

Pendekatan Analisis Diskriminan dalam Pengklasifikasian Indeks Prestasi Manusia (IPM) di Provinsi Jawa Timur

Eva Khoirun Nisa*

¹ Program Studi Matematika, UIN Walisongo Semarang, Indonesia.

* *Corresponding author's e-mail:* evakn@walisongo.ac.id

Abstrak

Kemajuan suatu daerah seringkali diukur dengan banyaknya pembangunan fasilitas dan angka kemiskinan di daerah tersebut. Padahal dapat kita tentukan dari satu indikator kependudukan yaitu Indeks Pembangunan Manusia (IPM). IPM selalu dipantau tiap tahun karena menjadi tolok ukur keberhasilan suatu daerah. Termasuk Provinsi Jawa Timur yang selalu memantau IPM sehingga dari tahun 2022 sampai tahun 2024 ini IPM mengalami kenaikan bahkan percepatan. Semua aspek kehidupan diantaranya kesehatan dan hidup layak mengalami kenaikan. Namun, untuk IPM berdasarkan kota/kabupaten masih menyisakan IPM yang rendah. Penelitian ini bertujuan mengklasifikasikan IPM tiap kota/kabupaten di Provinsi Jawa Timur pada tahun 2023 berdasarkan faktor-faktor yang mempengaruhinya menggunakan analisis diskriminan. Analisis diskriminan merupakan statistika multivariat yang digunakan untuk mengklasifikasikan objek berdasarkan beberapa variabel yang mempengaruhinya. Hasil riset menyatakan Kota/Kabupaten Situbondo, Probolinggo, Sampang dan Sumenep termasuk kedalam kabupaten/kota dengan status IPM menengah bawah .

Kata Kunci: IPM; analisis diskriminan; statistika multivariat

Abstract

The progress of an area is often measured by the number of facilities built and the poverty rate in that area. In fact, we can determine it from one population indicator, namely the Human Development Index (HDI). HDI is always monitored every year because it is a benchmark for the success of a region. Including East Java Province which always monitors the HDI so that from 2022 to 2024 the HDI will increase and even accelerate. All aspects of life, including health and decent living, have increased. However, the HDI based on city/district still has a low HDI. This research aims to classify the HDI for each city/district in East Java Province in 2023 based on the factors that influence it using discriminant analysis. Discriminant analysis is multivariate statistics used to classify objects based on several variables that influence them. The research results state that the cities/regencies of Situbondo, Probolinggo, Sampang and Sumenep are included in the regencies/cities with lower middle HDI status.

Keywords: HDI; discriminant analysis; multivariate statistics

Pendahuluan

Jawa Timur merupakan salah satu provinsi besar di Indonesia. Kemajuan Jawa Timur tentunya selalu menjadi sorotan sehingga IPM yang menjadi salah satu tolok ukur keberhasilan suatu daerah perlu selalu ditingkatkan. Indeks Pembangunan Manusia (IPM) merupakan indikator penting yang menggambarkan pencapaian pembangunan manusia dalam tiga dimensi utama: kesehatan, pendidikan, dan standar hidup layak (BPS Provinsi Jawa Timur, 2023). Tiga tahun terakhir, IPM Jawa Timur menunjukkan tren

peningkatan yang konsisten dalam beberapa tahun terakhir. Pada tahun 2024, Badan Pusat Statistik (BPS) Jawa Timur melaporkan bahwa IPM provinsi ini mencapai 75,35, meningkat 0,938 poin persen dari 74,65 pada tahun 2023. Capaian ini menempatkan Jawa Timur dalam kategori "tinggi" (rentang 70-80) dan melampaui rata-rata nasional yang sebesar 75,02 (Biro Administrasi Pimpinan Jawa Timur, 2024). Peningkatan IPM 2023 terjadi pada semua dimensi, baik pada aspek umur panjang, hidup sehat, pengetahuan, dan standar hidup layak (BPS Provinsi Jawa Timur, 2023).

Menurut Rindiyan & Mubaraq (2023), dua faktor yang berpengaruh signifikan terhadap IPM adalah tingkat kesehatan dan pendidikan. Oleh karena itu, untuk dapat memastikan setiap kota/kabupaten di Jawa Timur berada pada status IPM yang aman (menengah ke atas) atau menengah ke bawah maka diperlukan analisis statistika yang tepat. Salah satu metode statistika yang dapat mengklasifikasikan data yang berbentuk kategorik dikotomi adalah analisis diskriminan.

Menurut Johnson & Winchern (2007) analisis diskriminan merupakan teknik multivariat yang berkaitan dengan pemisahan sekumpulan objek yang berbeda dengan mengalokasikan objek baru kedalam kelompok yang sudah diklasifikasikan. Karakteristik analisis diskriminan yaitu mengklasifikasikan sekumpulan data ke dalam kelompok berdasarkan sejumlah variabel yang penting dalam membedakan antar kelompok tersebut (Greenacre, 2010). Dengan kata lain tujuan dari analisis diskriminan ini adalah mencari cara terbaik untuk mengalokasikan objek berdasarkan kelompok klasifikasinya dengan membentuk fungsi diskriminan. Dalam analisis diskriminan, variabel respon berupa data kualitatif dan variabel penjelas/prediktor berupa data kuantitatif. Harapannya setelah diperoleh klasifikasi maka dapat dengan mudah mengidentifikasi kota/kabupaten yang perlu didampingi dalam menaikkan IPM.

Metode

Jenis penelitian ini merupakan penelitian kuantitatif dengan teknik sampling yaitu *purposive sampling* karena membutuhkan semua kota/kabupaten yang ada di Provinsi Jawa Timur dalam analisis. Salah satu kelebihan *purposive sampling* adalah dapat diperoleh informasi yang tepat sesuai tujuan penelitian (Sugiyono, 2017). Data yang digunakan adalah Indeks Pembangunan Manusia (IPM) pada Kota/Kabupaten di Propinsi Jawa Timur pada tahun 2023 yang dikelompokkan ke dalam dua kelompok. Kelompok pertama dengan kode 0 adalah kota yang status pembangunan manusianya menengah bawah ($50 \leq \text{IPM} < 66$) dan kelompok kedua dengan kode 1 adalah kota yang status pembangunan manusianya menengah atas ($66 \leq \text{IPM} < 80$). Untuk variabel sebagai dasar klasifikasi adalah angka harapan hidup (X_1), angka melek huruf (X_2), rata-rata lama sekolah (X_3), pengeluaran per kapita (X_4).

Untuk metode statistika yang digunakan adalah analisis diskriminan. Dalam analisis diskriminan diasumsikan bahwa setiap populasi yang terbentuk dicirikan oleh sebaran peluang bersama dari p variabel acak yang digunakan dalam pengukuran misal π_{ij} ($i = 1, 2, 3, \dots, k$) adalah populasi yang terbentuk $X' = (X_1, X_2, \dots, X_p)$ adalah vektor variabel

acak yang diukur, $f_i(x)$ adalah sebaran peluang variabel X untuk populasi. Vektor X didefinisikan sebagai suatu ruang sampel misalnya R_1, R_2, \dots, R_k maka akan dipersoalkan untuk menentukan batas ruang sampel tersebut sehingga suatu obyek pengamatan X_0 berasal dari R_i berarti obyek tersebut akan dimasukkan pada populasi π_i apabila :

P_i = peluang suatu obyek terpilih dalam π_i

$P_{i|j}$ = peluang bersyarat obyek ditempatkan di π_i , jika anggota obyek tersebut anggota π_j

P_{ij} = peluang bersama suatu obyek dari π_j ditempatkan di π_i

Dalam ruang sampel R_1, R_2, \dots, R_k seluruh kesalahan pengelompokan (misclassification) yang terjadi adalah penjumlahan dari P_{ij} pada $i = 1, 2, \dots, k, i \neq j$ sehingga peluang bahwa suatu obyek yang berasal dari π_j dan salah satu kelompok adalah

$$E_j = \sum_{i=1}^k P_{ij} = P_j \sum_{i=1}^k P_{ij} = P_j(1 - P_{jj}) \text{ untuk setiap } i \neq j, i = 1, 2, \dots, k$$

Bila $f_j(x)$ adalah sebaran peluang variabel X untuk populasi π_j dan peluang bahwa suatu obyek terlanjur dimasukkan ke dalam ruang sampel R_i adalah daerah di bawah R_i pada

$$P_{i,j} = \int_{R_i} f_j(x) dx$$

$$E_j = P_j \sum_{i=1, i \neq j}^K \int_{R_i} f_j(x) dx = P_j \left[1 - \int_{R_j} f_j(x) dx \right]$$

Peluang kesalahan pengelompokan pada k populasi (E) adalah penjumlahan dari E_j , sehingga dapat diperoleh :

$$E = \sum_{j=1}^K E_j = \sum_{j=1}^K P_j \left[1 - \int_{R_j} f_j(x) dx \right] \text{ dimana } \sum_{j=1}^K P_j = 1$$

$$E = 1 - \sum_{j=1}^K P_j \int_{R_j} f_j(x) dx$$

Permasalahan pengelompokan adalah memilih R_1, R_2, \dots, R_k sehingga dapat diminimumkan kesalahan E . Hal ini berarti pula memaksimumkan peluang untuk alokasi yang benar adalah

$$1 - E = \sum_{j=1}^K P_j \int_{R_j} f_j(x) dx$$

Jika $K=2$ maka menjadi

$$1 - E = P_1 \int_{R_1} f_1(x) dx + P_2 \int_{R_2} f_2(x) dx$$

dimana : $P_1 + P_2 = 1$

R_1 dan R_2 adalah ruang sampel dari X

$R_2 = R_1^*$ (komplemen dari R_1)

Jika $P_1, P_2, f_1(x)$ dan $f_2(x)$ diketahui, maka R_1 dapat ditentukan sedemikian hingga :

$$P_1 \int_{R_1} f_1(x) dx + (1 - P_2) \int_{R_2} f_2(x) dx \text{ maksimum}$$

Sehingga R1 dan R2 dapat diturunkan menjadi :

$$R_1 = \left[X \mid \frac{f_1(x)}{f_2(x)} \geq \frac{P_1}{P_2} \right] \text{ dan } R_2 = \left[X \mid \frac{f_2(x)}{f_1(x)} \geq \frac{P_2}{P_1} \right]$$

Fungsi inilah yang disebut fungsi diskriminan untuk kasus $f_1(x)$ dan $f_2(x)$ normal variabel ganda dengan rata-rata μ_1 dan μ_2 diketahui dan matrik kovarian Σ_1 dan Σ_2 diketahui, maka :

$$R_1 = \left[X \mid X^T \Sigma_2^{-1} (X - 2\mu_2) - X^T \Sigma_1^{-1} (X - 2\mu_1) \geq C \right]$$

$$R_2 = \left[X \mid X^T \Sigma_2^{-1} (X - 2\mu_2) - X^T \Sigma_1^{-1} (X - 2\mu_1) < C \right]$$

$$C = \mu_1^T \Sigma_1^{-1} \mu_1 - \mu_2^T \Sigma_2^{-1} (X - 2\mu_1) + 2 \ln \left[\frac{P_2 |\Sigma_1|^{1/2}}{P_1 |\Sigma_2|^{1/2}} \right]$$

Fungsi diskriminannya adalah :

$$D = X^T \Sigma_2^{-1} (X - 2\mu_2) - X^T \Sigma_1^{-1} (X - 2\mu_1)$$

Untuk kasus $\Sigma_1 = \Sigma_2$, $f_1(x)$ dan $f_2(x)$ berbeda hanya pada μ_1 dan μ_2 maka

$$R_1 = \left[x \mid x^T \Sigma^{-1} (\mu_1 - \mu_2) \geq C \right]$$

$$R_2 = \left[x \mid x^T \Sigma^{-1} (\mu_1 - \mu_2) < C \right]$$

dimana :

$$C = \ln \left[\left(\frac{P_2}{P_1} \right) \cdot \frac{(\mu_2^T \Sigma^{-1} \mu_2 - \mu_1^T \Sigma^{-1} \mu_1)}{2} \right]$$

Fungsi diskriminannya adalah :

$$D = X^T \Sigma^{-1} (\mu_1 - \mu_2)$$

Dalam bentuk matematis fungsi dapat ditulis :

$$D_i = \beta_{x1} X_1 + \beta_{x2} X_2 + \dots + \beta_{ip} X_p$$

di mana :

Di = nilai diskriminan ke-i, $i = 1, 2, \dots, k-1$

k = jumlah kelompok

Nilai dugaan dari $\beta_1, \beta_2, \dots, \beta_r$ yaitu b_1, b_2, \dots, b_p adalah

$$B = S^{-1}d$$

di mana :

$$b' = (b_1, b_2, b_3, \dots, b_p)$$

S^{-1} = matrik kebalikan (inves) dari kovarian gabungan (Spooled)

d = vektor perbedaan rata-rata dari kelompok yang di bandingkan.

Berdasarkan dari hasil pembentukan fungsi diskriminan, maka dapat digunakan untuk mengelompokkan objek baru ke dalam kelompok tertentu (Hair et al., 2010). Jika

ℓ_1 dan ℓ_2 adalah nilai diskriminan untuk rata-rata kelompok ke-1 dan ke-2 maka nilai diskriminan batas untuk dua kelompok tersebut yaitu $(\ell_1 + \ell_2)/2$ dan jika $\ell_1 > \ell_2$ maka

Obyek baru akan masuk ke kelompok ke 1 jika $D \geq \frac{1}{2}(L_1 + L_2)$

Obyek baru akan masuk ke kelompok ke 2 jika $D < \frac{1}{2}(L_1 + L_2)$

Untuk perubah yang distandarkan, fungsi diskriminan dapat diturunkan sebagai berikut :

$$Y_i = d_{i1}X_1 + d_{i2}X_2 + \dots + d_{ip}X_p$$

dimana :

Y_i = nilai diskriminan ke-i

d_i = koefisien pembobot fungsi diskriminan ke-i

X = variabel yang terstandarisasi dari p variabel pembeda

Untuk mendapatkan vektor d_i , dapat diperoleh dari persamaan $(\omega^{-1}\beta - \lambda) = 0$ dan $(\omega^{-1}\beta - \lambda)d = 0$ dengan syarat $d' S_{gab} d = 1$

di mana :

ω = matrik jumlah kuadrat dalam kelompok

β = matrik jumlah kuadrat antar kelompok

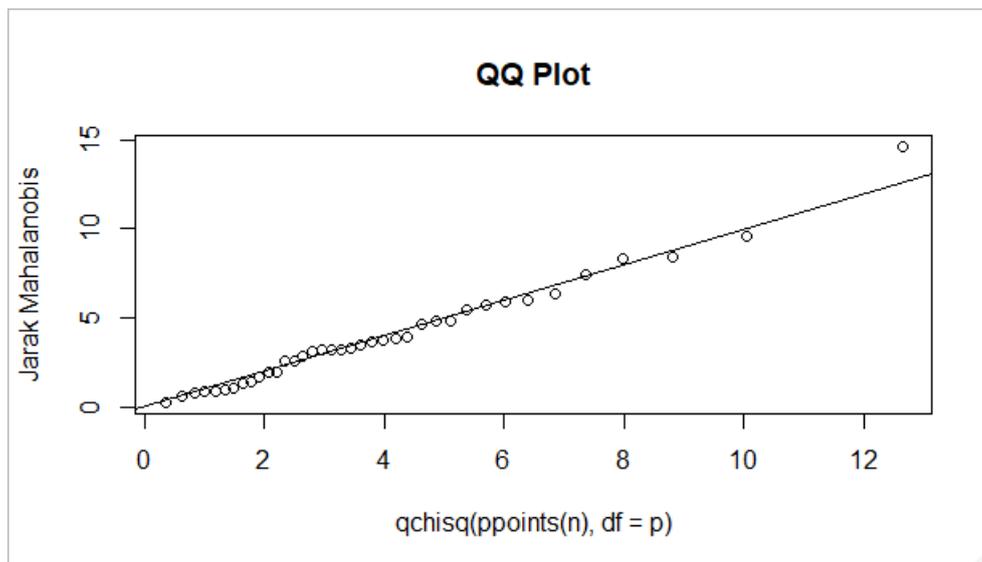
λ = akar karakteristik yang memenuhi persamaan diatas

S_{gab} = matrik kovarian sampel gabungan

Analisis ini berusaha mengeluarkan variabel yang kurang berguna dalam pembentukan fungsi diskriminan sebelum dilakukan analisis selanjutnya (Sharma, 2017). Kriteria seleksi variabel yang akan masuk dalam fungsi diskriminan dalam memilih variabel pembeda yang paling berarti yaitu mempunyai nilai F besar atau nilai Wilk's Lambda yang terkecil. Selanjutnya variabel pembeda kedua dipilih yang paling berarti berikutnya, dimana variabel tersebut mampu untuk meningkatkan kriteria perbedaan setelah dikombinasikan dengan variabel terpilih pertama. Demikian seterusnya, sampai tidak ada lagi variabel yang mampu untuk meningkatkan perbedaan setelah dikombinasikan dengan variabel terpilih sebelumnya atau pada setiap langkah, variabel terpilih dapat dikeluarkan dari analisis, jika ternyata dengan masuknya variabel tersebut justru melemahkan daya perbedaan setelah dikombinasikan dengan variabel sebelumnya.

Hasil dan Pembahasan

Langkah pertama yang dilakukan adalah menguji asumsi normalitas multivariat. Pengujian asumsi normalitas multivariat dapat menggunakan QQ Plot.



Gambar 1. QQ Plot Data IPM

Berdasarkan Gambar 1 terlihat titik-titik data disekitar garis lurus. Hanya ada satu titik data yang jauh dari garis lurus sehingga dapat dikatakan data berdistribusi normal multivariat. Begitu juga nilai jarak mahalanobis d^2 dan $qchisquare$ diperoleh lebih dari 50% nilai $d^2 \leq qchisquare$ maka H_0 gagal ditolak. Kesimpulannya asumsi normalitas multivariat terpenuhi. Dilanjutkan uji asumsi homogenitas.

Hipotesis:

H_0 : varian homogen

H_1 : varian tidak homogen

Statistik Uji:

Dengan uji Box's M diperoleh p-value 1,573 sehingga p-value lebih dari α ($1,573 > 0,05$) maka H_0 gagal ditolak. Artinya asumsi homogenitas terpenuhi. Berikut output nilai VIF untuk pengujian kebebasan antar variabel.

x1	x2	x3	x4
2.644574	6.894303	6.826421	2.029693

Output di atas diperoleh nilai VIF antar variable prediktor kurang dari 10 maka tidak terjadi multikolinieritas. Selanjutnya melihat perbedaan antar kategori atau kelompok dengan Wilks Lambda. Berdasarkan output yang diperoleh, nilai p-value $< \alpha$ ($4.816e-05 < 0,05$) maka H_0 ditolak yang artinya terdapat perbedaan antar kategori IPM. Berikut output koefisien parameter model diskriminan:

Coefficients of linear discriminants:

LD1	
x1	0.14685076
x2	0.28293484
x3	-0.63369420
x4	-0.02734747

berdasarkan output di atas diperoleh estimasi fungsi diskriminan sebagai berikut:

$$\hat{y} = 0,147X_1 + 0,283X_2 - 0,634X_3 - 0,027X_4$$

Selanjutnya diperoleh output analisis diskriminan dengan bantuan software RStudio sebagai berikut.

```
$class
```

```
[1] 1 1 1 1 1 1 1 1 1 1 1 0 0 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 1 1 1 1 1 1 1
```

```
Levels: 0 1
```

Dari output di atas diperoleh alokasi objek baru. Dapat kita lihat data ke-12, 13, 27, dan 29 masuk kedalam kategori 0 sehingga Kab/Kota Situbondo, Probolinggo, Sampang dan Sumenep termasuk kedalam kabupaten/kota dengan status IPM menengah bawah. Kemudian untuk memastikan bahwa klasifikasi hasil analisis diskriminan sudah tepat maka diuji ketepatannya. Karena total persentase ketepatan klasifikasi sebesar 94,74% maka dikatakan bahwa tepat fungsi diskriminan di atas dalam mengklasifikasikan IPM berdasarkan Angka Harapan Hidup(X1), Angka Melek Huruf (X2), Rata-rata lama sekolah (X3), Pengeluaran per kapita (X4).

Simpulan

Analisis diskriminan yang diperoleh mengklasifikasikan bahwa Kab/Kota Situbondo, Probolinggo, Sampang dan Sumenep termasuk kedalam kabupaten/kota dengan status IPM menengah bawah sedangkan sisanya yaitu 34 kabupaten/kota lainnya adalah kota/kabupaten yang status pembangunan manusianya menengah atas.

Daftar Pustaka

- Greenacre, M. (2010). Discriminant Analysis Biplots. *Fundacion BBVA, September*. <http://www.multivariatestatistics.org/biplots.html>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate Data Analysis*. Prentice Hall. <https://doi.org/10.1016/j.foodchem.2017.03.133>
- Johnson, R. A., & Winchern, D. W. (2007). *Applied Multivariate Statistical Analysis*. Pearson Prentice Hall.
- Rindiyani, & Mubaraq, A. (2023). Analisis Pengaruh Indeks Pembangunan Manusia Terhadap Pertumbuhan Ekonomi di Provinsi Kalimantan Barat. *Prosiding Seminar Nasional Program Studi Ekonomi Islam, 1*.
- Sharma, S. K. (2017). *A Comparative Analysis of Rate of Convergence For Linear And Quadratic Approximations in N-R Method*. 5, 94–96.
- Sugiyono. (2017). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Alfabeta.