

Metode Regresi *Elastic-net* untuk Mengatasi Masalah Multikolinearitas pada Kasus Tingkat Pengangguran Terbuka di Provinsi Jawa Barat

Astri Handayani*, Lisnur Wachidah

Prodi Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Universitas Islam Bandung, Indonesia.

*astrihandayani45@gmail.com, wachidah.lisnur07@gmail.com

Abstract. Linear regression is used to study the relationship between dependent variables and one or more independent variables. The most widely used linear regression parameter estimation is the Least Squares Method (MKT). Multiple linear regression involves more than one independent variable. The common problem that is often encountered in multiple linear regression is the problem of multicollinearity. Multicollinearity in multiple linear regression occurs when there is a correlation between independent variables, with multicollinearity it is often difficult to see the effect of the independent variables on the dependent variables. Multicollinearity handling can be done using the elastic-net regression method where this method can reduce the regression coefficient to exactly zero, besides that the regression nets also select variables simultaneously and can select groups of correlated variables. In 2020 the open response rate in West Java Province was relatively high, namely 10.46, exceeding the national average of 7.07. This study uses data regarding the open response rate in West Java Province in 2020. Based on the results of testing the independent variables that influence the open response rate in West Java, namely the dependency ratio, average length of schooling, human development index, percentage of poor people, wages minimum district/city, and population density.

Keywords: *Linear Regression, Least Square Metode, Multicollinearity, Elastic-net, Open Unemployment Rate.*

Abstrak. Regresi linier digunakan untuk mempelajari hubungan antara peubah tak bebas dengan satu atau lebih peubah bebas. Pendugaan parameter regresi linier yang paling banyak digunakan yaitu Metode Kuadrat Terkecil (MKT). Regresi linier berganda melibatkan lebih dari satu peubah bebas. Masalah umum yang sering dijumpai pada regresi linier berganda yaitu adanya masalah multikolinearitas. Multikolinearitas pada regresi linier berganda terjadi apabila terdapat korelasi antar peubah bebas, keberadaan multikolinearitas seringkali menyulitkan untuk melihat pengaruh antara peubah bebas terhadap peubah tak bebas. Penanganan multikolinearitas dapat dilakukan menggunakan metode regresi *elastic-net* dimana metode ini dapat menyusutkan koefisien regresi tepat nol, selain itu regresi *elastic-net* juga melakukan seleksi peubah secara simultan dan dapat memilih kelompok peubah yang berkorelasi. Pada tahun 2020 tingkat pengangguran terbuka di Provinsi Jawa Barat relatif tinggi yaitu sebesar 10,46 melebihi nilai rata-rata nasional sebesar 7,07. Penelitian ini menggunakan data mengenai tingkat pengangguran terbuka di Provinsi Jawa Barat pada tahun 2020. Berdasarkan hasil pengujian peubah bebas yang memberikan pengaruh terhadap tingkat pengangguran terbuka di Jawa Barat yaitu *dependency ratio*, rata-rata lama sekolah, indeks pembangunan manusia, presentase penduduk miskin, upah minimum kabupaten/kota, dan kepadatan penduduk.

Kata Kunci: *Regresi Linier, MKT, Multikolinearitas, Elastic-net, Tingkat Pengangguran Terbuka.*

A. Pendahuluan

Teknik yang digunakan untuk mempelajari hubungan antara peubah tak bebas dengan satu atau lebih peubah bebas disebut dengan regresi linier [1]. Regresi linier dengan peubah bebas yang berjumlah lebih dari satu dikenal dengan regresi linier berganda. Dalam regresi linier berganda umumnya sering dijumpai adanya masalah multikolinearitas. Multikolinearitas dapat terjadi apabila terdapat korelasi antara dua ataupun lebih peubah bebas, dengan adanya multikolinearitas pada peubah bebas dapat menyebabkan ragam penduga menjadi besar [2].

Pendugaan parameter regresi linier yang paling banyak digunakan yaitu Metode Kuadrat Terkecil (MKT) [3]. MKT memiliki tujuan yaitu untuk meminimumkan jumlah kuadrat dari sisaan (*error of sum square*). Pada kondisi-kondisi tertentu pendugaan koefisien menggunakan MKT sering memberikan hasil yang tidak memuaskan, misalnya saat terjadi multikolinearitas. Hal ini dapat menyebabkan masalah pada keakuratan prediksi yang berdampak pada penduga MKT yang memiliki bias yang rendah tetapi memiliki ragam besar [3].

Permasalahan multikolinearitas dapat ditangani menggunakan metode regresi *elastic-net*, metode ini merupakan metode regularisasi gabungan antara regresi *ridge* dan *Least Absolute Shrinkage and Selection Operator* (LASSO). Regresi *elastic-net* mampu menyusutkan koefisien regresi menjadi tepat nol, selain itu regresi *elastic-net* juga mampu melakukan seleksi pada peubah secara simultan serta dapat memilih kelompok peubah yang saling berkorelasi [4].

Permasalahan yang kerap ditemui di Indonesia sebagai negara berkembang yaitu belum tercapainya kesejahteraan sosial bagi masyarakat [5]. Adapun upaya yang dilakukan pemerintah Indonesia sebagai upaya meningkatkan kesejahteraan sosial bagi masyarakat diantaranya dengan mengadakan berbagai program yang ditujukan untuk membantu masyarakat dalam memenuhi kebutuhan dasar hidupnya. Namun pemberian bantuan ini tidak serta merta dapat menjamin kesejahteraan masyarakat, sebab masih banyak faktor-faktor yang lebih signifikan lainnya yang harus menjadi fokus pemerintah.

Upaya yang dapat dilakukan pemerintah untuk meningkatkan tingkat kesejahteraan sosial bagi masyarakat salah satunya yaitu dengan cara menekan angka pengangguran. Sebab pengangguran merupakan salah satu dari permasalahan kesejahteraan sosial yang sering terjadi di Indonesia. Pengangguran di Indonesia dapat disebabkan oleh banyak faktor, beberapa diantaranya yaitu disebabkan oleh kurangnya kompetensi yang dimiliki oleh pencari kerja serta jumlah pencari kerja yang melebihi permintaan dari pasar kerja.

Dibandingkan dengan nilai rata-rata nasional tingkat pengangguran terbuka di Jawa Barat masih tergolong relatif tinggi. Bersumber dari hasil rilis Badan Pusat Statistik Jawa Barat, tingkat pengangguran terbuka di Jawa Barat mengalami fluktuasi sejak tahun 2014 hingga tahun 2020. Pada tahun 2020 tingkat pengangguran terbuka di Jawa Barat mencapai nilai 10,46 melebihi nilai rata-rata nasional sebesar 7,07. Berdasarkan hal tersebut dapat disimpulkan bahwa terdapat 10 orang angkatan kerja yang belum memperoleh pekerjaan dari total 100 orang angkatan kerja. Untuk meminimumkan nilai tersebut pemerintah diharapkan dapat membuat kebijakan yang optimum untuk menekan angka pengangguran di Jawa Barat. Mengingat permasalahan pengangguran ini dapat berimbas pada sektor lain, seperti keamanan dan juga kenyamanan dalam bermasyarakat apabila tidak ditangani dengan serius oleh pemerintah.

Berdasarkan latar belakang tersebut, maka pada penelitian ini akan dilakukan analisis untuk mengetahui bagaimana penerapan metode regresi *elastic-net* dalam menangani multikolinearitas pada data tingkat pengangguran terbuka di Jawa Barat serta untuk mengetahui model regresi *elastic-net* pada tingkat pengangguran terbuka di Jawa Barat pada tahun 2020.

B. Metodologi Penelitian

Jenis dan Sumber Data

Pada penelitian ini menggunakan data sekunder, yaitu sebuah data yang diperoleh dari sumber yang sudah ada. Data yang digunakan adalah data yang diperoleh dari Badan Pusat Statistik (BPS) Provinsi Jawa Barat dan Bank Indonesia. Unit observasi yang digunakan adalah kabupaten/kota di Jawa Barat pada tahun 2020 yang terdiri dari 18 kabupaten dan 9 kota. Sehingga secara keseluruhan terdapat sebanyak 27 kabupaten/kota.

Data yang diperlukan meliputi tingkat pengangguran terbuka sebagai sebagai peubah tak bebas (Y), *Dependency Ratio* (X_1), rata-rata lama sekolah (X_2), indeks pembangunan manusia (X_3), presentase penduduk miskin (X_4), upah minimum kabupaten/kota (X_5), kepadatan penduduk (X_6).

Metode

Dalam penelitian ini akan dilakukan analisis menggunakan metode regresi *elastic-net*, adapun tahapan analisis diuraikan sebagai berikut:

1. Melakukan standarisasi data dengan Z-score
2. Melakukan pemodelan dengan menggunakan MKT

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ij} + \varepsilon_i \quad (1)$$

Keterangan:

Y_i : Peubah tak bebas

X_i : Peubah bebas

β_j : Koefisien regresi yang tidak diketahui dengan $j=0, 1, 2, \dots, k$

ε_i : Sisaan

i : $1, 2, \dots, n$

Persamaan diatas dapat dituliskan dengan persamaan matriks:

$$Y = X\beta + \varepsilon \quad (2)$$

3. Pemeriksaan asumsi regresi linier hasil MKT

a. Uji normalitas

b. Uji homogenitas dilakukan uji Harley

$$F = \frac{\text{varians terbesar}}{\text{varians terkecil}} \quad (3)$$

- c. Mendeteksi multikolinearitas dilakukan dengan melihat nilai *Variance Inflation Factor* (VIF). Nilai VIF dengan indikasi terjadi multikolinearitas yaitu memiliki nilai lebih dari 10.

$$VIF = \frac{1}{1 - R_j^2} \quad (4)$$

4. Membagi data menjadi 2 bagian, 80% data *training* dan 20% data *testing*.

5. Melakukan analisis dengan menggunakan metode regresi *elastic-net*.

a. Menentukan rentang nilai alpha yaitu $0 \leq \alpha \leq 1$

b. Menentukan nilai α dan λ yang paling optimum dengan data *training* menggunakan *5-fold cross validation*. Pada proses ini akan menghasilkan dugaan kuadrat tengah galat (*mean square error*-MSE), $MSE_1, MSE_2, \dots, MSE_k$. Nilai MSE dapat diperoleh dengan rumus berikut:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (5)$$

Keterangan:

n : banyaknya observasi

y_i : data sebenarnya

$\hat{y}_i$: nilai prediksi dari peubah y

Dugaan galat validasi silang (*cross validation error*-CVE) dapat diperoleh dengan menghitung rata-rata dari persamaan sebagai berikut:

$$CVE_{(k)} = \frac{1}{k} \sum_{i=1}^k MSE_i \quad (6)$$

Keterangan:

k : nilai fold

MSE : *mean square error*

Tujuan dari *cross validation* yaitu untuk mencari nilai alpha (α) dan lambda (λ) yang paling optimum yang akan digunakan sebagai penalti pada regresi *elastic-net*. Dimana nilai optimum tersebut diperoleh berdasarkan nilai CVE minimum. Nilai lambda dapat diperoleh dengan rumus berikut:

$$\hat{\lambda}_k = \underset{\lambda \in \{\lambda_1, \dots, \lambda_k\}}{\operatorname{argmin}} CVE \quad (7)$$

Melakukan pendugaan parameter model regresi.

$$\beta = (\beta_0, \beta) \operatorname{argmin} [i=1n(y_i - \beta_0 - \sum_{j=1}^p x_{ij}\beta_j)^2 + \lambda \sum_{j=1}^p (\alpha\beta_j + (1-\alpha)\beta_j^2)] \quad (8)$$

Keterangan:

y_i : nilai pengamatan ke- i , $i = 1, 2, \dots, n$

β_0 : intersep

x_{ij} : matriks peubah bebas berukuran $n \times (p+1)$

β_j : koefisien kemiringan peubah ke- j

λ : parameter penyusutan

α : koefisien regresi *ridge*

$(1 - \alpha)$: koefisien regresi LASSO

Melakukan pemodelan regresi *elastic-net* menggunakan persamaan

$$Y^* = X^* \beta + \varepsilon^* \quad (9)$$

6. Mengukur kebaikan model

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \quad \text{dimana } 0 \leq R^2 \leq 1 \quad (10)$$

C. Hasil Penelitian dan Pembahasan

Analisis Regresi Linier

Analisis regresi linier berganda dilakukan untuk mengetahui arah dan seberapa besar pengaruh dari peubah bebas terhadap peubah tak bebas. Secara umum regresi linier dibagi menjadi dua jenis, yaitu regresi linier sederhana yang hanya memiliki satu peubah bebas dan regresi linier berganda yang memiliki dua atau lebih peubah bebas di dalamnya. Dugaan parameter regresi yang dihasilkan yaitu:

Tabel 1. Analisis Regresi

Peubah	Koefisien MKT
Intersep	-4,513E-11
X_1	0,540
X_2	0,173
X_3	0,530
X_4	0,521
X_5	0,895
X_6	1,055

Model regresi linier berganda yang dihasilkan berdasarkan Tabel 1 di atas yaitu:

$$Y = -4,513e^{-11} + 0,540X_1 + 0,173X_2 + 0,503X_3 + 0,521X_4 + 0,895X_5 + 1,055X_6$$

Nilai konstanta bertanda negatif, yaitu $-4,513e^{-11}$ artinya apabila *dependency ratio* (X_1), rata-rata lama sekolah (X_2), indeks pembangunan manusia (X_3), presentase penduduk miskin (X_4), upah minimum kabupaten/kota (X_5) dan kepadatan penduduk (X_6) bernilai nol maka tingkat pengangguran terbuka (Y) akan mengalami penurunan sebesar $4,513e^{-11}$.

Uji Normalitas

Uji normalitas digunakan untuk mengetahui apakah sisaan berdistribusi normal atau tidak. Pengujian normalitas dilakukan menggunakan uji Lilliefors dengan rumusan hipotesis sebagai berikut:

H_0 : Sisaan berdistribusi normal

H_1 : Sisaan tidak berdistribusi normal

1. Statistik uji

Tabel 2. Uji Normalitas

L_o	L_{tabel}	Kesimpulan
0,2	1,1665	Normal

2. Kriteria uji:

Tolak H_0 jika $L_o > L_{tabel}$

Kesimpulan:

Berdasarkan tabel di atas, dengan menggunakan uji normalitas diperoleh nilai L_o sebesar 0,2 lebih kecil dari nilai L_{tabel} 1,1665. Dengan demikian H_0 diterima, artinya sisaan berdistribusi normal.

Uji Homogenitas

Uji homogenitas variansi menggunakan uji Harley merupakan uji yang sangat sederhana karena cukup membandingkan variansi terbesar dengan variansi terkecil. Langkah-langkah pengujian dengan uji Harley sebagai berikut:

1. Hipotesis

 $H_0 : \sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma_4^2 = \sigma_5^2 = \sigma_6^2$ (varians data homogen) $H_1 : \text{salah satu } \sigma^2 \text{ tidak sama (varians data tidak homogen)}$

2. Statistik uji

Tabel 3. Uji Homogenitas

F_{hitung}	F_{tabel}	Kesimpulan
1,0549	2,68	Homogen

3. Kriteria uji

Tolak H_0 jika $F_{hitung} \geq F_{tabel}$

Kesimpulan: Berdasarkan hasil perhitungan uji homogenitas pada Tabel 3 diperoleh nilai $F_{hitung} < F_{tabel 0,05(5,21)}$ atau $1,0549 < 2,68$ maka dapat disimpulkan bahwa H_0 diterima, artinya varians data bersifat homogen.

Multikolienearitas

Multikolienearitas dapat terjadi jika antar peubah bebas dalam suatu model regresi memiliki korelasi. Multikolienearitas hanya terjadi pada regresi linier berganda yang melibatkan dua atau lebih peubah bebas dan hanya terjadi pada hubungan linier.

Terdapatnya permasalahan multikolienearitas dapat mengakibatkan ragam penduga menjadi besar, akibat lainnya yaitu akan menyulitkan untuk melihat pengaruh peubah bebas terhadap peubah tak bebas. Multikolienearitas dapat di deteksi dengan melihat nilai dari VIF pada masing-masing peubah bebas. Nilai VIF dengan indikasi terjadi multikolienearitas yaitu memiliki nilai lebih dari 10.

Tabel 4. Nilai VIF

Peubah	Koefisien MKT	Sig.	VIF
Intersep	-4,513e-11	1,000	
x1	0,540	0,004	6,493
x2	0,173	0,035	15,469
x3	0,530	0,042	24,026
x4	0,521	0,041	2,755
x5	0,895	0,000	1,915
x6	1,055	0,017	8,049

Berdasarkan Tabel 4 diperoleh informasi bahwa pada peubah bebas X_2 dan X_3 memiliki nilai VIF lebih dari 10, artinya terdapat masalah multikolinearitas pada peubah bebas. Dampak dari adanya multikolinearitas yaitu menyebabkan model yang dihasilkan kurang efisien.

Regresi Elastic-net

Regularisasi *elastic-net* menggabungkan dua penalti penyusutan, yaitu penalti LASSO ($\sum |\beta_j|$) dan penalti *ridge* ($\sum \beta_j^2$). Parameter untuk penalti LASSO adalah $\lambda \times \alpha$ sedangkan parameter untuk penalti *ridge* adalah $\lambda \times (1-\alpha)$. Penentuan nilai α dan λ yang akan digunakan untuk pemodelan ditentukan melalui *5-fold cross validation* pada data *training*. Pembagian data *training* sebanyak 21 amatan dan data *testing* sebanyak 6 amatan dilakukan secara acak pada data yang tersedia untuk menduga galat prediksi. Kriteria λ optimum diperoleh dengan mencari nilai CVE minimum di antara $\alpha \in (0, 1)$, dimana nilai α yang digunakan untuk pemodelan adalah 0,1 hingga 0,9 [3].

Tabel 5. Nilai Parameter *Elastic-net* Optimum Hasil CV

Alpha (α)	MSE	CVE	Lambda (λ)
0,1	2,0996	0,4199	0,4607
0,2	2,0731	0,4146	0,2304
0,3	2,3617	0,4723	0,2228
0,4	2,2884	0,4577	0,2424
0,5	2,2303	0,4461	0,1940
0,6	2,3803	0,4761	0,1616
0,7	2,3126	0,4625	0,1385
0,8	2,3678	0,4736	0,0694
0,9	2,2089	0,4418	0,0743

Nilai α optimum adalah 0,2 karena menghasilkan CVE paling minimum dibandingkan CVE pada α lainnya yaitu sebesar 0,4146. Nilai λ optimum yang digunakan yaitu sebesar 0,2304.

Model *elastic-net* dibangun menggunakan *software* RStudio versi 4.0.3 dengan *package* glmnet. Tabel 6 menunjukkan koefisien peubah-peubah bebas hasil analisis dengan regresi *elastic-net*.

Tabel 6. Koefisien Hasil *Elastic-net*

Peubah	MKT	<i>Elastic-net</i>
(Intercept)	$-4,513e^{-11}$	0,107
X_1	0,540	0,031
X_2	0,173	0,088
X_3	0,530	0,718
X_4	0,521	0,245
X_5	0,895	0,546
X_6	1,055	0,301

Parameter penyusutan (λ) pada regresi *elastic-net* menyebabkan dugaan koefisien regresi yang dihasilkan lebih kecil dibandingkan MKT. Penyusutan koefisien hanya dilihat untuk peubah bebas saja dan tidak mencakup intersep. Berdasarkan Tabel 6 diketahui bahwa nilai dugaan koefisien regresi tidak ada yang bernilai nol, artinya seluruh peubah bebas memberikan pengaruh terhadap tingkat pengangguran terbuka.

Berdasarkan hasil taksiran regresi pada Tabel 6 maka model regresi *elastic-net* yang diperoleh adalah sebagai berikut:

$$Y^* = 0,107 + 0,031X_1 + 0,088X_2 + 0,718X_3 + 0,245X_4 + 0,546X_5 - 0,301X_6$$

Model di atas memiliki pengertian bahwa apabila terjadi peningkatan pada *dependency ratio*, rata-rata lama sekolah, indeks pembangunan manusia, presentase penduduk miskin dan upah minimum kabupaten/kota dan kepadatan penduduk sebesar 1% maka tingkat pengangguran terbuka akan mengalami peningkatan sebesar 0,107%.

Ukuran Kebaikan Model

Pemilihan model terbaik dari MKT dan regresi *elastic-net* berdasarkan pada nilai koefisien determinasi (R^2) dari data *testing*.

MKT	<i>Elastic-net</i>
58,80	85,26

Elastic-net menghasilkan nilai R^2 yang lebih besar dari MKT, sehingga *elastic-net* menghasilkan model terbaik. Nilai R^2 sebesar 85,26 yang dihasilkan oleh *elastic-net* artinya sebanyak 85,26% tingkat pengangguran terbuka di Jawa Barat dapat dijelaskan oleh peubah bebas pada model sedangkan sisanya 14,74% dijelaskan oleh peubah lain diluar model.

D. Kesimpulan

Berdasarkan pembahasan dalam penelitian ini, peneliti menyimpulkan beberapa hasil penelitian sebagai berikut:

1. Penerapan analisis regresi *elastic-net* diperoleh nilai CVE minimum sebesar 0,4146 serta nilai α optimum sebesar 0,2 dan nilai λ optimumnya sebesar 0,2304. Peubah bebas yang berpengaruh terhadap tingkat pengangguran terbuka di Jawa Barat tahun 2020, yaitu *dependency ratio* (X_1), rata-rata lama sekolah (X_2), indeks pembangunan manusia (X_3), presentase penduduk miskin (X_4), upah minimum kabupaten/kota (X_5) dan kepadatan penduduk (X_6).
2. Model regresi yang diperoleh yaitu:

$$Y^* = 0,107 + 0,031X_1 + 0,088X_2 + 0,718X_3 + 0,245X_4 + 0,546X_5 - 0,301X_6$$

Acknowledge

Peneliti mengucapkan terima kasih kepada seluruh dosen Program Studi Statistika Universitas Islam Bandung yang telah berdedikasi dan memberikan banyak ilmu terutama Ibu Dr. Lisnur Wachidah, DRA., M.Si., selaku dosen pembimbing sehingga penelitian ini dapat diselesaikan dengan baik.

Daftar Pustaka

- [1] Camartya, Dia, Achmad, Anneke Iswani. (2022). Analisis Korespondensi pada Jumlah Pengangguran Terbuka Menurut Kabupaten/Kota Berdasarkan Pendidikan Tertinggi. *Jurnal Riset Statistika* 2(2). 119-128.
- [2] Fanny, Resty. (2018). *Pendugaan Produktivitas Penangkapan Bagan Perahu di Pantai Carocok Tarusan Sumatera Barat dengan Regresi gulud, LASSO dan Elastic-net*. Skripsi tidak dipublikasikan. Bogor: Departemen Statistika, Fakultas Matematika dan Ilmu Pengetahuan Alam, Institut Pertanian Bogor.
- [3] Greene, W. H. (2012). *Econometric Analysis Ed ke-7*. New York: Pearson.
- [4] Mela, C.F. & P.K. Kopalle. (2002). *The Impact of Collinearity on Regression Analysis: The Asymmetric Effect of Negative and Positive Correlation*. *Applied Economics* 34, 667-677.
- [5] Wijaya, A. F. H. (2018). *Analisis Faktor-Faktor Yang Mempengaruhi Tingkat Pengangguran Terbuka (TPT) di Provinsi Aceh dengan Regresi Nonparamterik Spline Truncated*. Skripsi tidak dipublikasikan. Surabaya: Jurusan Statistika, Fakultas Matematika, Komputasi dan Sains Data, Institut Teknologi Sepuluh November.
- [6] Zou, H. & T, Hastie. (2005). *Regularization and variable selection via the elastic-net*. *Journal of the Royal Statisical Societly. Series B* 67, 301-302.