

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. Multicollinearity is a correlation between independent variables, the existence of multicollinearity is often difficult to see the effect of independent variables on dependent variables. Handling multicollinearity can be done using the *Elastic-net* regression method where this method can shrink the regression coefficient to zero, besides that *Elastic-net* regression also performs simultaneous selection of variables and can select groups of correlated variables. In 2020, the open unemployment rate in West Java Province was relatively high, at 10.46, exceeding the national average of 7.07. This study uses data on the open unemployment rate in West Java Province in 2020. Based on the test results of the seven independent variables used, only five independent variables have an influence on the open unemployment rate in West Java, namely the dependency ratio, the average length of schooling, percentage of poor population, district/city minimum wage, and population density. While the independent variable that has no effect on the open unemployment rate is the human development index. Based on the value of the importance variable, the independent variable that has the greatest influence on the open unemployment rate in West Java Province is the district/city minimum wage.

Keywords: *Multicollinearity, Elastic-Net, Open Unemployment Rate, Importance Variable, District/City Minimum Wage.*

Abstrak. Multikolinearitas merupakan 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 dari tujuh peubah bebas yang digunakan hanya lima peubah bebas saja yang memberikan pengaruh terhadap tingkat pengangguran terbuka di Jawa Barat yaitu dependency ratio, rata-rata lama sekolah, presentase penduduk miskin, upah minimum kabupaten/kota, dan kepadatan penduduk. Sedangkan peubah bebas yang tidak memberikan pengaruh terhadap tingkat pengangguran terbuka yaitu indeks pembangunan manusia. Berdasarkan nilai importance variable peubah bebas yang memberikan pengaruh paling besar terhadap tingkat pengangguran terbuka di Provinsi Jawa Barat adalah upah minimum kabupaten/kota.

Kata Kunci: *Multikolinearitas, Elastic-Net, Tingkat Pengangguran Terbuka, Importance Variable, Upah Minimum Kabupaten/Kota.*

A. Pendahuluan

Multikolinearitas merupakan korelasi antara dua ataupun lebih peubah bebas, adanya multikolinearitas dapat menyebabkan ragam penduga menjadi besar (Mela dan Kopalle, 2002). 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* dapat menyusutkan koefisien regresi tepat nol, selain itu regresi *Elastic-net* juga melakukan seleksi peubah secara simultan serta dapat memilih kelompok peubah yang berkorelasi (Zou dan Hastie, 2005).

Permasalahan yang kerap ditemui di Indonesia sebagai negara berkembang yaitu belum tercapainya kesejahteraan sosial bagi masyarakat (Wijaya, 2018). Salah satu upaya yang dapat dilakukan pemerintah untuk meningkatkan tingkat kesejahteraan sosial bagi masyarakat yaitu dengan cara menekan angka pengangguran. Sebab pengangguran merupakan salah satu dari permasalahan kesejahteraan sosial yang sering terjadi di Indonesia. Pengangguran dapat disebabkan oleh banyak hal, beberapa diantaranya yaitu karena kompetensi yang dimiliki oleh pencari kerja yang tidak sesuai dengan permintaan di pasar kerja serta pencari kerja yang jumlahnya melebihi permintaan 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 yang telah diuraikan, maka perumusan masalah dalam penelitian ini sebagai berikut: “Bagaimana model regresi *Elastic-net* pada tingkat pengangguran terbuka di Jawa Barat serta faktor apa yang memberikan pengaruh paling signifikan terhadap tingkat pengangguran terbuka di Jawa Barat?”. Selanjutnya, tujuan dalam penelitian ini diuraikan dalam pokok-pokok sbb.

1. Untuk mengetahui model regresi *Elastic-net* pada tingkat pengangguran terbuka di Jawa Barat.
2. Untuk mengetahui faktor apa yang memberikan pengaruh paling signifikan terhadap tingkat pengangguran terbuka di Jawa Barat.

B. Metodologi Penelitian

Penelitian ini menggunakan data sekunder yang bersumber dari Badan Pusat Statistik 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 adalah sebanyak 27 kabupaten/kota.

Data yang diperlukan meliputi tingkat pengangguran terbuka sebagai peubah 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), produk domestik regional bruto (X_6), kepadatan penduduk (X_7).

Tahapan analisis:

1. Melakukan eksplorasi data dengan membuat plot tebaran Y dan X untuk mengetahui pola hubungan.
2. Mendeteksi multikolinearitas, jika $VIF > 10$ maka terdapat multikolinearitas.
3. Melakukan standarisasi data pada seluruh peubah bebas karena tidak memiliki satuan yang sama menggunakan rumus: $Z = \frac{x_i - \bar{X}}{s}$
4. Melakukan cross validation untuk mencari nilai α dan λ yang paling optimum, dengan pertimbangan nilai RMSE yang paling kecil dengan menggunakan rumus:

$$RMSE_j = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{n}}$$

5. Melakukan analisis dengan menggunakan metode regresi *Elastic-net*.
 - a. Melakukan pendugaan parameter model regresi menggunakan rumus:

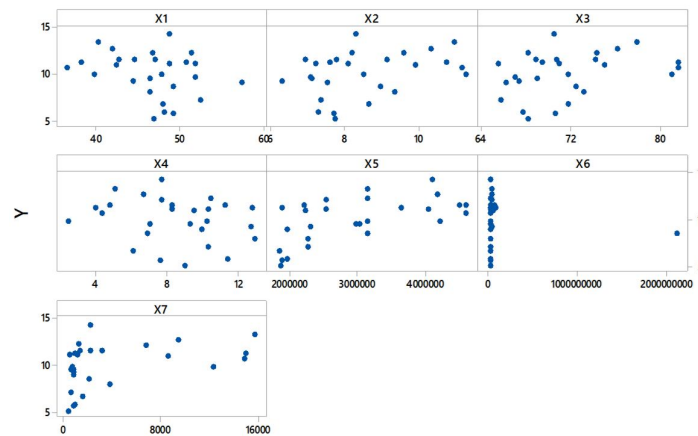
$$\hat{\beta} = \arg \min_{(\beta_0, \beta)} [\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p (\alpha |\beta_j| + (1 - \alpha) \beta_j^2)].$$
 - b. Melakukan pemodelan regresi *Elastic-net* menggunakan rumus:

$$Y = X\beta + \varepsilon$$
 - c. Mencari peubah bebas yang memberikan pengaruh paling besar dengan menggunakan importance variable.
6. Mengukur kebaikan model dengan menggunakan.

C. Hasil Penelitian dan Pembahasan

Eksplorasi Data

Berikut adalah pola hubungan yang terbentuk dari masing-masing peubah berdasarkan pengujian dengan diagram pencar:



Gambar 1. Diagram Pencar Data

Berdasarkan Gambar 4.1 tidak semua peubah bebas memiliki hubungan yang linier dengan tingkat pengangguran terbuka. Secara eksploratif peubah bebas yang memiliki hubungan linier dengan tingkat pengangguran terbuka 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_7), hal ini dikarenakan pola sebaran data dari ke-6 peubah bebas tersebut tidak membentuk pola tertentu. Sedangkan peubah bebas yang tidak memiliki hubungan linier dengan tingkat pengangguran terbuka adalah produk domestik regional bruto (X_6). Maka pada tahapan selanjutnya produk domestik regional bruto tidak akan diikutsertakan dalam proses analisis.

Multikolinearitas

Pemeriksaan multikolinearitas dapat dilakukan melalui tahapan-tahapan berikut:

1. Hipotesis
 - H_0 : Terdapat masalah multikolinearitas pada data tingkat pengangguran terbuka di Jawa Barat.
 - H_1 : Tidak terdapat masalah multikolinearitas pada data tingkat pengangguran terbuka di Jawa Barat.
2. Statistik uji diperoleh dengan menggunakan Persamaan (2.1) sehingga nilai VIF dari masing-masing peubah bebas yaitu:

Tabel 1. Nilai VIF Dari Setiap Peubah Bebas

Peubah	VIF	Keterangan
X_1	6,49	Kurang dari 10
X_2	15,47	Lebih dari 10
X_3	24,03	Lebih dari 10
X_4	2,75	Kurang dari 10
X_5	1,91	Kurang dari 10
X_7	8,05	Kurang dari 10

3. Kriteria Uji

H_0 ditolak jika seluruh peubah bebas memiliki nilai VIF kurang dari 10.

4. Kesimpulan

Berdasarkan Tabel 1 diperoleh informasi bahwa pada peubah bebas X_2 dan X_3 memiliki nilai VIF lebih dari 10, maka H_0 diterima artinya terdapat masalah multikolinearitas pada peubah bebas. Masalah multikolinearitas selanjutnya akan ditangani dengan menggunakan regresi *Elastic-net*.

Standarisasi Data

Standarisasi data dilakukan sebelum dilakukan proses analisis pada data penelitian. Ketika data yang akan diteliti memiliki satuan yang berbeda maka perlu dilakukan standarisasi data terlebih dahulu, standarisasi data tidak akan merubah pola distribusi dari data. Hasil dari standarisasi data disajikan dalam Tabel 2.

Tabel 2. Standarisasi Data Penelitian

Kabupaten/ Kota	ZY	ZX1	ZX2	ZX3	ZX4	ZX5	ZX7
Kabupaten Bogor	1,810543	0,439719	-0,16783	-0,26321	-0,261676098	1,17922	-0,367778724
Kabupaten Sukabumi	-0,15946	1,077708	-1,02529	-1,00862	-0,475343072	0,068473	-0,694939566
Kabupaten Cianjur	0,449602	1,073552	-0,94861	-1,3305	0,68914194	-0,45128	-0,696129242
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
Kota Cimahi	1,3947	-1,34956	1,686532	1,310187	-1,180444089	0,185053	2,319302329
Kota Tasikmalaya	-0,83573	-0,0341	0,550214	0,295841	1,618593281	-0,73625	-0,045971423
Kota Banjar	-1,36499	0,277624	0,062225	0,012078	-0,831454697	-1,19124	-0,492892961

Cross Validation

Pada tahapan cross validation data akan dibagi menjadi dua bagian yaitu 80% untuk data training dan 20% untuk data testing. Tahapan ini menggunakan k-fold cross validation dengan $k = 5$. Pada cross validation akan dicari nilai α dan λ yang paling optimum yang akan digunakan sebagai penalti dalam *Elastic-net* dengan pertimbangan nilai RMSE yang paling kecil. Berikut

merupakan hasil dari cross validation:

Tabel 3. Hasil Cross Validation

α	λ	RMSE
0,10	0,00122	0,8736286
0,10	0,01218	0,8318183
0,10	0,12179	0,7888864
0,55	0,00122	0,8717983
0,55	0,01218	0,8174044
0,55	0,12179	0,8139015
1,00	0,00122	0,8715519
1,00	0,01218	0,8051218
1,00	0,12179	0,8581309

Berdasarkan Tabel 3 nilai RMSE yang terkecil yaitu sebesar 0,7888864 sehingga nilai α yang paling optimum yang akan digunakan sebagai penalti pada regresi *Elastic-net* yaitu sebesar 0,10 sementara nilai λ optimum yang diperoleh yaitu sebesar 0,121790470.

Regresi *Elastic-net*

Pendugaan parameter koefisien regresi *elastic-net* diperoleh menggunakan bantuan software RStudio dengan package *glmnet* yang memiliki kemampuan dalam pemilihan dan prediksi peubah dari kumpulan data berdimensi tinggi serta mampu menerima prediksi koefisien bias dengan varian yang lebih rendah sehingga mendapat akurasi prediksi yang lebih baik. taksiran koefisien regresi sebagai berikut:

Tabel 4. Regresi *Elastic-net*

Peubah	Taksiran Koefisien
(Intercept)	0,10799765
X_1	0,21939457
X_2	0,09991640
X_3	0
X_4	0,48823106
X_5	0,73204560
X_7	0,51813083

Berdasarkan Tabel 4 diketahui bahwa peubah bebas indeks pembangunan manusia (X_3) memiliki nilai taksiran nol, artinya peubah ini tidak memberikan pengaruh terhadap tingkat pengangguran terbuka sehingga tidak akan diikutsertakan ke dalam model. Sedangkan peubah bebas yang memiliki pengaruh terhadap tingkat pengangguran terbuka adalah dependency ratio (X_1), rata-rata lama sekolah (X_2), presentase penduduk miskin (X_4), upah minimum kabupaten/kota (X_5), dan kepadatan penduduk (X_7). Maka model regresi *Elastic-net* yang diperoleh adalah sebagai berikut:

$$Y = 0,10799765 + 0,21939457 X_1 + 0,09991640 X_2 + 0,48823106 X_4 + 0,73204560 X_5 + 0,51813083 X_7$$

Untuk mengetahui faktor yang memberikan pengaruh paling besar terhadap tingkat pengangguran terbuka dapat dilihat melalui perbandingan tingkat kepentingan peubah atau importance variable yang diperoleh menggunakan bantuan software RStudio dengan syntax “varImp()”, nilai importance variable tersaji pada Tabel 5 sebagai berikut:

Tabel 5. Tingkat Kepentingan Peubah

Peubah	importance variable
Upah Minimum Kabupaten/Kota	100,000
Kepadatan Penduduk	68,130
Presentase Penduduk Miskin	63,638
Dependency Ratio	23,534
Rata-rata Lama Sekolah	5,799
Indeks Pembangunan Manusia	0,000

Upah minimum kabupaten/kota merupakan peubah yang memberikan pengaruh paling besar terhadap tingkat pengangguran terbuka di Provinsi Jawa Barat. Sedangkan peubah bebas yang dianggap tidak memiliki pengaruh terlalu penting terhadap tingkat pengangguran terbuka di Provinsi Jawa Barat adalah indeks pembangunan manusia, hal ini dapat dilihat dari tingkat kepentingannya yang bernilai nol.

Ukuran Kebaikan Model

Ukuran kebaikan model diperoleh dari nilai koefisien determinasi (R^2) pada data testing. Semakin besar nilai R^2 maka kemampuan model dalam menjelaskan peubah tak bebas akan semakin baik. Berdasarkan Persamaan (2.9) nilai koefisien determinasi (R^2) yang dihasilkan yaitu sebesar 50,45% artinya model regresi *Elastic-net* mampu menjelaskan data sebesar 50,45% sedangkan sisanya 49,55% dijelaskan oleh peubah lainnya diluar model.

D. Kesimpulan

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

1. Berdasarkan hasil pengujian dari tujuh peubah bebas hanya terdapat lima peubah bebas yang memberikan pengaruh terhadap tingkat pengangguran terbuka di Jawa Barat yaitu dependency ratio (X_1), rata-rata lama sekolah (X_2), presentase penduduk miskin (X_4), upah minimum kabupaten/kota (X_5), dan kepadatan penduduk (X_7). Sedangkan peubah bebas yang tidak memberikan pengaruh terhadap tingkat pengangguran terbuka yaitu indeks pembangunan manusia (X_3). Sehingga model regresi yang diperoleh yaitu sebagai berikut:

$$Y = 0,10799765 + 0,21939457 X_1 + 0,09991640 X_2 + 0,48823106 X_4 + 0,73204560 X_5 + 0,51813083 X_7$$
2. Faktor yang memberikan pengaruh paling signifikan terhadap tingkat pengangguran terbuka di Provinsi Jawa Barat yaitu upah minimum kabupaten/kota (X_5).

Acknowledge

Terima kasih kepada seluruh pihak yang membantu memberi saran maupun masukan sampai terlaksananya penelitian ini.

Daftar Pustaka

- [1] 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.
- [2] 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.
- [3] Zou, H. & T, Hastie. (2005). Regularization and variable selection via the *Elastic-net*. *Journal of the Royal Statisical Sociecty*. Series **B**.
- [4] Febyanti, Fifia. (2022). *Pemodelan Faktor-Faktor yang Mempengaruhi Harga Rumah di Jabodetabek Menggunakan Metode Regresi Probit*, *Jurnal Riset Statistika*, 2(1), 50-56.