------|-----------|
| Jumlah Kelahiran | | 1830 | 3565 | 1140 |
| Jumlah Kematian | | 868 | 1691 | 464 |
| Jumlah Pindah Keluar | | 826 | 1586 | 488 |
| Jumlah Pindah Datang | | 830 | 1595 | 536 |
| Pertumbuhan Penduduk (%) | | 1,59 | 9,824 | 0,547 |
| Klasifikasi Pertumbuhan Penduduk | | Sedang | Cepat | Lambat |
| Total Kecamatan | | 13 | 5 | 13 |

Tabel 7.
Uji Independensi

| Cluster | χ^2_{hitung} | χ^2_{tabel} |
|------------|-------------------|------------------|
| 2 (Sedang) | 500,730 | 50,998 |
| 3 (Cepat) | 579,747 | 21,026 |
| 4 (Lambat) | 726,601 | 50,998 |

independensi yang di gunakan yaitu uji Chi-Square yang merupakan salah satu alat uji statistic. Berikut adalah syarat-syarat dalam variabel yang harus terpenuhi sebagai berikut.

1) Homogen

Homogeny merupakan suatu sel di mana setiap sel harus menggambarkan objek yang sama. Sehingga jika data heterogen maka tidak bisa dianalisis menggunakan tabel kontingensi.

2) Mutually Exclusive dan Mutually Exhaustive

Mutually exclusive merupakan suatu hubungan antar level yang saling asing. Sedangkan mutually exhaustive merupakan suatu proses perubahan secara lengkap yang menjadi lebih sederhana, sehingga jika klasifikasi satu unsur, maka yang dapat diklasifikasikan hanya satu unit saja, dengan semua nilai harus terpenuhi dalam klasifikasi yang dilakukan.

3) Skala Nominal dan Skala Ordinal

Skala nominal merupakan skala yang membedakan kategori berdasarkan jenisnya. Sedangkan skala ordinal merupakan skala yang membedakan kategori berdasarkan tingkatan.

Langkah-langkah uji Chi-Square dalam uji independensi dengan hipotesis,

H0: Tidak ada hubungan antara dua variabel

H1: Terdapat hubungan antara dua variabel

Tabel 8.
Raduksi Dimensi Cluster Sedang

| Dimensi | Inersia | Proporsi | Proporsi Kumulatif |
|---------|---------|----------|--------------------|
| 1 | 0,005 | 0,616 | 0,616 |
| 2 | 0,003 | 0,342 | 0,957 |
| 3 | 0,000 | 0,043 | 1,000 |

Tabel 9.
Reduksi Dimensi Cluster Cepat

| Dimensi | Inersia | Proporsi | Proporsi Kumulatif |
|---------|---------|----------|--------------------|
| 1 | 0,012 | 0,861 | 0,861 |
| 2 | 0,001 | 0,108 | 0,969 |
| 3 | 0,000 | 0,031 | 1,000 |

Tabel 10.
Reduksi Dimensi Cluster Lambat

| Dimensi | Inersia | Proporsi | Proporsi Kumulatif |
|---------|---------|----------|--------------------|
| 1 | 0,014 | 0,671 | 0,671 |
| 2 | 0,005 | 0,257 | 0,928 |
| 3 | 0,002 | 0,072 | 1,000 |

Tabel 11.
Jarak Euclidean Cluster Sedang

| Kecamatan | E ₁ | E ₂ | E ₃ | E ₄ |
|----------------|----------------|----------------|----------------|----------------|
| Karang Pilang | 0,62773 | 0,59266 | 0,23123 | 0,38539 |
| Wonocolo | 0,33551 | 0,69792 | 0,91059 | 0,89311 |
| Rungkut | 0,50672 | 0,86367 | 0,66105 | 0,12434 |
| Tegalsari | 0,66900 | 0,24001 | 0,21273 | 0,77261 |
| Gubeng | 0,69127 | 0,13565 | 0,42569 | 0,93541 |
| Sukolilo | 0,27884 | 0,61798 | 0,50236 | 0,29310 |
| Simokerto | 0,04965 | 0,54372 | 0,65122 | 0,61567 |
| Pabean Cantian | 0,19304 | 0,39592 | 0,44342 | 0,52973 |
| Bubutan | 0,54118 | 0,28157 | 0,12777 | 0,61366 |
| Tandes | 0,24232 | 0,31509 | 0,48145 | 0,65690 |
| Krembangan | 0,45443 | 0,72477 | 0,50589 | 0,12376 |
| Mulyorejo | 0,33446 | 0,51233 | 0,36274 | 0,33504 |
| Sukomanunggal | 0,09014 | 0,50802 | 0,545 | 0,50506 |

Statistic u_{ji} yang digunakan dapat berupa statistic uji Chi-Square terdapat pada persamaan (8).

$$\chi^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(x_{ij} - e_{ij})^2}{e_{ij}} \tag{8}$$

Nilai harapan (e_{ij}) pada persamaan (8) dirumuskan pada persamaan (9).

$$e_{ij} = \frac{x_i \cdot x_j}{x_{..}} \tag{9}$$

G. Analisis Korespondensi

Analisis korespondensi merupakan prosedur grafis yang digunakan untuk memperagakan baris dan kolom secara serentak dalam tabel kontingensi dua arah yang dapat diperluas dengan tabel kontingensi multi arah. Analisis korespondensi ini memproyeksikan baris dan kolom yang dalam analisis diperoleh tabel kontingensi dengan menggunakan jarak Chi-Square.

1) Matriks Data

Langkah pertama yang dilakukan dalam analisis korespondensi yaitu perhitungan matriks N dengan elemen X_{ij} yang tersusun dalam tabel kontingensi dua dimensi (I × J) (Tabel 1). Apabila X_{..} merupakan total frekuensi dari matriks N, langkah pertama yang dilakukan yaitu menyusun matriks proporsi P = (p_{ij}) terdapat pada persamaan (10).

Tabel 12.
Jarak Euclidean Cluster Cepat

| Kecamatan | E ₁ | E ₂ | E ₃ | E ₄ |
|-------------|----------------|----------------|----------------|----------------|
| Wonokromo | 0,66463 | 0,54129 | 0,19016 | 0,58404 |
| Sawahan | 0,52864 | 0,24443 | 0,40202 | 0,88235 |
| Tambak Sari | 0,35488 | 0,39818 | 0,25683 | 0,67271 |
| Semampir | 0,13336 | 0,86804 | 0,56074 | 0,46694 |
| Kenjeran | 0,47362 | 1,08944 | 0,67059 | 0,21437 |

Tabel 13.
Jarak Euclidean Cluster Lambat

| Kecamatan | E ₁ | E ₂ | E ₃ | E ₄ |
|--------------|----------------|----------------|----------------|----------------|
| Genteng | 0,96954 | 0,21741 | 0,64257 | 1,01738 |
| Lakar Santri | 0,14700 | 0,60803 | 0,72405 | 0,67798 |
| Benowo | 0,64850 | 0,71772 | 0,30976 | 0,16786 |
| Wiyung | 0,30632 | 0,5946 | 0,53782 | 0,46815 |
| Dukuh Pakis | 0,22605 | 0,5331 | 0,72937 | 0,74038 |
| Gayungan | 1,05748 | 0,87082 | 0,29490 | 0,44055 |
| Jambangan | 0,69209 | 0,30493 | 0,28503 | 0,59600 |
| Tenggilis | | | | |
| Mejoyo | 0,70156 | 0,41300 | 0,18860 | 0,49777 |
| Gunung Anyar | 0,64542 | 0,74498 | 0,33916 | 0,14101 |
| Asemrowo | 0,44819 | 1,14734 | 1,03320 | 0,71743 |
| Bulak | 0,20670 | 0,57171 | 0,64744 | 0,61073 |
| Pakal | 0,44919 | 0,81290 | 0,56498 | 0,28668 |
| Sambikerep | 0,36003 | 0,78868 | 1,08573 | 1,06473 |

$$P = p_{ij} = \frac{x_{ij}}{x_{..}} \tag{10}$$

Matriks P merupakan matriks korespondensi dengan dimensi ukuran (I × J). Mencari baris yang terdapat pada elemen-elemen yang berada pada matriks r dan kolom yang terdapat pada elemen-elemen yang berada pada matriks c, kemudian diagonal matriks Dr elemen r dan Dc elemen c pada diagonal, terdapat pada persamaan (11) dan (12).

$$r_i = \sum_{j=1}^J p_{ij} = \sum_{j=1}^J \frac{x_{ij}}{x_{..}} \tag{11}$$

$$c_j = \sum_{i=1}^I p_{ij} = \sum_{i=1}^I \frac{x_{ij}}{x_{..}} \tag{12}$$

r_i merupakan massa baris dan c_j merupakan massa kolom. Vektor baris r dan kolom c terdapat pada persamaan (13).

$$r = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_i \end{bmatrix} \text{ dan } c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_j \end{bmatrix} \tag{13}$$

Langkah selanjutnya membentuk diagonal massa matriks baris dan kolom dari matriks P.

$$\text{Inersia (baris)} = \sum_i r_i (\tilde{r} - c)^T D_c^{-1} (\tilde{r}_i - c)$$

$$\text{Inersia (kolom)} = \sum_i c_i (\tilde{c} - r)^T D_r^{-1} (\tilde{c}_j - r)$$

Kontribusi relative atau korelasi baris ke-i atau kolom ke-j dengan komponen k merupakan kontribusi axis ke inersia baris ke-i atau kolom ke-j dalam dimensi ke-k dinyatakan dalam persen inersia baris ke-i atau kolom ke-j. Kontribusi baris ke-i menuju inersia $\frac{r_i d_{ik}^2}{\lambda_k}$. Kontribusi kolom ke-j

$$\text{menuju inersia} = \frac{c_j d_{jk}^2}{\lambda_k}$$

2) Jarak Euclidean

Jarak Euclidean merupakan salah satu cara untuk mengetahui ukuran jarak antar objek pada titik yang berbeda (similarity). Hal ini digunakan untuk mengukur perbedaan

yang menggambarkan karakteristik dan pola dalam objek [8].

$$d(\mathbf{F}, \mathbf{G}) = \sqrt{\sum_{i=1}^k (\mathbf{F}_i - \mathbf{G}_i)^2} \quad (14)$$

Nilai $d(\mathbf{F}, \mathbf{G})$ merupakan jarak *Euclidean* antara titik koordinat profil baris dengan titik koordinat profil kolom, dimana nilai \mathbf{F} yaitu nilai koordinat titik pada profil, \mathbf{G} yaitu nilai koordinat titik pada kolom, dan k merupakan banyaknya solusi dimensi. Nilai \mathbf{F}_i merupakan nilai koordinat profil baris dimensi ke- i dan nilai \mathbf{G}_i merupakan nilai koordinat profil kolom dimensi ke- i [8].

III. METODOLOGI PENELITIAN

A. Sumber Data

Penelitian menggunakan data indikator pertumbuhan penduduk Kota Surabaya tahun 2020 dengan jumlah 31 Kecamatan yang didapatkan dari Dinas Kependudukan dan Pencatatan Sipil Surabaya.

B. Variabel Penelitian

Variabel penelitian yang digunakan dalam penelitian yang disajikan dalam Tabel 2.

C. Struktur Data

Struktur data yang digunakan dalam penelitian ini terdapat pada Tabel 3.

D. Langkah Analisis Data

Langkah-langkah analisis yang digunakan dalam penelitian ini menggunakan dua metode yaitu analisis cluster dan analisis korespondensi adalah sebagai berikut.

1) Data Sekunder

Mengumpulkan data sekunder berupa data indikator pertumbuhan penduduk per wilayah Kecamatan di Kota Surabaya tahun 2020.

2) Indikator Pertumbuhan Penduduk

Mendeskripsikan indikator pertumbuhan penduduk, yaitu jumlah kelahiran, jumlah kematian, jumlah pindah keluar, dan jumlah pindah datang pada tahun 2020 per wilayah Kecamatan di Kota Surabaya.

3) Metode Analisis Cluster

Berdasarkan langkah analisis dengan menggunakan metode analisis *cluster* adalah sebagai berikut: (a) Melakukan analisis *cluster hierarki* menggunakan metode *Ward's* dengan memperoleh jumlah optimum menggunakan *Pseudo F-statistic*. (b) Melakukan interpretasi hasil analisis *cluster*.

4) Analisis Korespondensi

Berdasarkan langkah analisis dengan menggunakan analisis korespondensi diantaranya, melakukan uji Independen pada data indikator pertumbuhan penduduk per wilayah Kecamatan di Kota Surabaya tahun 2020.

Melakukan uji Korespondensi pada data indikator pertumbuhan penduduk per wilayah Kecamatan di Kota Surabaya tahun 2020. Langkah-langkah dalam mengetahui kecenderungan dengan analisis korespondensi sebagai berikut: (a) Menyusun matriks profil baris dan kolom; (b) Menghitung profil baris dan kolom; (c) Menentukan nilai singular dekomposisi (SVD); (d) Menghitung koordinat profil baris dan kolom; (e) Menentukan nilai inersia; (f)

Menentukan nilai kontribusi *relative* dan kontribusi mutlak; (g) Menghitung jarak *Euclidean*.

5) Visualisasi

Visualisasi plot yang terbentuk pada masing-masing wilayah Kecamatan untuk mendeskripsikan indikator pertumbuhan penduduk. Kemudian menginterpretasi hasil analisis korespondensi dan menarik kesimpulan dan saran.

IV. ANALISIS DAN PEMBAHASAN

A. Deskripsi Kecamatan di Kota Surabaya Berdasarkan Indikator Pertumbuhan Penduduk

Pertumbuhan penduduk terbanyak dari indikator pertumbuhan penduduk adalah Kecamatan Kenjeran dengan jumlah kelahiran sebanyak 3443 jiwa, kematian sebanyak 1122 jiwa, serta pindah keluar sebanyak 1313 jiwa, dan pindah datang sebanyak 1828 jiwa. Pertumbuhan penduduk terbanyak dari indikator jumlah kelahiran adalah Kecamatan Sawahan sebanyak 3934 jiwa, sedangkan pertumbuhan penduduk terbanyak dari jumlah pindah datang adalah Kecamatan Kenjeran sebanyak 1828 jiwa.

B. Analisis Cluster Terhadap Indikator Pertumbuhan Penduduk Kota Surabaya Tahun 2020

Analisis *cluster* digunakan untuk mengelompokkan Kecamatan di Kota Surabaya berdasarkan indikator pertumbuhan penduduk. Pengelompokan *cluster* menggunakan metode *Ward's*. Berikut hasil analisis *cluster* pada data indikator pertumbuhan penduduk Kota Surabaya tahun 2020.

Gambar 1. menunjukkan *cluster-cluster* yang terbentuk berdasarkan indikator pertumbuhan penduduk secara visual dengan dendrogram menggunakan data hasil *Z-score*.

Setelah dilakukan analisis *cluster* menggunakan metode *Ward's*, dendrogram tersebut belum dapat menunjukkan banyaknya *cluster* optimum yang terbentuk. Untuk menentukan banyaknya *cluster* optimum yang terbentuk digunakan metode *Pseudo F-statistic*. *Cluster* yang memiliki nilai *Pseudo F-statistic* tertinggi menunjukkan *cluster* tersebut mamou memberikan hasil yang optimum.

Nilai *Pseudo F-statistic* dengan *cluster* yang beranggotakan 2, 3, dan 4 yang dapat dilihat pada Tabel 4.

Nilai *Pseudo F-statistic* paling besar ditunjukkan oleh *cluster* dengan jumlah 3 kelompok. Anggota masing-masing kelompok ditampilkan pada Tabel 5. Klasifikasi dari setiap indikator yang menunjukkan deskripsi kelompok wilayah baru yang terbentuk terdapat pada Tabel 6.

Persentase pertumbuhan penduduk dihitung dengan cara (kelahiran – kematian) × jumlah penduduk / 100% sehingga menghasilkan *cluster 1* sebesar 1,59% maka *cluster 1* dapat diklasifikasikan sedang karena pertumbuhannya antara 1%-2%, *cluster 2* sebesar 9,824% maka dapat diklasifikasikan cepat karena pertumbuhan penduduknya lebih dari 2%, dan *cluster 3* sebesar 0,547% maka dapat diklasifikasikan lambat karena pertumbuhannya kurang dari 1%.

C. Analisis Korespondensi Terhadap Indikator Pertumbuhan Penduduk Kota Surabaya Tahun 2020

Analisis korespondensi hasil pengelompokan metode *Ward's* indikator pertumbuhan penduduk Kota Surabaya tahun 2020. Sebelum dilakukan analisis korespondensi,

dilakukan pengujian asumsi independensi untuk mengetahui apakah data indikator pertumbuhan penduduk dengan wilayah Kecamatan di Kota Surabaya memiliki hubungan atau tidak. Berikut merupakan pengujian asumsi dan analisis korespondensi terhadap indikator pertumbuhan penduduk Kota Surabaya tahun 2020.

1) Uji Independensi

Uji independensi *cluster* sedang, cepat, dan lambat yang akan dijelaskan dengan hipotesis,

H_0 : tidak ada hubungan antara indikator pertumbuhan penduduk dengan wilayah per Kecamatan

H_1 : ada hubungan antara indikator pertumbuhan penduduk dengan wilayah per Kecamatan

Pengujian *Chi-Square cluster* sedang diperoleh nilai χ^2_{hitung} sebesar 500,730 lebih besar dari nilai $\chi^2_{0,05;36}$ sebesar 50,998 selanjutnya, pengujian *Chi-Square cluster* cepat diperoleh nilai χ^2_{hitung} sebesar 579,747 lebih besar dari nilai $\chi^2_{0,05;36}$ sebesar 21,026, dan pengujian *Chi-Square cluster* lambat diperoleh nilai χ^2_{hitung} sebesar 726,601 lebih besar dari nilai $\chi^2_{0,05;36}$ sebesar 50,998. Maka keputusan dari *cluster* sedang, cepat, dan lambat H_0 tolak sehingga ada hubungan antara indikator pertumbuhan penduduk dengan wilayah per Kecamatan (Tabel 7).

2) Reduksi Dimensi

Reduksi dimensi digunakan untuk mereduksi dimensi indikator pertumbuhan penduduk dari *cluster* sedang, cepat, dan lambat untuk mengetahui hasil kecenderungan pertumbuhan penduduk dengan wilayah Kecamatan Kota Surabaya tahun 2020 (Tabel 8).

Dimensi 1 memiliki nilai inersia/variansi sebesar 0,005 dengan proporsi sebesar 0,616 yang berarti bahwa dimensi 1 dapat menjelaskan keragaman data sebesar 61,6% dari keseluruhan data. Dimensi 2 memiliki nilai inersia/variansi sebesar 0,003 dengan proporsi sebesar 0,342 yang berarti bahwa dimensi 2 dapat menjelaskan keragaman data sebesar 34,2% dari keseluruhan data. Jadi, dengan menggunakan 2 dimensi didapatkan total kumulatif sebesar 95,7% untuk kedua dimensi (Tabel 9).

Dimensi 1 memiliki nilai inersia/variansi sebesar 0,012 dengan proporsi sebesar 0,861 yang berarti bahwa dimensi 1 dapat menjelaskan keragaman data sebesar 86,1% dari keseluruhan data. Dimensi 2 memiliki nilai inersia sebesar 0,001 dengan proporsi sebesar 0,108 yang berarti bahwa dimensi 2 dapat menjelaskan keragaman data sebesar 10,8% dari keseluruhan data. Jadi, dengan menggunakan 2 dimensi didapatkan total kumulatif sebesar 96,9% untuk kedua dimensi (Tabel 10).

Dimensi 1 memiliki nilai inersia sebesar 0,014 dengan proporsi sebesar 0,671 yang berarti dimensi 1 dapat menjelaskan keragaman data sebesar 67,1% dari keseluruhan data. Dimensi 2 memiliki nilai inersia sebesar 0,005 dengan proporsi sebesar 0,257 yang berarti dimensi 2 menjelaskan keragaman data sebesar 25,7% dari keseluruhan data. 2 dapat dimensi didapatkan total kumulatif sebesar 92,8% untuk kedua dimensi.

3) Plot Korespondensi

Gambaran mengenai pola kecenderungan antara indikator pertumbuhan penduduk dengan wilayah Kecamatan Kota

Surabaya tahun 2020 pada *cluster* sedang, cepat, dan lambat. (Gambar 2, Gambar 3, dan Gambar 4).

Gambar 5, Gambar 6, Gambar 7 kecenderungan dapat diketahui secara visual dan hanya beberapa saja yang dapat diketahui dengan jelas melalui jarak kedekatan antar indikator pertumbuhan penduduk Kota Surabaya tahun 2020. Untuk mengetahui jarak kedekatannya maka perlu dilakukan perhitungan jarak *Euclidean*.

4) Jarak Euclidean

Jarak *Euclidean* digunakan untuk melihat keterkaitan antara indikator pertumbuhan penduduk dengan wilayah per Kecamatan secara matematis. Berikut merupakan jarak *Euclidean cluster* sedang, cepat, dan lambat. E_1 merupakan jumlah kelahiran, E_2 merupakan jumlah kematian, E_3 merupakan jumlah pindah keluar, dan E_4 merupakan jumlah pindah datang (Tabel 11).

Gambar 5 menunjukkan indikator pertumbuhan penduduk dengan jumlah kelahiran yang memiliki warna hijau tua dan indikator pertumbuhan penduduk dengan jumlah pindah datang yang memiliki warna hijau muda. Indikator pertumbuhan penduduk dengan jumlah kematian yang memiliki warna *orange* dan indikator pertumbuhan penduduk dengan jumlah pindah keluar yang memiliki warna merah. Warna abu-abu merupakan Kecamatan yang bukan *cluster* klasifikasi sedang.

Gambar 6 dan Tabel 12 menunjukkan indikator pertumbuhan penduduk dengan jumlah kelahiran yang memiliki warna hijau tua dan indikator pertumbuhan penduduk dengan jumlah pindah datang yang memiliki warna hijau tua. Indikator pertumbuhan penduduk dengan jumlah kematian yang memiliki warna merah dan indikator pertumbuhan penduduk dengan jumlah pindah keluar yang memiliki warna *orange*. Warna abu-abu merupakan Kecamatan yang bukan *cluster* klasifikasi cepat.

Gambar 7 dan Tabel 13 menunjukkan indikator pertumbuhan penduduk dengan jumlah kelahiran yang memiliki warna hijau tua dan indikator pertumbuhan penduduk dengan jumlah pindah datang yang memiliki warna hijau muda. Indikator pertumbuhan penduduk dengan jumlah kematian yang memiliki warna merah dan indikator pertumbuhan penduduk dengan jumlah pindah keluar yang memiliki warna *orange*. Warna abu-abu merupakan Kecamatan yang bukan *cluster* klasifikasi lambat.

D. Hasil Kecenderungan Indikator Pertumbuhan Penduduk dengan Menggunakan Analisis Cluster dan Analisis Korespondensi

Hasil kelompok yang terbentuk dari analisis *cluster* dan hasil kecenderungan dari analisis korespondensi memiliki hasil yang saling mendukung. Pada analisis *cluster* yang terbentuk 3 kelompok berdasarkan indikator pertumbuhan penduduk selanjutnya dilakukan analisis korespondensi menggunakan hasil pengelompokan analisis *cluster* berdasarkan jarak *Euclidean* yang terdekat maka dihasilkan pengelompokan yang dapat memperkuat hasil kecenderungan analisis korespondensi.

Gambar 8 didapatkan 3 *cluster* yang terbentuk. Kecamatan yang memiliki warna kuning merupakan pertumbuhan penduduk yang memiliki nilai persentase 1,59% dengan klasifikasi sedang. Kecamatan yang memiliki warna hijau merupakan pertumbuhan penduduk yang memiliki nilai

persentase 9,824% dengan klasifikasi cepat. Dan Kecamatan yang memiliki warna merah merupakan pertumbuhan penduduk yang memiliki nilai persentase 0,547% dengan klasifikasi lambat. Masing-masing anggota *cluster* klasifikasi sedang terdapat 13 Kecamatan cenderung jumlah kelahiran, *cluster* klasifikasi cepat terdapat 5 Kecamatan cenderung jumlah pindah keluar, dan *cluster* klasifikasi lambat terdapat 13 Kecamatan cenderung jumlah kelahiran.

Gambar 9 yang terbentuk dari hasil pengelompokan menggunakan analisis *cluster* berdasarkan indikator pertumbuhan penduduk dengan wilayah Kecamatan di Kota Surabaya yaitu indikator dengan jumlah kelahiran cenderung pada wilayah Kecamatan Wonocolo, Sukolilo, Simokerto, Pabean Cantian, Tandes, Mulyorejo, Sukomanunggal, Semampir, Bulak, Lakar Santri, Wiyung, Dukuh Pakis, Asemrowo, dan Sambikerep. Jumlah kematian cenderung pada wilayah Kecamatan Gubeng, Sawahan, dan Genteng. Jumlah pindah keluar cenderung pada wilayah Kecamatan Karang Pilang, Tegalsari, Bubutan, Wonokromo, Tambak Sari, Jambangan, Tenggilis Mejoyo, dan Gayungan. Jumlah pindah datang cenderung pada wilayah Kecamatan Rungkut, Krembangan, Kenjeran, Gunung Anyar, Pakal, dan Benowo.

V. KESIMPULAN/RINGKASAN

Kecamatan yang memiliki jumlah kelahiran, jumlah kematian, jumlah pindah keluar paling banyak adalah Kecamatan Sawahan, dan jumlah pindah datang paling banyak adalah Kecamatan Kenjeran. Kecamatan yang memiliki jumlah kelahiran terendah adalah Kecamatan Gayungan, Kecamatan yang memiliki jumlah kematian terendah adalah Asemrowo. Dan Kecamatan yang memiliki jumlah pindah keluar dan pindah datang terendah adalah Kecamatan Bulak.

Pengelompokan data indikator pertumbuhan penduduk Kota Surabaya tahun 2020 yang dihasilkan dari analisis *cluster* metode Ward's adalah 3 *cluster* yang terbentuk. Cluster 1 mempunyai pertumbuhan penduduk terklasifikasi sedang dengan jumlah anggota 13 Kecamatan yaitu Kecamatan simokerto, Tandes, Wonocolo, Tegalsari, Bubutan, Sukolilo, Sukomanunggal, Pabean Cantian, Mulyorejo, Karang Pilang, Rungkut, Krembangan, dan Gubeng. Cluster 2 mempunyai pertumbuhan penduduk terklasifikasi cepat dengan jumlah anggota 5 Kecamatan terdiri dari Kecamatan Sawahan, Tambak Sari, Wonokromo, Semampir, dan Kenjeran. Cluster 3 mempunyai pertumbuhan

penduduk terklasifikasi lambat dengan jumlah anggota 13 Kecamatan, terdiri dari Kecamatan Genteng, Lakar Santri, Benowo, Wiyung, Dukuh Pakis, Gayungan, Jambangan, Tenggilis Mejoyo, Gunung Anyar, Asemrowo, Bulak, Pakal, dan Sambikerep.

Hasil analisis korespondensi yang dihasilkan dari pengelompokan menggunakan analisis *cluster* metode Ward's pada visualisasi plot jarak Euclidean didapatkan *cluster* 1 cenderung memiliki kedekatan jarak dengan indikator jumlah kelahiran yaitu 7 Kecamatan, *cluster* 2 cenderung memiliki kedekatan jarak dengan indikator jumlah pindah keluar yaitu 2 Kecamatan, dan *cluster* 3 cenderung memiliki kedekatan jarak dengan indikator jumlah kelahiran yaitu 6 Kecamatan.

Saran yang dapat diberikan kepada pemerintah Kota Surabaya agar memberikan informasi untuk mengontrol pertumbuhan penduduk, sebagai pertimbangan upaya program Kota Surabaya untuk mengendalikan pertumbuhan penduduk. Untuk penelitian selanjutnya dapat menggunakan data pertumbuhan penduduk beberapa tahun terakhir.

UCAPAN TERIMA KASIH

Penulis mengucapkan terima kasih kepada Dinas Kependudukan dan Pencatatan Sipil Kota Surabaya yang sudah mengizinkan menggunakan data untuk bahan penelitian.

DAFTAR PUSTAKA

- [1] BPS Provinsi Jawa Timur, *Provinsi Jawa Timur dalam Angka 2018*, 1st ed. Surabaya: BPS Provinsi Jawa Timur, 2018.
- [2] U. Bambang, *Geografi Membuka Cakrawala Dunia*, 2nd ed. Jakarta: Pusat Departemen Pendidikan Nasional, pp. 202, 2009, isbn: 978-979-068-776-9.
- [3] BPS, *Kota Surabaya dalam Angka Tahun 2020*, 1st ed. Surabaya: Badan Pusat Statistik Kota Surabaya, 2021.
- [4] A. J. R. and W. W. D., *Applied Multivariate Statistical Analysis*, 6th ed. United States: Prentice Hall, 2007, isbn: 9780131877153.
- [5] J. Sarwono, *Statistik Itu Mudah: Panduan Lengkap untuk Belajar Komputasi Statistik Menggunakan SPSS 16*, 1st ed. Yogyakarta: CV. Andi Offset, 2009.
- [6] J. Supranto, *Analisis Multivariat: Arti dan Interpretasi*, 1st ed. Jakarta: PT Rineka Cipta, 2018.
- [7] A. R. Orpin and V. E. Kostylev, "Towards a statistically valid method of textural sea floor characterization of benthic habitats," *Mar. Geol.*, vol. 225, no. 1-4, pp. 209-222, 2006, doi: 10.1016/j.margeo.2005.09.002.
- [8] M. Greenacre, *Correspondence Analysis in Practise*, 2nd ed. Barcelona: Universitas Pompeu Febra, 2007, isbn: 1-58488-616-1.
- [9] A. Agresti, *Categorical Data Analysis*, 2nd ed. New York: John Wiley & Sons, Inc, 2002, doi: 10.1002/0471249688, isbn: 9780471360933.