

Received: 23 Agustus 2021

Revised: 21 Desember 2021

Accepted: 30 Desember 2021

Published: 31 Desember 2021

## Pemetaan Jenis Tindak Kriminal di Indonesia Berdasarkan Karakteristik Wilayah Menggunakan *Canonical Correspondence Analysis*

Luthfi Ghiffari<sup>1, a)</sup>, Nurul Gusriani<sup>2, b)</sup>, Kankan Parmikanti<sup>3, c)</sup>

<sup>1,2,3</sup> *Departemen Matematika, Fakultas MIPA, Universitas Padjadjaran  
Jl. Raya Bandung Sumedang KM 21 Jatinangor Sumedang 45363*

Email: <sup>a)</sup>[luthfi17002@mail.unpad.ac.id](mailto:luthfi17002@mail.unpad.ac.id), <sup>b)</sup>[nurul.gusriani@unpad.ac.id](mailto:nurul.gusriani@unpad.ac.id), <sup>c)</sup>[parmikanti@unpad.ac.id](mailto:parmikanti@unpad.ac.id)

### Abstract

Criminal acts are all forms of actions that are economically and psychologically harmful that violate the law, social norms, and religion. The occurrence of criminal acts in an area can be related to various community aspects including education, demography, and economy. Each region in Indonesia must have different characteristics, so that the tendency for crime to occur will be different in each region. This research was conducted with the aim of mapping the types of crimes based on regional characteristics in Indonesia. The unit of observation in this study were 34 provinces in Indonesia. The data used is secondary data obtained from the Badan Pusat Statistik Indonesia in 2019. The method used in this study is the Canonical Correspondence Analysis (CCA). Based on the research, a triplot map that explains the tendency of types of crime based on regional characteristics in Indonesia is obtained. The resulting triplot map can explain 90.6% of the total inertia captured by CCA. Through the Monte Carlo permutation test, it is shown that there is a significant relationship between regional characteristics and the distribution of types of crime.

**Keywords:** canonical correspondence analysis, crime, monte carlo permutation test, triplot map, inertia

### Abstrak

Tindak kriminal merupakan segala macam bentuk tindakan yang merugikan secara ekonomis dan psikologis yang melanggar hukum, norma-norma sosial, serta agama. Terjadinya tindak kriminal di suatu wilayah dapat berhubungan dengan berbagai karakter masyarakat diantaranya pendidikan, kependudukan, dan ekonomi. Setiap wilayah di Indonesia pasti memiliki karakteristik yang berbeda-beda, sehingga kecenderungan terjadinya tindak kriminal akan berbeda pula di setiap wilayahnya. Penelitian ini dilakukan dengan tujuan untuk membuat pemetaan jenis tindak kriminal berdasarkan karakteristik wilayah di Indonesia. Unit pengamatan pada penelitian ini adalah 34 provinsi di Indonesia. Data yang digunakan adalah data sekunder yang didapat dari Badan Pusat Statistik Indonesia pada tahun 2019. Metode yang digunakan pada penelitian adalah metode *Canonical Correspondence Analysis* (CCA). Hasil penelitian didapatkan pemetaan jenis-jenis tindak kriminal berdasarkan karakteristik wilayah di Indonesia pada peta triplot. Peta triplot yang dihasilkan dapat menjelaskan 90.6% dari total inersia yang ditangkap oleh CCA. Melalui uji permutasi Monte Carlo, ditunjukkan bahwa terdapat hubungan yang signifikan antara karakteristik wilayah dengan penyebaran jenis tindak kriminal.

**Kata-kata kunci:** *canonical correspondence analysis*, kriminal, uji permutasi monte carlo, peta triplot, inersia

## PENDAHULUAN

Tindak kriminal merupakan salah satu fenomena sosial yang sangat rentan dipengaruhi oleh dinamika kehidupan di masyarakat. Tindak kriminal di masyarakat dapat terjadi akibat faktor kebutuhan yang semakin meningkat, sementara pendapatan masyarakat tidak mampu untuk memenuhi kebutuhan tersebut. Tingkat kriminalitas di suatu daerah juga dapat meningkat karena dipengaruhi oleh berbagai aspek di masyarakat seperti politik, ekonomi, sosial, dan hal-hal lain yang berhubungan dengan keamanan negara.

Angka kriminalitas di Indonesia selalu mengalami perubahan seiring dengan dinamika kehidupan masyarakat. Data dari publikasi Badan Pusat Statistik Indonesia memperlihatkan bahwa banyaknya kejadian kejahatan yang terjadi di Indonesia pada tahun 2019 adalah 269.324 kejadian. Sementara tingkat risiko penduduk untuk terkena tindak kejahatan adalah 103 orang per 100.000 orang. Banyaknya kejadian kejahatan tersebut merupakan gambaran seluruh peristiwa kejahatan secara umum dan merupakan akumulasi dari data yang tercatat di seluruh kepolisian tingkat daerah (Polda). Adapun berdasarkan Kitab Undang-Undang Hukum Pidana (KUHP) Republik Indonesia, pengelompokan tindak kejahatan didasari atas tiga kriteria yaitu target dari kejadian kejahatan (orang, harta benda, ketertiban umum, dan sebagainya), tingkat keseriusan kejahatan (kejahatan terhadap hak milik, kejahatan terhadap fisik, dan sebagainya), serta bagaimana kejahatan dilakukan (dengan kekerasan, tanpa kekerasan, dan sebagainya). Berbagai tindak kejahatan tersebut terjadi di Indonesia dan tersebar di seluruh provinsi dengan angka yang beragam. Angka kejahatan tersebut kemudian perlu diteliti untuk melihat bagaimana penyebarannya, serta hubungannya dengan karakteristik wilayah di Indonesia.

Penelitian mengenai faktor-faktor yang memengaruhi tindak kriminal pernah dilakukan oleh Hardianto (2009) dengan menggunakan analisis regresi berganda, yang menyatakan bahwa tingkat upah berpengaruh negatif sementara pengeluaran pembangunan pemerintah berpengaruh positif terhadap tingkat kriminalitas di Indonesia. Penelitian serupa juga dilakukan oleh Purwanti (2019), yang menyatakan bahwa faktor yang signifikan dalam memengaruhi tindak kriminal di Provinsi Jawa Timur diantaranya adalah Produk Domestik Regional Bruto (PDRB) per kapita, tingkat pengangguran, kepadatan penduduk, dan tingkat kemiskinan. Nurfitasari (2017) melakukan pemetaan tindak kriminal di wilayah Madiun dengan menggunakan analisis korespondensi berganda. Penelitian tersebut berhasil memetakan jenis-jenis tindak kriminal di wilayah Madiun berdasarkan pada kriteria pelaku serta waktu dan tempat kejadian perkara.

Pada penelitian ini, akan dilakukan pemetaan wilayah Indonesia, jenis-jenis tindak kriminal, serta karakteristik wilayah yang berkaitan. Pemetaan dilakukan dengan menggunakan metode *Canonical Correspondence Analysis* (CCA). CCA adalah metode multivariat yang umum digunakan untuk menjelaskan hubungan antara kumpulan spesies biologi dan lingkungannya (Braak, 1986). CCA merupakan bentuk kanonik dari analisis korespondensi (Legendre and Legendre, 1998). Sebagai perluasan dari analisis korespondensi, CCA memanfaatkan data tambahan berupa variabel independen untuk melakukan pemetaan variabel dependen. Seperti analisis korespondensi, CCA memaksimalkan korelasi antara variabel dependen dengan sampel, tetapi unit sampel pada CCA dibatasi sebagai kombinasi linear dari beberapa variabel independen. Dengan menggunakan metode CCA, sebaran tindak kriminal, unit sampel, serta karakteristik wilayah dapat dipetakan pada satu pemetaan. Dengan demikian, hubungan antara karakteristik wilayah dengan jenis tindak kriminal dapat terlihat secara langsung.

## METODOLOGI

### Data

Data variabel dependen yang digunakan pada penelitian ini adalah data sekunder berupa jenis-jenis tindak kriminal di berbagai provinsi di Indonesia. Data diperoleh dari publikasi Badan Pusat Statistik Indonesia pada tahun 2020. Jenis-jenis tindak kriminal sebagai variabel dependen terdiri atas kejahatan terhadap nyawa ( $Y_1$ ), kejahatan terhadap fisik ( $Y_2$ ), terhadap kesusilaan ( $Y_3$ ), terhadap kemerdekaan

orang ( $Y_4$ ), terhadap hak milik ( $Y_5$ ), kejahatan terkait narkoba ( $Y_6$ ), serta kejahatan terkait penipuan, penggelapan, dan korupsi ( $Y_7$ ).

Karakteristik setiap provinsi di Indonesia diwakili oleh berbagai variabel independen yang diduga memiliki hubungan dengan penyebaran tindak kriminal di suatu wilayah. Data variabel independen yang digunakan diperoleh dari Badan Pusat Statistik (2014, 2019a, 2019b, 2020a, 2020c, 2020d, 2020e, 2021a, 2021b, 2021c). Variabel independen yang digunakan pada penelitian disajikan pada TABEL 1.

**TABEL 1.** Variabel Independen Penelitian

Variabel	Keterangan	Skala Pengukuran
$X_1$	Angka Melek Huruf (%)	Rasio
$X_2$	Angka Putus Sekolah SMP (%)	Rasio
$X_3$	Angka Tidak Bersekolah Usia 16-18 Tahun (%)	Rasio
$X_4$	Rata-Rata Lama Sekolah (tahun)	Rasio
$X_5$	Kepadatan Penduduk (jumlah penduduk/km <sup>2</sup> )	Rasio
$X_6$	Rasio Jenis Kelamin Laki-laki Terhadap Perempuan (%)	Rasio
$X_7$	Persentase Penduduk Daerah Perkotaan (%)	Rasio
$X_8$	Persentase Penduduk Miskin (%)	Rasio
$X_9$	Tingkat Pengangguran Terbuka (%)	Rasio
$X_{10}$	Rata-Rata Upah Bersih Buruh/Karyawan/Pegawai (Rp)	Rasio
$X_{11}$	Inklusi Keuangan (%)	Rasio
$X_{12}$	Produk Domestik Regional Bruto (Rp)	Rasio
$X_{13}$	Pengeluaran Provinsi (Rp)	Rasio
$X_{14}$	Rata-Rata Pengeluaran Rumah Tangga (Rp)	Rasio
$X_{15}$	Indeks Gini	Interval
$X_{16}$	Indeks Pembangunan Manusia	Interval
$X_{17}$	Persentase Korban yang Melapor pada Polisi (%)	Rasio

### Metode Penelitian

#### *Canonical Correspondence Analysis (CCA)*

*Canonical Correspondence Analysis (CCA)* dapat dilakukan dengan melalui prosedur rata-rata timbal balik. Jika  $\mathbf{Y}$  merupakan matriks data terdiri dari sebanyak  $n$  sampel (baris) dan  $p$  variabel (kolom), dan  $y_{ij}$  merepresentasikan respon variabel ke- $j$  ( $j = 1, 2, \dots, p$ ) pada sampel ke- $i$  ( $i = 1, 2, \dots, n$ ), maka matriks  $\mathbf{Y}$  dapat dinyatakan seperti pada persamaan (1).

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \dots & y_{1p} \\ y_{21} & y_{22} & \dots & y_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \dots & y_{np} \end{bmatrix} \quad (1)$$

Tahapan pertama pada algoritma rata-rata timbal balik adalah menyatakan skor pada setiap sampel sebagai sembarang bilangan tak sama dan tak nol. Pemilihan sembarang bilangan tidak akan memengaruhi hasil perhitungan, atau dijamin akan mencapai konvergensi. Jika  $v_i^{(a;0)}$  merupakan skor sampel awal ke- $i$  pada sumbu ke- $a$  ( $a = 1, 2, \dots, r$ ), maka dapat dihitung skor variabel  $u_j^{(a;t)}$  untuk iterasi ke- $t$  ( $t = 1, \dots, m$ ), sebagai rata-rata terbobot dari skor sampel awal melalui persamaan (2) yaitu:

$$u_j^{(a;t)} = \sum_{i=1}^n \frac{y_{ij} v_i^{(a;t-1)}}{y_{i+}} \quad (2)$$

dimana  $y_{i+} = y_{i1} + y_{i2} + \dots + y_{ip}$  merupakan jumlah respon seluruh variabel dari pada sampel ke- $i$ .

Secara berkoresponden, skor sampel LC dapat diperoleh dengan menghitung rata-rata terbobot dari skor variabel, dimana bobot yang digunakan merupakan jumlah respon relatif variabel  $y_{+j}$ , sehingga skor sampel LC  $v_i^{(a;t)}$  dapat dinyatakan seperti pada persamaan (3):

$$v_i^{(a;t)} = \sum_{j=1}^p \frac{y_{ij}u_j^{(a;t)}}{y_{+j}} \tag{3}$$

dimana  $y_{+j} = y_{1j} + y_{2j} + \dots + y_{nj}$  merupakan jumlah respon variabel ke- $j$  dari seluruh sampel. Setelah menghitung skor sampel, tahap berikutnya adalah menghitung skor sampel *linear combination* (LC). Jika  $\mathbf{R}$  merupakan matriks diagonal berukuran  $n \times n$  dengan entri  $y_{i+}$  pada kolom ke- $i$  dan baris ke- $i$ , dan  $\mathbf{X} = \{x_{ik}\}$  merupakan matriks berukuran  $n \times (q + 1)$  terdiri atas respon variabel independen ke- $k$  ( $k = 1, 2, \dots, q$ ) pada sampel ke- $i$ , ditambah sebuah kolom dengan entri 1 (untuk mendapatkan koefisien *intercept* pada regresi), maka  $\mathbf{R}$  dan  $\mathbf{X}$  secara berturut-turut dapat dinyatakan pada persamaan (4) dan persamaan (5).

$$\mathbf{R} = \begin{bmatrix} y_{1+} & 0 & \dots & 0 \\ 0 & y_{2+} & \dots & \vdots \\ \vdots & \vdots & \ddots & 0 \\ 0 & \dots & 0 & y_{n+} \end{bmatrix} \tag{4}$$

$$\mathbf{X} = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1q} \\ 1 & x_{21} & x_{22} & \dots & x_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{nq} \end{bmatrix} \tag{5}$$

Jika  $\mathbf{v} = (v_1^{(a;t)}, \dots, v_n^{(a;t)})$  merupakan vektor berisi skor sampel dari persamaan (3), maka koefisien regresi dapat dihitung melalui operasi matriks menggunakan persamaan (6) yaitu:

$$\mathbf{b}^{(a;t)} = (\mathbf{X}'\mathbf{R}\mathbf{X})^{-1}\mathbf{X}'\mathbf{R}\mathbf{v}^{(a;t)} \tag{6}$$

dengan  $\mathbf{b} = (b_0^{(a;t)}, b_1^{(a;t)}, \dots, b_q^{(a;t)})'$  merupakan vektor berisi koefisien regresi, atau selanjutnya disebut sebagai koefisien kanonik. Tahap selanjutnya dapat dihitung skor sampel LC  $\hat{v}_i^{(a;t)}$  sebagai kombinasi linier dari variabel independen seperti ditunjukkan pada persamaan (7).

$$\hat{v}_i^{(a;t)} = b_0^{(a;t)} + \sum_{j=1}^q b_j^{(a;t)} x_{ik}^{(a;t)} \tag{7}$$

Pada perhitungan sumbu kedua dan sumbu-sumbu berikutnya, harus diberikan kendala tambahan bahwa skor sampel yang dihitung adalah tegak lurus atau tidak berkorelasi dengan setiap sumbu sebelumnya. Proses tersebut dapat dilakukan melalui prosedur ortogonalisasi. Untuk menghitung skor sumbu ke- $a$  yang tegak lurus dengan sumbu-sumbu sebelumnya, dapat dihitung produk skalar  $\gamma^{(a)}$  pada persamaan (8):

$$\gamma^{(a)} = \sum_{i=1}^n \frac{y_{i+}}{y_{++}} v_i^{(a;t)} v_i^{(a-l;t)} \tag{8}$$

dengan  $l = 1, 2, \dots, (a - 1)$  jumlah sumbu-sumbu yang sudah dibuat. Untuk mendapatkan skor sampel  $\bar{v}_i^{(a)}$  yang tegak lurus dengan seluruh skor sampel  $v_i^{(a-l)}$ , maka dapat dilakukan perhitungan dengan menggunakan persamaan (9).

$$\bar{v}_i^{(a)} = v_i^{(a)} - \sum_{l=1}^{a-1} \gamma^{(a)} v_i^{(a-l)} \tag{9}$$

Setelah didapatkan skor variabel dan skor sampel, tahap berikutnya adalah melakukan normalisasi skor sampel dan mendapatkan estimasi nilai eigen. Legendre and Legendre (1998) menyajikan prosedur normalisasi melalui perhitungan *centroid* dan jumlah kuadrat dari skor sampel berturut-turut pada persamaan (10) dan persamaan (11).

$$d^{(a;t)} = \frac{\sum_{i=1}^n y_{i+} v_i^{(a;t)}}{y_{++}} \tag{10}$$

$$S^2^{(a;t)} = \frac{\sum_{i=1}^n y_{i+} (v_i^{(a;t)} - d^{(a;t)})^2}{y_{++}} \tag{11}$$

Skor sampel normal dapat dihitung dengan mengurangkan skor sampel dengan *centroid*, kemudian membagi hasil perhitungan dengan  $S^{(a;t)} = \sqrt{S^2^{(a;t)}}$ . Skor sampel normal  $v_i^{*(a;t)}$  dapat dinyatakan pada persamaan (12).

$$v_i^{*(a;t)} = \frac{v_i^{(a;t)} - d^{(a;t)}}{S^{(a;t)}} \tag{12}$$

Perhitungan diulangi kembali ke persamaan (2) dengan menggunakan skor sampel normal  $v_i^{*(a;t)}$ . Pada setiap akhir iterasi ke- $t$ ,  $S^{(a;t)}$  menyatakan estimasi untuk mendapatkan nilai eigen, yang merupakan ukuran untuk mengetahui jumlah informasi yang dipertahankan pada suatu sumbu. Selisih estimasi nilai eigen  $\delta$  dapat dihitung melalui persamaan (13):

$$\delta = |S^{(a;t)} - S^{(a;t-1)}| < \varepsilon \tag{13}$$

dengan  $\varepsilon$  merupakan nilai toleransi. Iterasi dihentikan jika tidak terjadi atau hampir tidak terjadi perubahan pada nilai  $S^{(a;t)}$ , sesuai dengan nilai toleransi yang digunakan.

Setelah mendapatkan skor dependen dan skor sampel, tahap selanjutnya adalah menghitung korelasi interseset. Korelasi interseset merupakan ukuran seberapa baik variasi variabel dependen dapat dijelaskan oleh setiap variabel independen, yang ekuivalen dengan korelasi antara skor sampel  $\bar{v}_i^{*(a;t)}$  hasil iterasi terakhir dengan variabel independen.

Untuk melakukan plot pada peta korespondensi, skor dependen dari sumbu pertama sampai terakhir digunakan sebagai koordinat setiap variabel dependen, sementara skor sampel digunakan sebagai koordinat setiap unit sampel. Variabel independen pada peta korespondensi kanonik dibuat dalam bentuk garis panah, dengan korelasi interseset sebagai koordinat ujung panah. Pada peta berdimensi rendah, koordinat-koordinat setiap titik merupakan hasil proyeksi dari peta dengan dimensi yang lebih tinggi. Untuk mengukur jumlah informasi yang ditangkap dari suatu peta triplot, dapat digunakan perhitungan total inersia. Sebelum melakukan perhitungan total inersia, terlebih dahulu dilakukan perhitungan nilai *chi-square*. Nilai *chi-square* didefinisikan sebagai jumlah kuadrat deviasi antara nilai observasi dengan nilai harapan, dibagi dengan nilai harapan seperti pada persamaan (14) yaitu:

$$\chi^2 = \sum_{i=1}^n \sum_{j=1}^p \frac{(y_{ij} - \hat{y}_{ij})^2}{\hat{y}_{ij}} \tag{14}$$

dengan nilai harapan  $\hat{y}_{ij} = (y_{i+} \times y_{+j})/y_{++}$ . Total inersia  $I$  didapatkan dengan membagi statistik *chi-square* dengan jumlah nilai observasi, atau dinyatakan pada persamaan (15).

$$I = \frac{\chi^2}{y_{++}} \tag{15}$$

*Monte Carlo Permutation Test*

Menurut Legendre and Legendre (1998), *Monte Carlo Permutation Test* merupakan uji yang menggunakan angka secara acak untuk mempelajari himpunan data asli atau perilaku dari metode statistika melalui simulasi. Uji permutasi ini tidak memerlukan asumsi terkait dengan distribusi variabel respon. Pada metode pemetaan, nilai observasi statistik uji akan dibandingkan dengan distribusi harapan statistik uji dibawah asumsi hipotesis nol. Berdasarkan perbandingan tersebut, dihitung peluang untuk mendapatkan hasil yang berbeda dari distribusi harapan statistik uji yang didapat melalui permutasi (Leps and Smilaur, 2003). Nilai *P-value* (peluang menolak  $H_0$  benar) dihitung melalui persamaan (16):

$$P_V = \frac{np_x + 1}{NP + 1} \quad (16)$$

dengan  $np_x$  merupakan banyaknya perulangan permutasi jika nilai statistik uji yang diperoleh lebih besar dari nilai observasi, dan  $NP$  menunjukkan banyaknya perulangan permutasi keseluruhan.

Hipotesis penelitian yang digunakan pada uji permutasi Monte Carlo serentak menyatakan bahwa respon (variabel dependen) dan variabel penjelas tidak memiliki hubungan (independen), atau dapat dinyatakan sebagai berikut:

$H_0$  : tidak terdapat hubungan antara variabel independen dan dependen

$H_1$  : terdapat hubungan antara variabel independen dan dependen

Dengan berlakunya asumsi  $H_0$ , maka variabel penjelas dapat secara acak digunakan untuk menjelaskan variabel respon manapun. Dengan melakukan hal tersebut berulang-ulang melalui permutasi, serta mencatat hasil statistik uji yang dihasilkan, maka akan didapat distribusi harapan dari statistik uji.

Statistik uji yang dihitung pada setiap perulangan disebut *pseudo-F*. *Pseudo-F* menyatakan rasio antara varians yang dijelaskan oleh pemetaan (dibagi dengan jumlah parameter) dengan varians residual (dibagi dengan derajat bebas). *Pseudo-F* dapat dihitung melalui persamaan (17) yaitu:

$$F_{pseudo} = \frac{\sum_{a=1}^r \lambda_a / q}{RSS / (n - q - 1)} \quad (17)$$

dengan  $\lambda_a$  menyatakan nilai eigen dari sumbu ke- $a$ , dan  $RSS$  (*residual sum squares*) merupakan selisih total inersia dengan jumlah nilai eigen keseluruhan sumbu. Setelah mendapatkan nilai observasi *pseudo-F*, maka tahap berikutnya adalah membandingkan nilai tersebut dengan distribusi statistik uji yang didapat melalui permutasi. Untuk menghitung *P-value* atau peluang menolak  $H_0$  benar, maka digunakan persamaan (16).

Uji permutasi Monte Carlo parsial digunakan untuk menilai kontribusi dari setiap variabel independen dalam menjelaskan variansi dari data. Hipotesis penelitian yang digunakan akan bersesuaian dengan variabel yang diuji. Jika terdapat sebanyak  $q$  variabel, maka kombinasi variabel yang diuji adalah sebanyak  $q$  kombinasi. Hipotesis yang digunakan adalah sebagai berikut:

1. Model 1 (Y bergantung pada  $X_1$ )  
 $H_0$ : tidak terdapat kontribusi variabel  $X_1$  dalam menjelaskan variansi data  
 $H_1$ : terdapat kontribusi variabel  $X_1$  dalam menjelaskan variansi data
2. Model 2 (Y bergantung pada  $X_1$  dan  $X_2$ )  
 $H_0$ : tidak terdapat kontribusi  $X_2$  pada model 1 dalam menjelaskan variansi data  
 $H_1$ : terdapat kontribusi  $X_2$  pada model 1 dalam menjelaskan variansi data
3. Model 3 (Y bergantung pada  $X_1$ ,  $X_2$ , dan  $X_3$ )  
 $H_0$ : tidak terdapat kontribusi  $X_3$  pada model 2 dalam menjelaskan variansi data  
 $H_1$ : terdapat kontribusi  $X_3$  pada model 2 dalam menjelaskan variansi data
4. Dilakukan hal serupa sehingga seluruh variabel yang digunakan masuk ke dalam model pemetaan.

Untuk melakukan seleksi pada variabel yang digunakan, dilakukan perhitungan *conditional effect* dari setiap variabel. *Conditional effect* suatu variabel menyatakan selisih antara varians yang dapat dijelaskan model awal (tanpa variabel terkait), dengan varians yang dapat dijelaskan model setelah variabel ditambahkan (Leps and Smilauer, 2003). Urutan penambahan variabel ke dalam model akan memengaruhi besarnya *conditional effect* variabel tersebut. Perhitungan *conditional effect* variabel ke- $k$  dapat dilakukan melalui persamaan (18).

$$\lambda_k = \sum_{a=1}^{r_k} \lambda_a^{(k)} - \sum_{a=1}^{r_{k-1}} \lambda_a^{(k-1)} \tag{18}$$

Statistik uji yang digunakan pada uji permutasi parsial juga disebut sebagai pseudo- $F$ . Pseudo- $F$  pada uji parsial menyatakan rasio antara kontribusi setiap variabel (diukur melalui *conditional effect*) dengan varians residual (dibagi dengan derajat bebas). Perhitungan pseudo- $F$  uji parsial dapat dilakukan menggunakan persamaan (19):

$$F_{pseudo_k} = \frac{\lambda_k}{RSS/(n - q - 1)} \tag{19}$$

dengan  $\lambda_k$  menyatakan nilai eigen atau *conditional effect* variabel ke- $k$ , dan  $RSS$  (*residual sum squares*) merupakan selisih jumlah nilai eigen keseluruhan sumbu dengan total inersia. Setelah mendapatkan nilai observasi pseudo- $F$ , maka tahap berikutnya adalah melakukan perhitungan  $P$ -value menggunakan persamaan (16).

## HASIL DAN PEMBAHASAN

### Gambaran Terjadinya Tindak Kriminal di Indonesia

Data yang digunakan sebagai variabel dependen yaitu jumlah berbagai jenis tindak kriminal berdasarkan provinsi di Indonesia pada tahun 2019. Untuk memberikan gambaran umum mengenai sebaran jenis tindak kriminal di Indonesia, disajikan statistika deskriptif dari data pada TABEL 2.

**TABEL 2.** Statistika Deskriptif Variabel Dependen

	$Y_1$	$Y_2$	$Y_3$	$Y_4$	$Y_5$	$Y_6$	$Y_7$
Nilai Maksimum	136	465	4817	399	8378	6338	5115
Nilai Minimum	1	33	108	0	107	2	49
Rata-rata	28.3	153.9	1146.5	67.7	2552.1	1072.8	1156.5
Median	18	119.5	666	34.5	1813.5	495.5	660.5
Deviasi Standar	28.9	107.2	1131	87.9	2082.4	1531.7	1319.6

TABEL 2 menunjukkan gambaran jumlah kasus tindak kriminal yang terjadi di Indonesia pada tahun 2019. Untuk  $Y_1$  atau kasus kejahatan terhadap nyawa, jumlah kasus terbanyak adalah 136 kasus yaitu di provinsi Sumatera Selatan. Untuk  $Y_2$  atau kasus kejahatan terhadap fisik, provinsi dengan kasus terbanyak yaitu 465 kasus adalah provinsi Jawa Barat.  $Y_3$  merepresentasikan kasus kejahatan terhadap kesusilaan, dimana jumlah kasus terbanyak adalah 4817 kasus yaitu di provinsi Sumatera Utara. Untuk  $Y_4$  atau kasus kejahatan terhadap kemerdekaan orang, jumlah kasus terbanyak adalah 399 kasus yaitu di provinsi Sulawesi Selatan. Untuk  $Y_5$  atau kasus kejahatan terhadap hak milik, provinsi dengan kasus terbanyak yaitu 8378 kasus adalah provinsi Sumatera Utara.  $Y_6$  atau kasus kejahatan terkait narkoba, dimana jumlah kasus terbanyak adalah 6338 kasus yaitu di provinsi DKI Jakarta. Kemudian  $Y_7$  atau kasus kejahatan terkait korupsi, jumlah kasus terbanyak adalah 5115 kasus yaitu di provinsi DKI Jakarta.

**Pemetaan Menggunakan *Canonical Correspondence Analysis (CCA)***

Dengan menggunakan *Canonical Correspondence Analysis (CCA)*, dihasilkan peta triplot yang memetakan jenis-jenis tindak kriminal, provinsi, serta karakteristik wilayah secara serentak. Ukuran seberapa besar informasi yang ditangkap oleh CCA dapat dihitung melalui proporsi jumlah nilai eigen terhadap total inersia. Hasil perhitungan disajikan pada TABEL 3.

**TABEL 3.** Nilai Eigen Peta Triplot

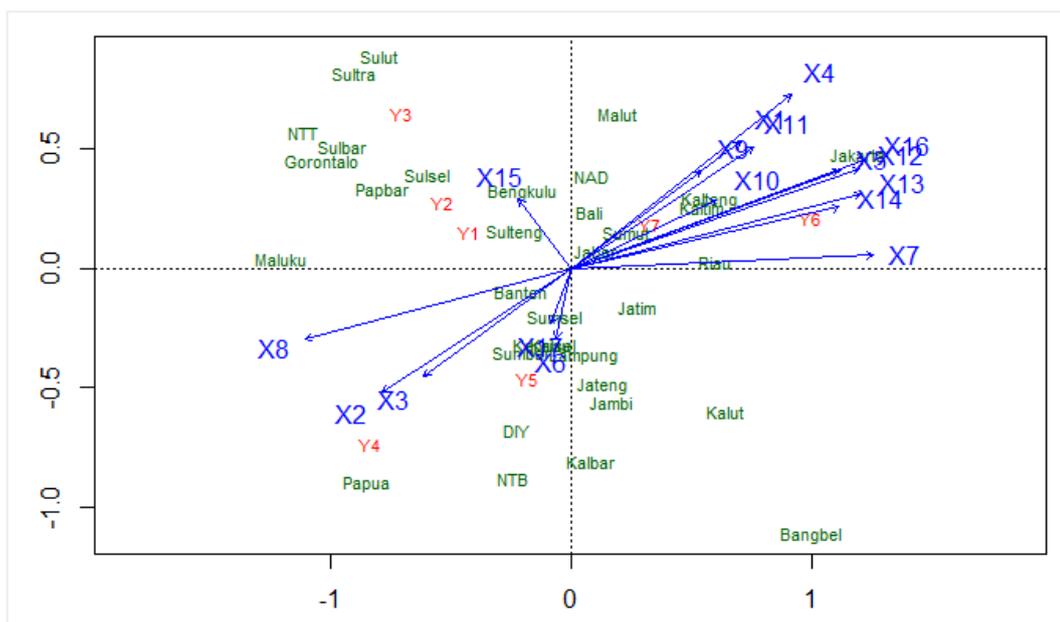
Dimensi	Nilai Eigen	Proporsi	Proporsi Kumulatif
<i>Constrained</i>	0.1558	78.24%	78.24%
<i>Unconstrained</i>	0.0433	21.76%	100%
Total	0.1992	100%	

Dimensi	Nilai Eigen	Proporsi	Proporsi Kumulatif
Dimensi 1	0.0987	63.39%	63.39%
Dimensi 2	0.0352	22.59%	85.98%
Dimensi 3	0.0131	8.41%	94.39%
Dimensi 4	0.0046	2.95%	97.34%
Dimensi 5	0.0035	2.25%	99.59%
Dimensi 6	0.0006	0.41%	100%
Total <i>Constrained</i>	0.1558	100%	

TABEL 3 menunjukkan bahwa jumlah variansi data yang dapat dijelaskan oleh peta triplot (*constrained*) adalah sebanyak 0.1558 atau sebesar 78.24% dari total inersia, sementara 21.76% sisanya tidak dapat ditangkap oleh peta triplot (*unconstrained*). Banyaknya dimensi maksimal yang dapat dibuat adalah sebanyak enam dimensi. Dimensi pertama memiliki nilai eigen sebesar 0.098 dan dapat menjelaskan variansi data sebanyak 63.39%. Dimensi kedua memiliki nilai eigen sebesar 0.035 atau dapat menjelaskan variansi data sebanyak 22.59%. Secara akumulatif, dimensi pertama dan kedua dapat menjelaskan sebagian besar informasi yang ditangkap oleh peta triplot atau sebesar 85.98%, maka peta yang digunakan adalah peta berukuran dua dimensi.

**Peta Triplot Awal**



**GAMBAR 1.** Peta Triplot Awal

GAMBAR 1 menunjukkan peta *Canonical Correspondence Analysis* (CCA) yang memetakan tiga komponen secara serentak. Terdapat tujuh variabel dependen, 34 sampel, dan 17 variabel independen. Banyaknya variabel yang digunakan pada penelitian mengakibatkan peta triplot memuat banyak komponen dan menjadi cenderung sulit untuk diinterpretasi. Oleh karena itu, akan dilakukan uji signifikansi menggunakan metode uji permutasi Monte Carlo untuk menentukan variabel yang berhubungan secara signifikan dengan penyebaran jenis-jenis tindak kriminal.

Uji permutasi Monte Carlo dilakukan untuk menguji signifikansi pemetaan/model baik secara serentak maupun menguji signifikansi dari setiap variabel yang digunakan. Statistik uji yang digunakan adalah *pseudo-F* seperti pada persamaan (17). Hasil pengujian dari uji permutasi Monte Carlo serentak ditunjukkan pada TABEL 4.

TABEL 4. Hasil Uji Permutasi Monte Carlo Serentak

	Df	Total Nilai Eigen	$F_{pseudo}$	P-Value
Model	17	0.1558	3.3845	0.001
Residual	16	0.04334		

Berdasarkan hasil uji permutasi Monte Carlo serentak sebanyak 999 kali permutasi, dapat dilihat bahwa *P-value* yang didapat adalah sebesar 0.001, lebih kecil dari taraf signifikansi  $\alpha = 0.05$ , maka hipotesis  $H_0$  ditolak. Hal ini menunjukkan bahwa terdapat hubungan yang signifikan antara penyebaran jenis-jenis tindak kriminal dengan karakteristik wilayah yang direpresentasikan melalui variabel independen.

Setelah melakukan uji permutasi Monte Carlo serentak, tahap selanjutnya adalah melakukan uji permutasi Monte Carlo parsial pada setiap variabel independen yang digunakan. Sebelum melakukan pengujian, terlebih dahulu dilakukan perhitungan *conditional effect* seperti pada persamaan (18) untuk mendapatkan nilai eigen dari setiap variabel. Hasil pengujian dari uji permutasi Monte Carlo parsial ditunjukkan pada TABEL 5.

TABEL 5. Hasil Uji Permutasi Monte Carlo Parsial

Variabel	Df	Nilai Eigen	$F_{pseudo}$	P-Value
$X_1$	1	0.0226	8.358	0.006
$X_2$	1	0.0102	3.763	0.027
$X_3$	1	0.0023	0.862	0.406
$X_4$	1	0.0156	5.776	0.008
$X_5$	1	0.0196	7.262	0.007
$X_6$	1	0.0086	3.164	0.040
$X_7$	1	0.0110	4.089	0.014
$X_8$	1	0.1185	4.376	0.017
$X_9$	1	0.0030	1.113	0.325
$X_{10}$	1	0.0148	5.488	0.007
$X_{11}$	1	0.0051	1.883	0.154
$X_{12}$	1	0.0048	1.783	0.138
$X_{13}$	1	0.0075	2.782	0.057
$X_{14}$	1	0.0075	2.787	0.070
$X_{15}$	1	0.0073	2.689	0.067
$X_{16}$	1	0.0014	0.532	0.646
$X_{17}$	1	0.0022	0.824	0.444
Residual	16	0.0433		

Berdasarkan TABEL 5, dapat dilihat nilai *P-value* yang dihasilkan untuk setiap variabel independen. Berdasarkan permutasi sebanyak 999 kali untuk setiap variabelnya, variabel  $X_1, X_2, X_4, X_5, X_6, X_7, X_8,$  dan  $X_{10}$  menunjukkan nilai *P-value* yang lebih kecil dari taraf signifikansi  $\alpha = 0.05$ . Hal ini menunjukkan bahwa variabel-variabel tersebut memiliki hubungan yang signifikan dengan data penyebaran jenis tindak kriminal. Sementara variabel  $X_3, X_9, X_{11}, X_{12}, X_{13}, X_{14}, X_{15}, X_{16},$  dan  $X_{17}$  akan dihilangkan dari penelitian, karena tidak memiliki hubungan yang signifikan dengan penyebaran jenis tindak kriminal.

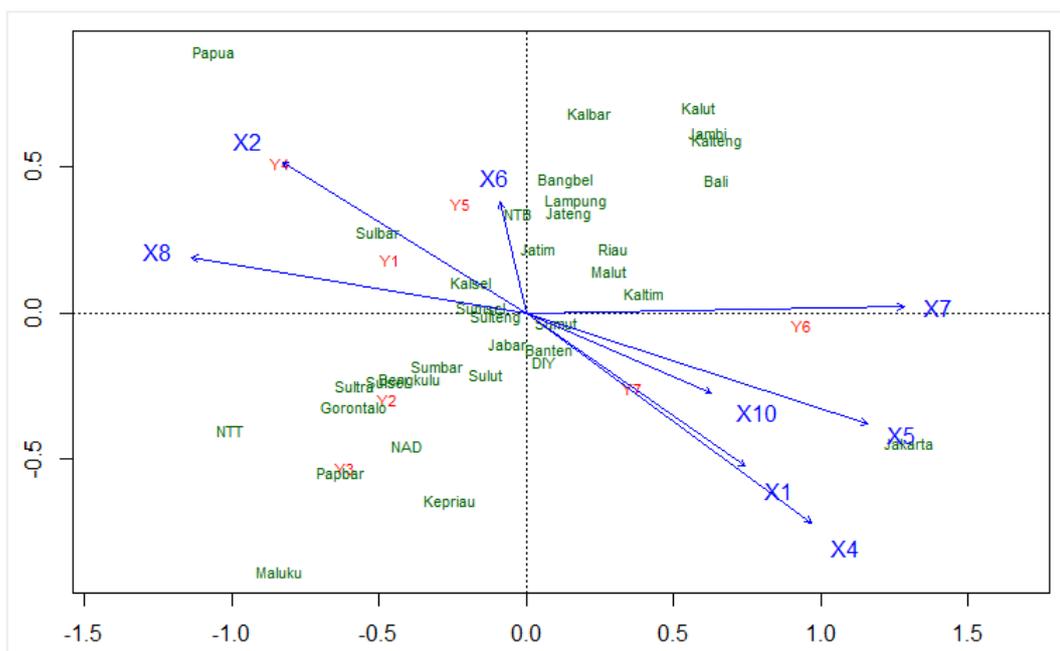
Setelah melakukan seleksi variabel, analisis dilakukan kembali menggunakan CCA dengan menggunakan variabel yang signifikan. Variabel independen yang digunakan diantaranya adalah  $X_1, X_2, X_4, X_5, X_6, X_7, X_8,$  dan  $X_{10}$ . Hasil perhitungan nilai eigen setiap dimensi ditunjukkan pada TABEL 6.

TABEL 6. Nilai Eigen Peta Triplot Akhir

Dimensi	Nilai Eigen	Proporsi	Proporsi Kumulatif
<i>Constrained</i>	0.1064	53.42%	53.43%
<i>Unconstrained</i>	0.0928	46.57%	100%
Total	0.1992	100%	
Dimensi	Nilai Eigen	Proporsi	Proporsi Kumulatif
Dimensi 1	0.0799	75.09%	75.09%
Dimensi 2	0.0165	15.51%	90.60%
Dimensi 3	0.0057	5.36%	95.96%
Dimensi 4	0.0027	2.54%	98.50%
Dimensi 5	0.0012	1.13%	99.72%
Dimensi 6	0.0003	0.28%	100%
Total <i>Constrained</i>	0.1064	100%	

TABEL 6 menunjukkan bahwa jumlah variansi data yang dapat dijelaskan oleh peta triplot (*constrained*) adalah sebanyak 0.1064 atau sebesar 53.42% dari variansi keseluruhan, sementara 46.57% sisanya tidak dapat ditangkap oleh peta triplot (*unconstrained*). Terdapat perbedaan yang cukup signifikan jika dibandingkan dengan peta yang dibuat sebelumnya, hal ini dikarenakan dari berkurangnya variabel independen sehingga jumlah informasi yang dapat ditangkap juga semakin berkurang. Kemudian banyaknya dimensi maksimal yang dapat dibuat adalah sebanyak enam dimensi. Dimensi pertama memiliki nilai eigen sebesar 0.0799 dan dapat menjelaskan variansi data sebanyak 75.09%. Dimensi kedua memiliki nilai eigen sebesar 0.0165 atau dapat menjelaskan variansi data sebanyak 15.51%. Secara akumulatif, dimensi pertama dan dimensi kedua dapat menjelaskan sebagian besar informasi yang ditangkap oleh peta triplot atau sebesar 90.60%, maka peta yang digunakan adalah peta dua dimensi. Hasil pemetaan menggunakan CCA ditunjukkan pada GAMBAR 2.

Peta Triplot Akhir



GAMBAR 2. Peta Triplot Akhir

Dari GAMBAR 2, dapat dilihat pemetaan jenis-jenis tindak kriminal dan provinsi di Indonesia. Berdasarkan pengamatan secara visual, dapat dilihat kecenderungan jenis tindak kriminal yang terjadi di suatu provinsi. Sebagai contoh, posisi variabel  $Y_4$  atau kasus kejahatan terhadap kemerdekaan orang berdekatan dengan provinsi Papua. Hal ini menunjukkan bahwa provinsi Papua cenderung sering terjadi kasus kejahatan terhadap kemerdekaan orang. Kemudian, dari GAMBAR 2 dapat dilihat kecenderungan karakteristik suatu provinsi berdasarkan posisinya dengan sumbu variabel independen. Contohnya, variabel  $X_1$ ,  $X_2$ , dan  $X_4$  berturut-turut menunjukkan angka melek huruf, angka putus sekolah SMP, dan rata-rata lama sekolah penduduk di setiap provinsi. Ketiga variabel tersebut merupakan indikator kualitas pendidikan di suatu wilayah. Pada GAMBAR 2, terlihat bahwa garis  $X_1$  dan  $X_4$  cenderung memiliki arah yang sama namun berlawanan dengan  $X_2$ . Hal ini menunjukkan hubungan antara variabel yaitu angka melek huruf berbanding lurus dengan rata-rata lama sekolah, tetapi keduanya berbanding terbalik dengan angka putus sekolah. Berdasarkan posisi provinsi-provinsi terhadap garis, provinsi yang memiliki tingkat pendidikan yang kurang diantaranya adalah Papua, Sulawesi Barat, Nusa Tenggara Timur, dan sebagainya, sedangkan provinsi dengan kualitas pendidikan yang baik adalah DKI Jakarta.

Berdasarkan GAMBAR 2, dapat dilihat pula hubungan antara jenis tindak kriminal dengan setiap karakteristik wilayah. Jika suatu garis panah diperpanjang sehingga menjadi sebuah sumbu, maka posisi setiap jenis tindak kriminal menunjukkan sebaran variabel dependen di sepanjang sumbu variabel independen tersebut. Sebagai contoh, jika pada GAMBAR 2 dibuat sumbu perpanjangan untuk variabel  $X_7$ , kemudian posisi setiap variabel dependen diproyeksikan terhadap sumbu tersebut, maka dapat diketahui hubungan antara variabel  $X_7$  dengan setiap jenis tindak kriminal. Karena variabel  $X_7$  menunjukkan persentase penduduk daerah perkotaan, dapat terlihat bahwa jenis kejahatan yang cenderung terjadi pada daerah dengan persentase penduduk perkotaan yang tinggi adalah kejahatan terkait narkoba ( $Y_6$ ). Sementara pada daerah dengan persentase penduduk perkotaan yang rendah, cenderung terjadi jenis kejahatan terhadap kemerdekaan orang ( $Y_4$ ).

## KESIMPULAN

Berdasarkan analisis dan uraian pada pembahasan sebelumnya, dapat disimpulkan bahwa dengan menggunakan *Canonical Correspondence Analysis* (CCA), jenis-jenis tindak kriminal di Indonesia berdasarkan karakteristik wilayah dapat dipetakan dalam satu peta triplot. Hasil pemetaan menunjukkan bahwa terdapat pola atau kemiripan tertentu dalam penyebaran jenis tindak kriminal berdasarkan karakteristik wilayah. Sebagai contoh, DKI Jakarta merupakan provinsi dengan tingkat pendidikan terbaik, sementara Papua menempati posisi paling rendah diantara provinsi lainnya. Jenis tindak kriminal yang cenderung terjadi di Jakarta adalah kejahatan terkait narkoba dan korupsi, sementara pada provinsi Papua cenderung terjadi kejahatan terhadap kemerdekaan orang. Hal ini menunjukkan bahwa kejahatan terkait narkoba dan korupsi cenderung terjadi di wilayah dengan pendidikan baik sementara kejahatan terhadap kemerdekaan orang terjadi pada wilayah yang berpendidikan kurang.

Total inersia atau ukuran tersebarnya data pada peta triplot adalah sebesar 0.1992. Variansi data yang dapat ditangkap oleh peta triplot adalah sebesar 0.1064, atau 53.4% dari total inersia. Secara akumulatif, 90.6% variansi data yang ditangkap oleh CCA dapat dijelaskan oleh peta triplot berdimensi dua. Hasil pengujian menggunakan uji permutasi Monte Carlo menunjukkan bahwa terdapat hubungan yang signifikan antara karakteristik wilayah dengan penyebaran jenis tindak kriminal. Beberapa variabel yang memiliki hubungan dengan penyebaran jenis tindak kriminal diantaranya adalah angka melek huruf, angka putus sekolah SMP, rata-rata lama sekolah, kepadatan penduduk, persentase penduduk miskin, rata-rata upah bersih buruh, rasio jenis kelamin, dan persentase penduduk daerah perkotaan.

## REFERENSI

- Badan Pusat Statistik, 2014, *Persentase Penduduk Daerah Perkotaan menurut Provinsi*, diakses 7 Maret 2021, <https://www.bps.go.id/statictable/2014/02/18/1276/persentase-penduduk-daerah-perkotaan-hasil-proyeksi-penduduk-menurut-provinsi-2015---2035.html>.
- Badan Pusat Statistik, 2019a, *Kepadatan Penduduk menurut Provinsi (jiwa/km<sup>2</sup>) 2015-2019*, diakses 3 Maret 2021, <https://www.bps.go.id/indicator/12/141/1/kepadatan-penduduk-menurut-provinsi.html>.
- Badan Pusat Statistik, 2019b, *Potret Pendidikan Indonesia, Statistik Pendidikan Indonesia 2019*, diakses 3 Maret 2021, <https://www.bps.go.id/publication/2019/11/29/1deb588ef5fdbfba3343bb51/potret-pendidikan-statistik-pendidikan-indonesia-2019.html>.
- Badan Pusat Statistik, 2020a, *Statistik Indonesia 2020*, diakses 1 Maret 2021, <https://www.bps.go.id/publication/2020/04/29/e9011b3155d45d70823c141f/statistik-indonesia-2020.html>.
- Badan Pusat Statistik, 2020b, *Statistik Kriminal 2020*, diakses 1 Maret 2021, <https://www.bps.go.id/publication/2020/11/17/0f2dfc46761281f68f11afb1/statistik-kriminal-2020.html>.
- Badan Pusat Statistik, 2020c, *Rata-Rata Lama Sekolah Penduduk Umur  $\geq 15$  Tahun Menurut Provinsi 2018-2020*, diakses 1 Maret 2021, <https://www.bps.go.id/indicator/28/1429/1/rata-rata-lama-sekolah-penduduk-umur-15-tahun-menurutprovinsi.html>.
- Badan Pusat Statistik, 2020d, *Indeks Pembangunan Manusia 2019-2020*, diakses 7 Maret 2021, <https://www.bps.go.id/indicator/26/413/1/-metode-baru-indeks-pembangunan-manusia.html>.
- Badan Pusat Statistik, 2020e, *Rata-rata Upah/Gaji Bersih Sebulan Buruh/Karyawan/Pegawai Menurut Provinsi dan Jenis Pekerjaan Utama, 2019*, diakses 7 Maret 2021, <https://www.bps.go.id/statictable/2019/08/02/2070/rata-rata-upahgaji-bersih-sebulan-buruh-karyaw-an-pegawai-menurut-provinsi-dan-jenis-pekerjaan-utama-2019.html>.
- Badan Pusat Statistik, 2021a, *Gini Rasio 2018-2019*, diakses 10 Maret 2021, <https://www.bps.go.id/indicator/23/98/2/gini-rasio.html>.

- Badan Pusat Statistik, 2021b, *Tingkat Pengangguran Terbuka Menurut Provinsi (Persen), 2018-2019*, diakses 4 Maret 2021 <https://www.bps.go.id/indicator/6/543/2/tingkat-pengangguran-terbuka-menurut-provinsi.html>.
- Badan Pusat Statistik, 2021c, *Persentase Penduduk Miskin Menurut Provinsi (Persen), 2018-2019*, diakses 8 Maret 2021, <https://www.bps.go.id/indicator/23/192/2/persentase-penduduk-miskin-menurut-provinsi.html>.
- Braak, T 1986, 'Canonical Correspondence Analysis: A New Eigenvector Technique for Multivariate Direct Gradient Analysis', *Ecology*, vol. 67, no. 5, hh. 1167-1179.
- Hardianto, F 2009, 'Analisis Faktor-Faktor yang Mempengaruhi Tingkat Kriminalitas di Indonesia dari Pendekatan Ekonomi', *Jurnal Ilmiah Fakultas Ekonomi Universitas Katolik Parahyangan*, vol. 13, no. 2, hh. 28-41.
- Legendre, P, Legendre, L 1998, *Numerical Ecology*, 2nd edition, Elsevier Science B V, Amsterdam.
- Leps, J, Smilauer, P 2003, *Multivariate Analysis of Ecological Data using CANOCO*, Cambridge University Press, New York.
- Nurfitasari, A, 2017, 'Pemetaan Tindak Kriminal di Wilayah Madiun dengan Analisis Korespondensi Berganda', *Media Statistika*, vol. 10, no. 2, hh. 131-143.
- Palmer, M 1993, "Putting things in even better order: The advantages of Canonical Correspondence Analysis", *Ecology*, vol. 74, no.8, hh. 2215-2230.
- Purwanti, E 2019, 'Analisis Faktor Ekonomi yang Mempengaruhi Kriminalitas di Jawa Timur', *Jurnal Ekonomi-Qu*, vol. 9, no. 2, hh. 154-177.