

Pemodelan Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur Dengan Menggunakan Metode Regresi Logistik Ridge

Dwi Maumere Putra dan Vita Ratnasari
Jurusan Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam,
Institut Teknologi Sepuluh Nopember (ITS)
Jl. Arief Rahman Hakim, Surabaya 60111 Indonesia
e-mail: vita_ratna@statistika.its.ac.id

Abstrak— Pembangunan manusia di suatu daerah merupakan upaya yang dilakukan oleh pemerintah daerah terkait dengan kesejahteraan masyarakat yang diukur dengan Indeks Pembangunan Manusia (IPM). Terdapat tiga sektor pembentuk IPM yaitu kesehatan, pendidikan dan ekonomi dimana faktor-faktor dalam setiap sektor cenderung memiliki kolinieritas yang tinggi yang menyebabkan adanya kasus multikolinieritas. Apabila kasus multikolinieritas tidak diatasi, maka dapat menyebabkan variansi dari hasil estimasi parameter menjadi besar yang dapat berakibat pada banyaknya variabel prediktor yang tidak signifikan meskipun nilai koefisien determinasi (R^2) tinggi. Sehingga untuk mengatasinya dilakukan pemodelan terhadap Indeks Pembangunan Manusia (IPM) Provinsi Jawa Timur menggunakan metode Regresi Logistik Ridge. Terdapat tiga variabel yang berpengaruh signifikan, yaitu angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5). Dengan metode backward elimination, didapatkan model terbaik dengan ketepatan klasifikasi sebesar 97,37% yang menghasilkan 5 kabupaten/kota tergolong IPM menengah bawah, yaitu Kabupaten Bangkalan, Sampang, Probolinggo, Situbondo dan Jember. Pada 33 kabupaten/kota yang lain tergolong dalam IPM menengah atas.

Kata Kunci— Indeks Pembangunan Manusia (IPM), Jawa Timur, Klasifikasi, Regresi Logistik Ridge.

I. PENDAHULUAN

JAWA TIMUR merupakan salah satu provinsi besar di Indonesia dengan jumlah penduduk mencapai 37 juta jiwa lebih pada tahun 2010 dan pada tahun 2013 meningkat menjadi 38 juta jiwa lebih [1]. Rujukan [2] pada artikel BKKbN yang menyatakan bahwa terdapat beberapa faktor yang mempengaruhi pertumbuhan ekonomi, dimana faktor-faktor tersebut terbagi menjadi dua, yaitu faktor ekonomi dan faktor nonekonomi. Sumber Daya Manusia (SDM) merupakan salah satu faktor ekonomi yang menentukan keberhasilan pembangunan nasional melalui jumlah dan kualitas penduduk. Namun, di lain sisi jika pertumbuhan penduduk tidak terkendali, maka dapat menjadi suatu beban yang berat dalam proses pembangunan manusia di suatu wilayah.

Pembangunan di suatu daerah merupakan suatu upaya yang dilakukan oleh pemerintah daerah terkait kesejahteraan masyarakat serta mewujudkan kemakmuran masyarakat. Salah satu pembangunan yang menjadi perhatian pemerintah daerah

adalah pembangunan manusia yang diukur menggunakan indikator yang disebut Indeks Pembangunan Manusia (IPM). IPM merupakan suatu indeks komposit yang digunakan untuk mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup manusia. IPM dibangun dari tiga dimensi dasar yang digunakan sebagai ukuran kualitas hidup manusia. Ketiga sektor pembentuk IPM adalah kesehatan, pendidikan dan ekonomi [3].

IPM Provinsi Jawa Timur pada tahun 2013 sebesar 73,54. BPS mengkategorikan nilai IPM Provinsi Jawa Timur tahun 2013 dalam kategori menengah atas, namun nilai tersebut masih berada di bawah rata-rata IPM Indonesia yang bernilai 73,81 dengan selisih sebesar 0,17. Pembangunan manusia pada setiap kabupaten/kota di Provinsi Jawa Timur masih belum merata dan terdapat kesenjangan yang cukup tinggi. Hal tersebut dapat dilihat dari nilai IPM Kota Surabaya sebesar 78,97 yang merupakan IPM tertinggi di Provinsi Jawa Timur, tetapi di sisi lain masih terdapat nilai IPM yang cukup jauh dari rata-rata IPM Jawa Timur, yaitu Kabupaten Sampang yang hanya sebesar 62,39 [4]. Sehingga perlu dilakukan penelitian untuk mengetahui faktor-faktor yang mempengaruhi nilai IPM untuk setiap kabupaten/kota di Provinsi Jawa Timur.

Faktor-faktor di setiap sektor pembentuk IPM akan cenderung memiliki hubungan yang kuat satu sama lainnya, karena antar faktor-faktor tersebut saling mempengaruhi. Hal tersebut dapat mengakibatkan adanya kasus multikolinieritas dimana terjadi korelasi yang tinggi di antara variabel prediktor atau dapat dikatakan antar variabel prediktor tidak bersifat saling bebas [5]. Apabila kasus multikolinieritas tidak diatasi, maka dapat menyebabkan variansi dari hasil estimasi parameter menjadi besar yang akan berakibat pada besarnya standar error serta interval kepercayaan juga akan menjadi lebar. Selain itu dapat berakibat sedikit variabel prediktor yang signifikan, meskipun nilai koefisien determinasinya (R^2) tinggi.

Berdasarkan adanya kasus multikolinieritas serta struktur data IPM Provinsi Jawa Timur yang telah dikategorikan, maka digunakan analisis Regresi Logistik Ridge dimana dilakukan pemodelan dengan menambahkan suatu bilangan positif kecil yang disebut *ridge parameter* pada estimasi parameter. Hasil penelitian ini diharapkan dapat memberikan informasi terkait

faktor-faktor yang secara signifikan mempengaruhi IPM pada masing-masing kabupaten/kota di Jawa Timur.

II. TINJAUAN PUSTAKA

A. Multikolinieritas

Multikolinieritas merupakan suatu kondisi dimana terjadi korelasi yang tinggi di antara variabel prediktor atau dapat dikatakan antar variabel prediktor tidak bersifat saling bebas. *Variance Inflation Factor (VIF)* dapat digunakan sebagai kriteria untuk mendeteksi kasus multikolinieritas pada regresi linier yang memiliki lebih dari dua variabel prediktor. Nilai *VIF* untuk parameter regresi ke-*j* diformulasikan pada persamaan (1) [5].

$$VIF_j = \frac{1}{1 - R_j^2} \quad (1)$$

Dengan R_j^2 merupakan koefisien determinasi antara X_j dengan variabel prediktor lainnya pada persamaan regresi, dimana $j = 1, 2, \dots, p$. Apabila nilai *VIF* > 10, maka dapat diindikasikan bahwa terdapat kasus multikolinieritas yang serius. Hal-hal yang akan terjadi apabila kasus multikolinieritas tidak diatasi adalah variansi estimasi menjadi besar, interval kepercayaan menjadi lebar, dikarenakan variansi dan standar error besar. Kemudian pengujian signifikansi secara parsial menjadi tidak signifikan. Serta koefisien determinasi (R^2) tinggi, tetapi sedikit variabel prediktor yang signifikan [5].

B. Regresi Logistik

Model Regresi Logistik digunakan untuk menggambarkan hubungan antara variabel respon dengan satu atau beberapa buah variabel prediktor. Merujuk [6], Variabel dependen dalam regresi logistik merupakan variabel kualitatif. Apabila kategori dalam variabel dependen berupa biner/dikotomis, maka variabel dependen (*Y*) mengikuti distribusi Binomial dengan parameter π_i , dimana untuk setiap pengamatan ke-*i* ditulis pada persamaan (2).

$$y_i \sim \text{Binomial}(1, \pi_i) \quad (2)$$

Dengan fungsi probabilitas $f(y_i) = (\pi_i(\mathbf{x}_i))^{y_i} (1 - \pi_i(\mathbf{x}_i))^{1 - y_i}$, dimana $y_i = 0, 1$. $\pi_i(\mathbf{x}_i)$ adalah probabilitas dari kejadian ke-*i*. Apabila $y_i = 1$, maka $f(y_i) = \pi_i(\mathbf{x}_i)$ dan apabila $y_i = 0$, maka nilai $f(y_i) = 1 - \pi_i(\mathbf{x}_i)$.

Bentuk persamaan Regresi Logistik ditunjukkan pada persamaan (3) [6].

$$\pi_i(\mathbf{x}_i) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})} \quad (3)$$

Dimana $\beta_0 =$ konstanta, $\beta_j =$ koefisien regresi dan $j =$ banyaknya variabel prediktor. Transformasi logit dari $\pi(\mathbf{x}_i)$ ditampilkan pada persamaan (4) [6].

$$\text{Logit}[\pi_i(\mathbf{x}_i)] = \mathbf{x}_i \boldsymbol{\beta} = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_j x_{ip} \quad (4)$$

Estimasi parameter dalam regresi logistik digunakan metode *Maximum Likelihood Estimator (MLE)* karena distribusi dari variabel respon telah diketahui. *MLE* didapatkan dengan cara memaksimalkan logaritma fungsi likelihood [6].

Didapatkan fungsi likelihood pada persamaan (5).

$$L(\mathbf{X}, \boldsymbol{\beta}) = \prod_{i=1}^n [\pi_i(\mathbf{x}_i)]^{y_i} [1 - \pi_i(\mathbf{x}_i)]^{1 - y_i}$$

$$\ln L(\mathbf{X}, \boldsymbol{\beta}) = \sum_{i=1}^n [y_i (\mathbf{x}_i^T \boldsymbol{\beta}) - \ln(1 + \exp(\mathbf{x}_i^T \boldsymbol{\beta}))] \quad (5)$$

Hasil estimasi parameter $\hat{\boldsymbol{\beta}}$ adalah pada persamaan (6) [7].

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{W} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{z} \quad (6)$$

Dimana \mathbf{z} merupakan vektor $n \times 1$ dengan :

$$z_i = \text{Logit}[\hat{\pi}_i(\mathbf{x}_i)] + \frac{y_i - \hat{\pi}_i(\mathbf{x}_i)}{\hat{\pi}_i(\mathbf{x}_i)[1 - \hat{\pi}_i(\mathbf{x}_i)]} \quad (7)$$

Matriks kovarian untuk $\hat{\boldsymbol{\beta}}$ ditampilkan pada persamaan (8) [7].

$$\text{Var}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}^T \text{diag}[\hat{\pi}_i(1 - \hat{\pi}_i)] \mathbf{X})^{-1} \quad (8)$$

C. Regresi Ridge

Regresi Ridge adalah salah satu metode yang dapat digunakan untuk mengatasi kasus multikolinieritas. Metode ini ditujukan untuk mengatasi kondisi buruk yang diakibatkan oleh korelasi yang tinggi antara beberapa variabel prediktor dalam model regresi, yang dapat menghasilkan hasil estimasi dari parameter model regresi menjadi tidak stabil [8]. Estimasi parameter regresi ridge menggunakan metode *least square (LS)* dengan menambahkan bilangan positif kecil θ pada diagonal matriks $\mathbf{X}^T \mathbf{X}$, sehingga bias yang terjadi dapat dikendalikan. Bilangan positif kecil θ bernilai antara 0 dan 1, sehingga estimasi ridge akan bias terhadap parameter β , tetapi cenderung lebih stabil [9].

Kebaikan estimator ridge diukur dengan menggunakan *Mean Square Error (MSE)* pada persamaan (9) [10].

$$MSE(\hat{\boldsymbol{\beta}}) = \text{Tr}[\text{Var}(\hat{\boldsymbol{\beta}})] + [\text{bias}(\hat{\boldsymbol{\beta}})]^T [\text{bias}(\hat{\boldsymbol{\beta}})] \quad (9)$$

Estimasi parameter Regresi Ridge didasarkan pada regresi linier dengan menambahkan dengan *ridge parameter* (θ) pada elemen diagonal matriks $\mathbf{X}^T \mathbf{X}$. Sehingga fungsi obyektif ditampilkan pada persamaan (10) [10].

$$\phi(\hat{\boldsymbol{\beta}}^*) = (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^*)^T (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}^*) + \theta \hat{\boldsymbol{\beta}}^{*T} \hat{\boldsymbol{\beta}}^* \quad (10)$$

Estimasi parameter untuk Regresi Ridge pada persamaan (11) [10].

$$\hat{\boldsymbol{\beta}}^* = (\mathbf{X}^T \mathbf{X} + \theta \mathbf{I})^{-1} \mathbf{X}^T \mathbf{y} \quad (11)$$

Nilai *MSE* regresi ridge adalah pada persamaan (12) [10].

$$MSE = \text{Tr}[\sigma^2 (\mathbf{X}^T \mathbf{X} + \theta \mathbf{I})^{-1}] = \sigma^2 \text{Tr}[(\mathbf{X}^T \mathbf{X} + \theta \mathbf{I})^{-1}]$$

$$= \sigma^2 \sum_{j=1}^p \frac{1}{(\lambda_j + \theta)} \quad (12)$$

D. Regresi Logistik Ridge

Fungsi obyektif untuk Regresi Logistik Ridge [9].

$$\phi(\hat{\boldsymbol{\beta}}^{\oplus}) = \sum_{i=1}^n y_i \ln[\pi_i(\mathbf{x}_i)] + \sum_{i=1}^n (1 - y_i) \ln[1 - \pi_i(\mathbf{x}_i)] - \theta \hat{\boldsymbol{\beta}}^{*\otimes T} \hat{\boldsymbol{\beta}}^{\oplus}$$

$$\phi(\hat{\boldsymbol{\beta}}^{\oplus}) = \sum_{i=1}^n [y_i (\mathbf{x}_i \hat{\boldsymbol{\beta}}^{\oplus}) - \ln(1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^{\oplus}))] - \theta \hat{\boldsymbol{\beta}}^{*\otimes T} \hat{\boldsymbol{\beta}}^{\oplus} \quad (13)$$

Dimana $\hat{\boldsymbol{\beta}}^{\oplus}$ merupakan koefisien parameter untuk Regresi Logistik Ridge. Sedangkan y_i merupakan respon berupa kategorik yang mengikuti distribusi Binomial ($1, \pi_i$) dan \mathbf{x}_i

merupakan vektor untuk setiap observasi yang diambil dari matriks variabel prediktor berukuran $n \times h+1$.

Selanjutnya diturunkan secara parsial terhadap $\hat{\beta}^{\oplus}$ [9]:

$$\frac{\partial \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus}} = \mathbf{X}^T (\mathbf{y} - \boldsymbol{\pi}) - 2\theta \hat{\beta}^{\oplus}$$

Kemudian dilanjutkan pada turunan kedua menjadi [9]:

$$\frac{\partial^2 \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus} \partial \hat{\beta}^{\oplus T}} = \frac{\partial}{\partial \hat{\beta}^{\oplus T}} \left[\frac{\partial \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus}} \right] = -\mathbf{X}^T \mathbf{W} \mathbf{X} - 2\theta \mathbf{I}$$

dimana : $\mathbf{W} = \text{diag}[\hat{\pi}_i(x_i)[1 - \hat{\pi}_i(x_i)]$.

Estimasi parameter Regresi Logistik Ridge menggunakan metode *MLE* dengan iterasi Newton-Raphson yang akan digunakan untuk memaksimalkan fungsi obyektif pada persamaan (13). Diekspansikan di sekitar β^{\oplus} menurut Deret Taylor dan didapatkan persamaan (14) [9].

$$\left. \frac{\partial \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus}} \right|_{\hat{\beta}^{\oplus} = \hat{\beta}_0^{\oplus}} = \left. \frac{\partial^2 \phi(\hat{\beta}^{\oplus})}{\partial \hat{\beta}^{\oplus} \partial \hat{\beta}^{\oplus T}} \right|_{\hat{\beta}^{\oplus} = \hat{\beta}_0^{\oplus}} (\hat{\beta}^{\oplus} - \hat{\beta}_0^{\oplus}) \quad (14)$$

Hasil penurunan disubstitusikan ke dalam persamaan (14) menghasilkan persamaan (15) [10].

$$\hat{\beta}^{\oplus} = (\mathbf{X}^T \mathbf{W} \mathbf{X} + 2\theta \mathbf{I})^{-1} \mathbf{X}^T \mathbf{W} \mathbf{z} \quad (15)$$

Dengan \mathbf{z} merupakan vektor berukuran $n \times 1$, dimana :

$$z_i = \text{Logit}[\hat{\pi}_i(x_i)] + \frac{y_i - \hat{\pi}_i(x_i)}{\hat{\pi}_i(x_i)[1 - \hat{\pi}_i(x_i)]}$$

Dengan menambahkan *ridge parameter* untuk Regresi Logistik Ridge pada elemen diagonal matriks kovarian Regresi Logistik, maka formula variansi pada persamaan (16) [10].

$$\text{Var}(\hat{\beta}^{\oplus}) = (\mathbf{X}^T \text{diag}[\hat{\pi}_i(1 - \hat{\pi}_i)] \mathbf{X} + \theta^{\oplus} \mathbf{I})^{-1} \quad (16)$$

E. Pengujian Signifikansi Parameter

Pengujian signifikansi parameter secara serentak dilakukan dengan menggunakan *Likelihood Ratio Test*, dimana hipotesis yang digunakan adalah :

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{Minimal ada satu } \beta_j \neq 0, j=1, 2, \dots, p$$

Statistik uji *Likelihood Ratio Test* ditampilkan pada persamaan (17) [6].

$$G = -2 \log \left(\frac{L_0}{L_1} \right) = -2[\ln(L_0) - \ln(L_1)] \quad (17)$$

Nilai $-2[\ln(L_0) - \ln(L_1)]$ mengikuti distribusi *chi-square* dengan $df = p$. Jika $-2[\ln(L_0) - \ln(L_1)] \geq \chi^2_{(p, \alpha)}$, maka diputuskan untuk menolak H_0 yang artinya model berpengaruh signifikan, sehingga uji signifikansi parameter dapat dilanjutkan dengan pengujian secara parsial untuk mengetahui variabel-variabel prediktor yang signifikan terhadap peluang sukses [6].

Pengujian signifikansi parameter secara parsial dilakukan dengan menggunakan *Wald test* [11]. Hipotesis yang digunakan adalah :

$$H_0 : \beta_j = 0$$

$$H_1 : \beta_j \neq 0, j=1, 2, \dots, p$$

Statistik uji *Wald test* ditampilkan pada persamaan (18).

$$t = \frac{\hat{\beta}_j}{SE(\hat{\beta}_j)} \quad (18)$$

Dimana : $SE(\hat{\beta}_j) = \sqrt{\text{Var}(\hat{\beta}_j)}$

Diputuskan untuk menolak H_0 apabila nilai $t > t_{(\alpha, p)}$ atau $p\text{-value} < \alpha$, yang artinya variabel ke- j berpengaruh signifikan terhadap pembentukan model [11].

F. Indeks Pembangunan Manusia (IPM)

Pembangunan manusia merupakan suatu proses untuk memperbanyak pilihan-pilihan yang dimiliki oleh manusia. Dimana pilihan-pilihan tersebut terdiri dari tiga komponen dasar, yaitu untuk berumur panjang dan sehat, untuk memiliki ilmu pengetahuan dan yang ketiga untuk mempunyai akses terhadap sumber daya yang dibutuhkan sehingga dapat menjalani kehidupan yang layak [12]. Oleh karena itu, Indeks Pembangunan Manusia (IPM) diartikan sebagai suatu indeks komposit yang digunakan untuk mengukur capaian pembangunan manusia berbasis sejumlah komponen dasar kualitas hidup manusia. Ketiga komponen dasar yang digunakan sebagai ukuran kualitas hidup tersebut diukur dengan menggunakan suatu indeks untuk masing-masing komponen, yaitu indeks harapan hidup, indeks pendidikan dan indeks standar hidup layak [2].

III. METODOLOGI PENELITIAN

A. Sumber Data

Data yang digunakan pada penelitian ini adalah data sekunder mengenai Indeks Pembangunan Manusia (IPM) beserta faktor-faktornya untuk setiap Kabupaten/Kota di Jawa Timur yang diambil di Badan Pusat Statistik (BPS) Jawa Timur. Unit observasi yang diteliti adalah 29 kabupaten dan 9 kota di Provinsi Jawa Timur.

B. Variabel Penelitian

Pada penelitian ini digunakan IPM Provinsi Jawa Timur yang telah dikategorikan ke dalam empat kategori, yaitu IPM rendah ($IPM \leq 50$), IPM menengah bawah ($50 < IPM \leq 66$), IPM menengah atas ($66 < IPM \leq 90$) dan IPM tinggi ($IPM > 90$), sedangkan variabel prediktor berdasarkan masing-masing sektor ditampilkan pada Tabel 1.

Tabel 1.

Variabel Penelitian Indikator IPM		
Sektor	Simbol	Variabel
-	Y	Indeks Pembangunan Manusia
Kesehatan	X ₁	Angka kematian bayi
	X ₂	Presentase keluhan kesehatan
	X ₃	Jumlah sarana kesehatan
Pendidikan	X ₄	Angka buta huruf (Usia 10 tahun ke atas)
	X ₅	Angka partisipasi sekolah (SMA)
	X ₆	Rasio guru-siswa (SMA)
	X ₇	Rasio sekolah-murid (SMA)
Ekonomi	X ₈	Presentase penduduk miskin
	X ₉	PDRB perkapita
	X ₁₀	Pertumbuhan ekonomi
	X ₁₁	Presentase penduduk usia 15 tahun ke atas yang bekerja
	X ₁₂	Tingkat pengangguran terbuka
	X ₁₃	Tingkat partisipasi angkatan kerja

C. Metode Analisis Data

Langkah-langkah dalam analisis data dalam penelitian ini adalah sebagai berikut.

1. Pengumpulan data IPM Provinsi Jawa Timur tahun 2012.
2. Analisis deskriptif pada IPM Provinsi Jawa Timur dan faktor-faktor di setiap sektor yang mempengaruhinya, yaitu kesehatan, pendidikan dan ekonomi.
3. Pemeriksaan kasus multikolinieritas dengan menggunakan nilai *Variance Inflation Factor (VIF)* untuk masing-masing variabel prediktor (X).
4. Pemilihan nilai *ridge parameter* (θ^{\oplus}):
 - a. Untuk $r = 1, \dots, p$ dapat dihitung nilai *ridge parameter* pada persamaan (20).

$$\theta_r^{\oplus} = \frac{r}{\hat{\alpha}_r^T \hat{\alpha}_r} \quad (19)$$

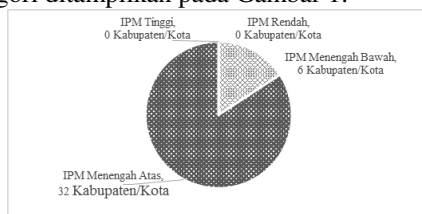
Dimana $\hat{\alpha}_r$ merupakan r koefisien pertama *Principal Component* Regresi Logistik (PCLR).

- b. Digunakan θ_r^{\oplus} yang meminimumkan perbedaan antara r dan derajat bebas dari model Regresi Logistik Ridge yang telah didapatkan.
5. Pemodelan menggunakan metode Regresi Logistik Ridge.
 - a. Pengujian signifikansi parameter Regresi Logistik Ridge secara serentak dan parsial.
 - b. Interpretasi model lengkap Regresi Logistik Ridge.
6. Pemilihan model Regresi Logistik Ridge terbaik dengan menggunakan metode *backward elimination*.
 - a. Pengujian signifikansi parameter Regresi Logistik Ridge secara serentak dan parsial.
 - b. Interpretasi model terbaik Regresi Logistik Ridge yang didapatkan dari hasil *backward elimination*.

IV. HASIL DAN PEMBAHASAN

A. Statistika Deskriptif Kabupaten/Kota di Provinsi Jawa Timur Berdasarkan Indeks Pembangunan Manusia (IPM)

Provinsi Jawa Timur terdiri dari 29 kabupaten dan 9 kota. IPM Jawa Timur tahun 2012 sebesar 72,83, tergolong kategori menengah atas. Jumlah kabupaten/kota yang tergolong dalam setiap kategori ditampilkan pada Gambar 1.



Gambar 1 Jumlah Kabupaten/Kota untuk Setiap Kategori IPM

Kabupaten/kota di Provinsi Jawa Timur yang memiliki IPM dengan kategori menengah atas dan menengah bawah masing-masing berjumlah 6 dan 32 kabupaten/kota. Sedangkan pada kategori yang lain, tidak terdapat kabupaten/kota yang tergolong dalam IPM rendah maupun tinggi.

Pada Tabel 2 ditampilkan deskripsi mengenai kabupaten/kota di Provinsi Jawa Timur.

Tabel 2

Deskripsi Kabupaten/Kota di Provinsi Jawa Timur Menurut setiap Sektor

Variabel	Y	Rataan	Varians	Min	Maks
Angka Kematian Bayi	0	56,29	13,16	53,93	63,51
	1	29,64	72,57	19,50	51,07

Keluhan Kesehatan (%)	0	30,00	41,69	17,82	37,18
	1	27,33	34,32	16,89	45,02

Lanjutan Tabel 2
Deskripsi Kabupaten/Kota di Provinsi Jawa Timur Menurut setiap Sektor

Variabel	Y	Rataan	Varians	Min	Maks
Jumlah Sarana Kesehatan	0	1641	602105	1050	3184
	1	1414	675816	186	3249
Angka Buta Huruf	0	18,630	9,650	15,35	24,01
	1	7,654	21,336	1,50	19,52
Angka Partisipasi Sekolah (SMA)	0	47,95	58,74	38,61	61,21
	1	67,80	86,15	47,84	82,07
Rasio Guru-Siswa (SMA)	0	863,2	26689,0	658	1047
	1	1011,7	176776,3	553	2245
Rasio Sekolah-Murid (SMA)	0	22,50	50,70	12	31
	1	41,16	1468,01	11	151
Rumah Tangga Miskin (%)	0	23,80	100,96	9,94	36,03
	1	13,85	102,24	4,30	52,96
PDRB Perkapita (juta)	0	13,42	11,89	8,69	17,70
	1	30,97	2528,76	8,32	290,79
Pertumbuhan Ekonomi (%)	0	66,120	0,1320	6,19	7,27
	1	69,881	0,2917	5,82	8,26
Penduduk Usia > 15 Tahun Yang Bekerja (%)	0	68,710	25,660	61,630	75,330
	1	66,651	25,014	58,332	78,801
Tingkat Pengangguran Terbuka (%)	0	3,342	1,739	1,78	5,32
	1	4,397	2,980	1,16	7,85
Tingkat Partisipasi Angkatan Kerja (%)	0	71,050	20,320	64,13	76,69
	1	69,678	17,951	62,53	79,73

Keterangan kategori IPM :

0 = IPM menengah bawah

1 = IPM menengah atas

Tabel 2 menunjukkan bahwa hampir semua variabel memiliki varians yang cukup besar kecuali pada variabel pertumbuhan ekonomi dan tingkat pengangguran terbuka. Hal tersebut menunjukkan bahwa pembangunan di Provinsi Jawa Timur dalam hal pertumbuhan ekonomi dan pengentasan jumlah pengangguran terbuka telah merata untuk setiap kabupaten/kota. Tetapi pada variabel-variabel lain menunjukkan bahwa masih terdapat perbedaan yang cukup jauh antar kabupaten/kota di Provinsi Jawa Timur, sehingga masih terjadi kesenjangan yang cukup tinggi di semua sektor pembentuk IPM baik sektor kesehatan, pendidikan maupun ekonomi.

Pada Tabel 3 ditampilkan nilai *Variance Inflation Factors (VIF)* pada masing-masing variabel prediktor terhadap variabel prediktor yang lain.

Tabel 3

Nilai *Variance Inflation Factors (VIF)*

Sektor	Variabel	VIF
Kesehatan	Angka Kematian Bayi	4,122
	Keluhan Kesehatan	1,436
	Jumlah Sarana Kesehatan	4,935
Pendidikan	Angka Buta Huruf	6,205
	Angka Partisipasi Sekolah (SMA)	2,591
	Rasio Guru-Siswa (SMA)	4,539
	Rasio Sekolah-Murid (SMA)	9,436
Ekonomi	Rumah Tangga Miskin	1,766
	PDRB Perkapita	1,841
	Pertumbuhan Ekonomi	2,652
	Penduduk Usia > 15 Tahun Yang Bekerja	5243,225
	Tingkat Pengangguran Terbuka	294,646
	Tingkat Partisipasi Angkatan Kerja	3530,604

Tabel 3 menunjukkan bahwa kasus multikolinieritas terjadi pada faktor-faktor di sektor pendidikan dan ekonomi. Pada

sektor pendidikan diindikasikan terdapat kasus multikolinieritas pada variabel angka buta huruf (X_4) dan rasio sekolah-murid (X_7) karena memiliki nilai *VIF* lebih dari 5, sedangkan kasus multikolinieritas yang serius terdapat pada sektor ekonomi, khususnya pada variabel prosentase penduduk usia 15 tahun ke atas yang bekerja (X_{11}), tingkat pengangguran terbuka (X_{12}) dan tingkat partisipasi angkatan kerja (X_{13}) yang memiliki nilai *VIF* lebih dari 10.

B. Pemodelan Regresi Logistik Ridge

Pemodelan regresi logistik ridge bertujuan untuk mengatasi kasus multikolinieritas yang terjadi pada faktor-faktor yang mempengaruhi IPM. Didapatkan bahwa jumlah dari *Principal Component* (r) yang memiliki selisih antara r dan derajat bebas (df) paling kecil adalah 4. Sehingga digunakan 4 *PC* dan didapatkan nilai *ridge parameter* sebesar 0,094133. Dengan menggunakan *Software R*, hasil koefisien model lengkap Regresi Logistik Ridge ditampilkan pada Tabel 4.

Tabel 4
Signifikansi Parameter Model Lengkap

Variabel	Koefisien	Koefisien (scaled)	Std. error	t-value (scaled)	P-value
Konstan	-0,1771	-	-	-	-
X ₁ *	-0,0442	-3,3999	0,8601	-3,9530	0,0001
X ₂	-0,0412	-1,4890	1,0355	-1,4380	0,1504
X ₃	0,0000	-0,0527	0,9210	-0,0572	0,9544
X ₄ *	-0,0754	-2,7364	0,7910	-3,4594	0,0005
X ₅ *	0,0371	2,6081	0,9219	2,8292	0,0047
X ₆	0,0003	0,6892	0,7965	0,8652	0,3869
X ₇	0,0016	0,3440	0,6406	0,5371	0,5912
X ₈	-0,0162	-1,0453	0,9847	-1,0615	0,2885
X ₉	0,0009	0,2585	0,6053	0,4270	0,6694
X ₁₀	0,1653	0,5336	0,8760	0,6091	0,5424
X ₁₁	0,0108	0,3284	0,7000	0,4692	0,6389
X ₁₂	0,0340	0,3516	0,8021	0,4383	0,6612
X ₁₃	0,0191	0,4937	0,7613	0,6485	0,5166

Pengujian signifikansi parameter secara serentak menggunakan statistik uji *likelihood ratio test* dengan hipotesis.

Hipotesis : $H_0 : \hat{\beta}_1^{\oplus} = \hat{\beta}_2^{\oplus} = \dots = \hat{\beta}_p^{\oplus} = 0$

$H_1 : \text{Minimal ada satu } \hat{\beta}_j^{\oplus} \neq 0, j=1, 2, \dots, p$

Nilai statistik uji *likelihood ratio test* sebesar 45,398. Daerah penolakan $\chi^2_{(13;0,10)}$ adalah 22,362. Didapatkan keputusan menolak H_0 karena nilai statistik uji $[G = 45,398] \geq [\chi^2_{(13;0,05)} = 22,362]$, sehingga dapat disimpulkan bahwa paling sedikit terdapat satu variabel prediktor yang berpengaruh signifikan terhadap variabel respon, yaitu IPM untuk masing-masing kabupaten/ kota di Provinsi Jawa Timur.

Pengujian signifikansi parameter regresi secara parsial menggunakan statistik uji *Wald* dengan hipotesis berikut ini.

Hipotesis : $H_0 : \hat{\beta}_j^{\oplus} = 0$

$H_1 : \hat{\beta}_j^{\oplus} \neq 0, j=1, 2, \dots, p$

Dengan melihat nilai statistik uji pada Tabel 4, dapat dilihat bahwa terdapat tiga variabel prediktor yang memiliki *p-value* < $[\alpha = 0,05]$. Sehingga disimpulkan bahwa variabel angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5) berpengaruh signifikan terhadap indeks pembangunan manusia (IPM) pada taraf $\alpha = 0,05$.

Didapatkan model lengkap Regresi Logistik Ridge untuk IPM kabupaten/kota di Provinsi Jawa Timur pada persamaan (20).

$$\pi_i(x_i) = \frac{\exp(x_i \hat{\beta}^{\oplus})}{1 + \exp(x_i \hat{\beta}^{\oplus})} \tag{20}$$

Dimana : $x_i \hat{\beta}^{\oplus} = -0,1771 - 0,0442x_1 - 0,0412x_2 + 0,0000x_3 - 0,0754x_4 + 0,0371x_5 + 0,0003x_6 + 0,0016x_7 - 0,0162x_8 + 0,0009x_9 + 0,1653x_{10} + 0,0108x_{11} + 0,0340x_{12} + 0,0191x_{13}$

Akurasi (ketepatan klasifikasi) dari hasil prediksi model lengkap Regresi Logistik Ridge dengan melibatkan 13 variabel prediktor sebesar 97,37%.

Interpretasi hasil model lengkap menggunakan nilai *odds ratio* variabel yang signifikan ditampilkan pada Tabel 5.

Tabel 5
Odds Ratio Hasil Pemodelan Regresi Logistik Ridge

Variabel	Koefisien	OR = Exp(10 × $\hat{\beta}^{\oplus}$)	1/OR
Konstan	-0,1771	-	-
Angka Kematian Bayi	-0,0442	0,6425	1,5564
Angka Buta Huruf	-0,0754	0,4707	2,1246
Angka Partisipasi Sekolah	0,0371	1,4485	0,6904

Pada Tabel 5 didapatkan bahwa setiap kenaikan 10% angka kematian bayi, maka kabupaten/kota berpeluang untuk tergolong IPM menengah bawah 1,5 kali lipat dibandingkan IPM menengah atas. Setiap kenaikan 10% angka buta huruf, maka kabupaten/kota akan berpeluang tergolong IPM menengah bawah dua kali lipat dibandingkan IPM menengah atas. Tetapi sebaliknya pada angka partisipasi sekolah, yaitu setiap kenaikan 10% angka partisipasi sekolah, maka kabupaten/kota akan berpeluang tergolong IPM menengah atas 1,4 kali lipat dibandingkan IPM menengah bawah.

C. Pemilihan Model Regresi Logistik Ridge Terbaik

Dengan menggunakan *backward elimination* didapatkan model terbaik dengan tiga variabel prediktor yang berpengaruh signifikan terhadap variabel respon. Didapatkan bahwa jumlah dari *principal component* (r) yang memiliki selisih antara r dan derajat bebas (df) paling kecil adalah 2. Perhitungan nilai *ridge parameter* pada model terbaik digunakan 2 *PC* dan didapatkan nilai *ridge parameter* sebesar 0,048791. Dengan menggunakan *Software R*, hasil koefisien model lengkap Regresi Logistik Ridge ditampilkan pada Tabel 6.

Tabel 6
Signifikansi Parameter Model Terbaik

Variabel	Koefisien	Koefisien (scaled)	Std. error	t-value (scaled)	P-value
Konstan	2,62665	-	-	-	-
X ₁ *	-0,06379	-4,90166	1,350	-3,631	0,000283
X ₄ *	-0,10456	-3,79675	1,380	-2,751	0,005942
X ₅ *	0,04869	3,42683	1,433	2,391	0,016792

Pengujian signifikansi parameter secara serentak menggunakan statistik uji *likelihood ratio test* dengan hipotesis.

Hipotesis : $H_0 : \hat{\beta}_1^{\oplus} = \hat{\beta}_2^{\oplus} = \dots = \hat{\beta}_p^{\oplus} = 0$

$H_1 : \text{Minimal ada satu } \hat{\beta}_j^{\oplus} \neq 0, j=1, 2, \dots, p$

Nilai statistik uji *likelihood ratio test* sebesar 26,101. Daerah penolakan $\chi^2_{(3;0,05)} = 7,815$ yang artinya diputuskan menolak H_0 karena nilai $[G = 26,1012] \geq [\chi^2_{(3;0,05)} = 7,815]$, sehingga disimpulkan bahwa paling sedikit terdapat satu variabel prediktor yang berpengaruh signifikan terhadap IPM Jawa Timur.

Pada pengujian signifikansi parameter secara parsial digunakan statistik uji *Wald* dengan hipotesis sebagai berikut.

Hipotesis : $H_0 : \hat{\beta}_j^{\otimes} = 0$

$H_1 : \hat{\beta}_j^{\otimes} \neq 0, j=1, 2, \dots, p$

Dengan melihat nilai statistik uji pada Tabel 6, dapat diketahui bahwa ketiga variabel prediktor memiliki *p-value* < [$\alpha = 0,05$]. Sehingga dapat disimpulkan bahwa variabel angka kematian bayi (X_1), angka buta huruf (X_4) dan angka partisipasi sekolah (X_5) berpengaruh signifikan terhadap Indeks Pembangunan Manusia (IPM) pada taraf $\alpha = 0,05$.

Model terbaik Regresi Logistik Ridge untuk IPM kabupaten/kota di Provinsi Jawa Timur ditampilkan pada persamaan (21).

$$\pi_i(x_i) = \frac{\exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^{\otimes})}{1 + \exp(\mathbf{x}_i \hat{\boldsymbol{\beta}}^{\otimes})} \quad (21)$$

Dimana : $\mathbf{x}_i \hat{\boldsymbol{\beta}}^{\otimes} = 2,62665 - 0,06379x_1 - 0,10456x_4 + 0,04869x_5$

Akurasi (ketepatan klasifikasi) dari hasil prediksi model Regresi Logistik Ridge terbaik dengan melibatkan 3 variabel prediktor sebesar 97,37%. Interpretasi hasil pemodelan menggunakan nilai *odds ratio* yang ditampilkan pada Tabel 7.

Tabel 7

Odds Ratio Hasil Pemodelan Regresi Logistik Ridge Terbaik			
Variabel	Koefisien	OR = $Exp(10 \times \hat{\beta}^{\otimes})$	1/OR
Konstan	2,626650	-	-
Angka Kematian Bayi	-0,063790	0,52840	1,89250
Angka Buta Huruf	-0,104560	0,35148	2,84511
Angka Partisipasi Sekolah	0,048690	1,62726	0,61453

Pada Tabel 7 didapatkan bahwa setiap kenaikan 10% angka kematian bayi, maka kabupaten/kota berpeluang untuk tergolong IPM menengah bawah dua kali lipat dibandingkan IPM menengah atas. Setiap kenaikan 10% angka buta huruf, maka kabupaten/kota akan berpeluang tergolong IPM menengah bawah tiga kali lipat dibandingkan IPM menengah atas. Tetapi pada angka partisipasi sekolah berlaku sebaliknya, yaitu setiap kenaikan 10% angka partisipasi sekolah, maka kabupaten/kota akan berpeluang tergolong IPM menengah atas 1,6 kali lipat dibandingkan IPM menengah bawah.

Hasil prediksi dengan menggunakan model terbaik Regresi Logistik Ridge menghasilkan 5 kabupaten/kota tergolong IPM menengah bawah dan 33 kabupaten/kota lain tergolong dalam IPM menengah atas. Terdapat satu kesalahan prediksi, yaitu pada Kabupaten Bondowoso dimana pada data observasi Kabupaten Bondowoso tergolong dalam IPM menengah bawah, sedangkan pada hasil prediksi menunjukkan bahwa Kabupaten Bondowoso tergolong dalam IPM menengah atas.

V. KESIMPULAN DAN SARAN

A. Kesimpulan

Berdasarkan pada hasil analisis dapat diambil kesimpulan :

1. Terdapat kasus multikolinieritas pada faktor-faktor yang mempengaruhi IPM di Provinsi Jawa Timur, yaitu pada variabel angka buta huruf (X_4), rasio sekolah-murid (X_7), prosentase penduduk usia 15 tahun ke atas yang bekerja (X_{11}), tingkat pengangguran terbuka (X_{12}) dan tingkat partisipasi angkatan kerja (X_{13}).

2. Model terbaik Regresi Logistik Ridge IPM di Provinsi Jawa Timur dengan *ridge parameter* sebesar 0,04879089 adalah sebagai berikut.

$$\pi_i(x_i) = \frac{\exp(2,62665 - 0,06379x_1 - 0,10456x_4 + 0,04869x_5)}{1 + \exp(2,62665 - 0,06379x_1 - 0,10456x_4 + 0,04869x_5)}$$

3. Ketepatan klasifikasi model terbaik IPM di Provinsi Jawa Timur sebesar 97,3684%.
4. Prediksi model terbaik Regresi Logistik Ridge menghasilkan 5 kabupaten/kota tergolong IPM menengah bawah. Sedangkan 33 kabupaten/kota lainnya tergolong IPM menengah atas.

B. Saran

Saran yang diberikan pada penelitian ini adalah :

1. Perlu adanya penelitian mengenai metode lain sebagai perbandingan dengan Regresi Logistik Ridge, untuk dapat diketahui kelebihan dan kekurangan masing-masing.
2. Pemerintah Provinsi Jawa Timur harus memperhatikan pemerataan pembangunan manusia untuk setiap sektornya.

DAFTAR PUSTAKA

- [1] Badan Pusat Statistik (BPS), *Penduduk Indonesia Menurut Provinsi 1971, 1980, 1990, 1995, 2000 dan 2010*. Diambil kembali dari Statistics Indonesia (2010): http://bps.go.id/tab_sub/view.php?kat=1&tabel=1&daftar=1&id_subyek=12¬ab=1.
- [2] D. Darwis, *Kependudukan Dalam Perspektif Pembangunan Ekonomi Guna Pembangunan Nasional*. Jawa Barat: BKKbN (2011).
- [3] Badan Pusat Statistik (BPS), *Konsep Indeks Pembangunan Manusia*. Diambil kembali dari Statistics Indonesia (2014): http://bps.go.id/menutab.php?kat=1&tabel=1&id_subyek=26.
- [4] Badan Pusat Statistik (BPS), *Indeks Pembangunan Manusia dan Komponennya*. Diambil kembali dari Statistics Indonesia (2013): http://bps.go.id/ipm.php?id_subyek=26¬ab=0.
- [5] X. Yan, & X. G. Su, *Linear Regression Analysis : Theory and Computing*. Singapore: World Scientific (2009).
- [6] A. Agresti, *An Introduction to Categorical Data Analysis Second Edition*. United State of America: A John Wiley & Sons, Inc (2007).
- [7] T. P. Ryan, *Modern Regression Methods*. New York: John Wiley & Sons (1997).
- [8] N. R. Draper, & H. Smith, *Applied Regression Analysis : Third Edition*. Canada: John Wiley & Sons (1998).
- [9] Sunyoto, *Regresi Logistik Ridge : Pada Keberhasilan Siswa SMA Negeri 1 Kediri Diterima Di Perguruan Tinggi Negeri*. Surabaya: Institut Teknologi Sepuluh Nopember (2009).
- [10] H. Vago, & S. Kemény, *Logistic Ridge For Clinical Data Analysis (A Case Study)*. Applied Ecology And Environmental Research 4(2) (2006), 171-179.
- [11] D. W. Hosmer, & S. Lemeshow, *Applied Logistic Regression Second Edition*. New York, United State of America: John Wiley & Sons, Inc (2000).
- [12] Human Development Report (HDR), *Sustaining Human Progress : Reducing Vulnerabilities and Building Resilience*. New York, United State of America: United Nations Development Programme (UNDP) (2014).