

Pendugaan Produktivitas Penangkapan Bagan Perahu dengan Regresi Gulud, LASSO dan *Elastic-net*

Resty Fanny*, Anik Djuraidah*, Aam Alamudi*,

*Departemen Statistika Institut Pertanian Bogor

Abstrak—Analisis regresi merupakan suatu teknik statistika untuk memeriksa dan memodelkan hubungan antara peubah bebas dengan peubah tak bebas. Regresi linear berganda melibatkan lebih dari satu peubah bebas. Multikolinearitas dalam regresi linear berganda terjadi bila antara peubah bebas terjadi korelasi. Multikolinearitas menyebabkan penduga dengan metode kuadrat terkecil (MKT) menjadi tidak stabil dan menghasilkan ragam besar. Multikolinearitas dapat diatasi dengan penambahan penalti penyusutan koefisien regresi. Parameter penyusutan diperoleh menggunakan validasi silang. Tujuan dari penelitian ini adalah untuk melakukan pemodelan menggunakan regresi gulud, LASSO, dan *elastic-net*. Data yang digunakan merupakan data hasil tangkapan nelayan di Pantai Carocok Tarusan Sumatera Barat sebagai peubah tak bebas dan jumlah tenaga kerja, jumlah bahan bakar, volume jaring/waring bagan perahu, jumlah jam tangkapan, ukuran kapal, jumlah watt perahu, pengalaman melaut, pendidikan dan umur sebagai peubah bebas. Model terbaik diberikan oleh LASSO berdasarkan nilai RMSEP validasi terkecil dibandingkan regresi gulud dan *elastic-net*. LASSO melakukan penyusutan pada peubah bebas jumlah tenaga kerja, jumlah bahan bakar dan jumlah watt perahu tepat nol sehingga peubah bebas (faktor produktivitas) yang mempengaruhi peubah tak bebas (perubahan produktivitas) yaitu volume jaring/waring dan ukuran perahu yang digunakan nelayan.

Kata kunci—*elastic-net*; LASSO; MKT; multikolinearitas; regresi gulud; regresi linier berganda

I. PENDAHULUAN

A. Latar Belakang

Regresi linear digunakan untuk mempelajari hubungan antara sebuah peubah tak bebas dan satu atau lebih peubah bebas (Greene (2012)). Penduga parameter regresi linear sering kali dilakukan dengan menggunakan metode kuadrat terkecil (MKT). Multikolinearitas pada peubah bebas dapat mengakibatkan ragam penduga menjadi besar (Mela and

Kopalle (2002)). Penanganan kasus multikolinearitas dapat dilakukan dengan beberapa metode, diantaranya dengan regresi komponen utama, regresi bertatar, regresi gulud, Least Absolute Shrinkage and Selection Operator (LASSO) dan regularisasi *elastic-net*.

Regresi gulud yang diperkenalkan pertama kali oleh Hoerl pada tahun 1962 untuk mengendalikan ketidakstabilan penduga MKT. Penduga yang dihasilkan dari regresi gulud adalah bias namun mempunyai kuadrat tengah galat yang lebih kecil daripada MKT. Peubah bebas dan peubah tak bebas dalam regresi gulud menggunakan peubah-peubah yang sudah dibakukan.

LASSO dikemukakan oleh Tibshirani (1996) pertama kali sebagai alternatif dari solusi permasalahan penduga parameter model MKT yang kurang akurat dalam prediksi dan sulit diinterpretasi. Berbeda dengan regresi gulud, LASSO menyusutkan koefisien regresi tepat nol sekaligus dapat digunakan untuk seleksi peubah bebas sehingga hanya peubah bebas yang penting saja yang masuk dalam model regresi. LASSO melakukan penyusutan nilai parameter sehingga ragam sisaan yang dihasilkan menjadi kecil (Tibshirani (1996)).

Elastic-net merupakan regularisasi penalti gabungan antara regresi gulud dan LASSO. Regularisasi ini menangani beberapa kekurangan LASSO. Serupa dengan LASSO, *elastic-net* melakukan seleksi peubah secara simultan, penyusutan, dan dapat memilih kelompok peubah yang berkorelasi (Zou and Hastie (2005)).

Putra (2017) melakukan penelitian mengenai hasil tangkapan nelayan di Pantai Carocok Tarusan Sumatera Barat. Putra (2017) menduga sembiilan peubah bebas yang berpengaruh terhadap hasil

tangkapan nelayan yaitu, jumlah tenaga kerja, jumlah bahan bakar, volume jaring/waring bagan perahu, jumlah jam tangkapan, ukuran kapal, jumlah watt lampu, pengalaman melaut, pendidikan dan umur. Pada peubah bebas tersebut terdapat multikolinearitas yang belum diatasi peneliti. Pada penelitian ini multikolinearitas tersebut akan diatasi dengan menambahkan penalti penyusutan koefisien pendugaan parameter regresi dengan regresi gulud, LASSO, dan *elastic-net*.

B. Tujuan

- 1) Menentukan model regresi gulud, LASSO dan *elastic-net* untuk pendugaan produktivitas penangkapan bagan perahu
- 2) Menentukan model terbaik yang dapat digunakan untuk pendugaan produktivitas penangkapan bagan perahu.
- 3) Menentukan faktor-faktor yang memengaruhi produktivitas penangkapan bagan perahu.

II. TINJAUAN PUSTAKA

A. Analisis Regresi Linier

Analisis regresi merupakan suatu teknik statistika untuk memeriksa dan memodelkan hubungan antar peubah (Montgomery and Peck (1992)). Regresi merupakan tempat kedudukan nilai tengah dari peubah Y untuk berbagai nilai atau selang nilai peubah X, serta merupakan usaha untuk mengepas suatu fungsi atau kurva terhadap pencaran titik-titik pada sumbu X-Y (Aunuddin (2005)). Dalam hal ini, Y adalah peubah tak bebas, sedangkan X adalah peubah bebas.

Regresi linier sederhana hanya melibatkan satu peubah bebas, sedangkan regresi linier berganda melibatkan p peubah bebas. Regresi linier sederhana memiliki beberapa asumsi, yaitu nilai harapan rata-rata sisaan sama dengan nol, ragam sisaan homogen, sisaan saling bebas, sisaan menyebar normal dengan rata-rata nol dan ragam (σ^2), serta sisaan bebas terhadap peubah bebas. Pada regresi linier berganda terdapat asumsi tambahan bahwa tidak ada multikolinearitas pada peubah bebas.

Model linier artinya linier dalam parameter (Draper dan Smith 1992). Misalkan terdapat vektor peubah bebas $x^T = (x_1, x_2, \dots, x_p)$ dan digunakan untuk menduga luaran nilai y yang berupa bilangan

riil, maka model regresi linier memiliki bentuk sebagai berikut:

$$y_i = \beta_0 + \sum_{j=1}^p x_{ij}\beta_j + \varepsilon_i \quad (1)$$

Keterangan :

- y_i : peubah tak bebas berukuran $n \times 1$
- β_0 : intersep
- x_{ij} : matriks peubah bebas berukuran $n \times (p+1)$
- β_j : slope atau kemiringan

B. Regularisasi Pada Model Regresi

1) *Regresi Gulud*: Regresi gulud yang diperkenalkan pertama kali oleh Hoerl pada tahun 1962 untuk mengendalikan ketidakstabilan penduga MKT. Regresi gulud serupa dengan MKT yang meminimumkan jumlah kuadrat sisaan (JKS) pada pendugaan koefisien regresi. Namun, dalam metode ini menambahkan penalti penyusutan dalam meminimumkan JKS. Pendugaan koefisien regresi gulud dilakukan dengan meminimumkan:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p \beta_j^2 = JKS + \sum_{j=1}^p \beta_j^2 \quad (2)$$

dengan $\lambda \geq 0$ adalah parameter penyusutan yang ditentukan secara terpisah. Persamaan (2) dibagi menjadi dua bagian yang berbeda. Bagian pertama sebagai kuadrat terkecil, regresi gulud mencari penduga koefisien yang mengepas data dengan baik, dengan meminimumkan JKS. Bagian kedua, $\sum_{j=1}^p \beta_j^2$ disebut penalti penyusutan yang bernilai kecil ketika β_1, \dots, β_p mendekati nol dan memberikan efek penyusutan koefisien β_j terhadap nol. Parameter penyusutan (λ) memberikan kontrol terhadap penyusutan koefisien regresi (James et al. (2013)).

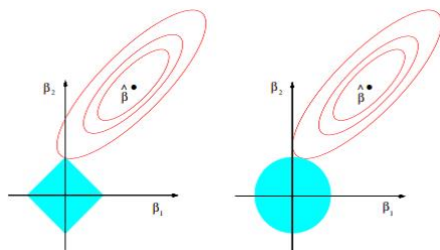
Jika $\lambda = 0$, maka penalti penyusutan tidak memberikan pengaruh apapun sehingga regresi gulud akan menghasilkan MKT. Namun, jika $\lambda \rightarrow \infty$ akan berdampak pada penalti penyusutan yang semakin besar dan koefisien pendugaan yang semakin mendekati nol. Pada MKT hanya dihasilkan satu set dugaan koefisien, sedangkan pada regresi gulud dihasilkan set dugaan koefisien berbeda-beda untuk setiap nilai λ yang dinotasikan dengan $(\hat{\beta}_\lambda^R)$. Parameter penyusutan (λ) yang optimal ditentukan dengan

metode validasi silang. Penyusutan hanya ditujukan untuk peubah bebas, tidak pada intersep. Regresi gulud diterapkan pada data yang dibakukan. Nilai dugaan pada intersep $\beta_0 = \bar{y} = \sum_{i=1}^n \frac{y_i}{n}$ (James et al. (2013)).

2) *LASSO: Least Absolute Shrinkage and Selection Operator* (LASSO) adalah salah satu metode penyusutan seperti gulud yang dapat mengatasi permasalahan multikolinearitas yang diperkenalkan oleh (Tibshirani (1996)). Koefisien LASSO, $(\hat{\beta}_\lambda^L)$ meminimumkan persamaan berikut:

$$\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})^2 + \lambda \sum_{j=1}^p |\beta_j| = JKS + \sum_{j=1}^p |\beta_j| \quad (3)$$

LASSO dan regresi gulud memiliki formula yang hampir sama, hanya penaltinya saja yang berbeda. Penalti penyusutan pada regresi gulud diganti dengan $|\beta_j|$ sebagai penalti penyusutan LASSO. LASSO sering disebut dengan menggunakan penalti L_1 dan regresi gulud sering disebut dengan penalti L_2 . Penalti L_1 dari vektor koefisien β adalah $\|\beta\|_1 = \sum |\beta_j|$ (James et al. (2013)). LASSO menyusutkan dugaan koefisien mendekati nol. Pada LASSO penalti L_1 memberikan efek beberapa dugaan koefisien menjadi tepat nol ketika λ cukup besar. Hal tersebut menyebabkan LASSO dapat menghasilkan model dengan peubah penjelas yang lebih sedikit.



Gambar 1. Pendugaan untuk LASSO (kiri) dan gulud (kanan) Sumber: Hastie et al. (2008)

Gambar 1 memberikan ilustrasi tentang penalti LASSO dalam membuat nilai koefisien menjadi 0 pada dua peubah penjelas. $\hat{\beta}$ adalah nilai dugaan dari MKT dan garis elips berwarna merah adalah sisaan dari MKT. Daerah penalti untuk LASSO $|\beta_1| + |\beta_2| \leq s$ memiliki bentuk belah ketupat,

sedangkan daerah penalti untuk gulud $\beta_1^2 + \beta_2^2 \leq s$ memiliki bentuk lingkaran. Nilai s memberikan dugaan koefisien LASSO yang sama untuk setiap nilai λ . Sama halnya dengan LASSO, untuk setiap nilai λ , terdapat nilai s yang menghasilkan dugaan koefisien yang sama. Belah ketupat memiliki sudut, maka ketika elips menyentuh sudut tersebut berarti salah satu koefisien bernilai 0. Lingkaran tidak memiliki sudut, sehingga elips menjadi lebih sulit bersinggungan dengan daerah kendala pada titik 0. Ketika $p > 2$, maka kendala LASSO memiliki sudut lebih banyak sehingga peluang suatu koefisien bernilai 0 lebih besar (Hastie et al. (2008)).

3) *Elastic-net*: Regularisasi LASSO seperti yang telah dipaparkan memiliki beberapa kekurangan, antara lain (Zou and Hastie (2005)):

- 1) Ketika $p \ll n$, maka LASSO hanya memilih n peubah yang diikuti dalam model.
- 2) Jika ada sekumpulan peubah dengan korelasi tinggi, maka LASSO hanya sembarang memilih salah satu peubah saja.
- 3) Ketika $p \gg n$, kinerja LASSO lebih didominasi oleh gulud.

Zou dan Hastie (2005) memperkenalkan penalti *elastic-net* sebagai berikut :

$$\sum_{j=1}^p [\alpha |\beta_j| + (1 - \alpha) \beta_j^2] \quad (4)$$

Jika $\alpha = 0$, maka *elastic-net* menjadi penalti gulud dan jika $\alpha = 1$ maka *elastic-net* menjadi penalti LASSO. Pada regularisasi *elastic-net* terdapat penyusutan koefisien bersama dari peubah-peubah penjelas yang berkorelasi seperti gulud dan seleksi peubah seperti LASSO (Hastie et al. 2008).

Penduga koefisien pada regularisasi *elastic-net* adalah (James et al. (2013)):

$$\hat{\beta} = [\arg \min_{\beta_0, \beta} (\sum_{i=1}^n (y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij})) + \quad (5)$$

$$\lambda \sum_{j=1}^p (\alpha |\beta_j| + (1 - \alpha) \beta_j^2)$$

4) *Validasi Silang*: Validasi silang (Cross Validation-CV) merupakan metode paling sederhana dan banyak dipakai secara luas untuk menduga galat prediksi. Idealnya, ketika data yang dimiliki memadai, akan dapat ditentukan suatu anak gugus data validasi dan digunakan untuk mengukur

ketepatan model yang dimiliki. Namun sering kali data yang dimiliki terlalu sedikit sehingga tidak memungkinkan untuk dilakukan validasi secara langsung. Solusinya adalah melakukan validasi silang yang menggunakan sebagian data yang tersedia untuk mengepas model dan sebagian data yang lain digunakan sebagai data pengujian model (Hastie et al. (2008)). Pada CV, data dibagi menjadi dua kelompok, yaitu data pemodelan dan data validasi. Pengelompokan data pemodelan dan data validasi dilakukan secara acak, dan setiap pengamatan memiliki peluang yang sama menjadi data validasi (James et al. (2013)).

Salah satu jenis validasi silang adalah validasi silang lipat-k. Metode ini baik digunakan ketika jumlah data amatan sedikit. Dalam validasi silang lipat-k, amatan dibagi ke dalam k anak gugus data secara acak. Amatan sebanyak n k digunakan sebagai data pemodelan dan sebanyak k amatan sebagai data validasi. Proses validasi silang dilakukan berulang sampai k kali, dengan masing-masing k anak gugus data digunakan satu kali sebagai validasi model. Proses ini menghasilkan k dugaan kuadrat tengah galat (mean square error-MSE), $MSE_1, MSE_2, \dots, MSE_k$. Dugaan galat validasi silang (cross validation error-CVE) lipat-k diperoleh dengan menghitung rata-rata dari,

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

Validasi silang lipat-k sering kali menggunakan $k = 5$ atau $k = 10$ dibandingkan nilai $k = n$, *leave-one-out cross validation* (LOOCV). Hal ini dikarenakan lebih menguntungkan dalam proses komputasi. LOOCV membutuhkan waktu yang lebih lama yang berpotensi menyulitkan proses komputasi (James et al. (2013)).

III. METODOLOGI

A. Data

Peubah tak bebas (Y) dalam penelitian ini adalah jumlah hasil tangkapan nelayan bagan perahu per trip (kg) yang disebut juga sebagai perubahan produktivitas penangkapan bagan perahu. Putra (2017) menduga sembilan peubah bebas yang berpengaruh terhadap peubah tak bebas yang disebut juga faktor-faktor yang memengaruhi produktivitas (X)

(Tabel I).

Pengumpulan data dilaksanakan pada bulan April-Mei 2017 di Pelabuhan Perikanan Pantai Carocok Tarusan Kabupaten Pesisir Selatan, Sumatera Barat. Pengumpulan data dilakukan dengan metode survei. Responden pada penelitian ini adalah nelayan bagan perahu sebanyak 92 orang.

Tabel I
DAFTAR PEUBAH YANG DIGUNAKAN

Peubah	Keterangan	Peubah	Keterangan
X_1	jumlah tenaga kerja	X_6	jumlah watt perahu
X_2	jumlah bahan bakar	X_7	pengalaman melaut
X_3	volume jaring/waring	X_8	pendidikan
X_4	bagan bagan perahu	X_9	umur
X_5	jumlah jam tangkapan		
	ukuran perahu		

B. Metode

Metode analisis yang digunakan dalam penelitian ini dapat dijelaskan secara rinci sebagai berikut:

- 1) Melakukan eksplorasi data dengan membuat plot tebaran antara peubah tak bebas (Y) dengan masing-masing peubah bebas (X) untuk mengetahui pola hubungan, serta korelasi antar peubah. Hasil plot tebaran menyatakan bahwa terdapat lima peubah bebas (X) yang memiliki pengaruh linier terhadap peubah tak bebas (Y) yaitu X_1, X_2, X_3, X_5 , dan X_6 sehingga hanya 5 peubah yang akan dianalisis. X_6 menggambarkan jumlah watt bagan perahu. Pada plot tebaran jumlah watt perahu (X_6) terbagi menjadi dua kelompok sehingga dalam analisis peubah ini diubah menjadi peubah *dummy* dengan jumlah watt perahu kurang dari 10000 diberi skor 0 dan jumlah watt perahu lebih dari 10000 diberi skor 1.
- 2) Melakukan pemodelan dengan MKT, yaitu meregresikan peubah Y dengan peubah X_1, X_2, X_3, X_5 , dan X_6 .
- 3) Mendeteksi adanya multikolinearitas dengan melihat nilai VIF (*Variance Inflation Factor*) dari masing-masing peubah bebas (X).
- 4) Membagi data menjadi 2 bagian, 70 % data pemodelan dan 30 % data validasi.
- 5) Melakukan analisis regresi gulud.
 - a) Memilih nilai λ optimal dengan validasi silang.

- b Analisis regresi menggunakan regresi gulud dengan λ optimal.
- 6) Melakukan analisis regresi dengan metode LASSO.
 - a Memilih nilai λ optimal dengan validasi silang.
 - b Analisis regresi menggunakan metode LASSO dengan λ optimal.
- 7) Melakukan analisis regresi dengan regresi *elastic-net*.
 - Menentukan rentang nilai α yaitu $0 \leq \alpha \leq 1$ dengan validasi silang.
 - Memilih nilai pasangan nilai α dan λ optimal yang menghasilkan CVE minimum.
 - Analisis regresi menggunakan metode regresi *elastic-net* dengan pasangan α dan λ optimal.
- 8) Mengukur kebaikan model dengan menghitung RMSEP pada data validasi menggunakan formula :

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

Model terbaik adalah model yang memiliki nilai RMSEP validasi terkecil.

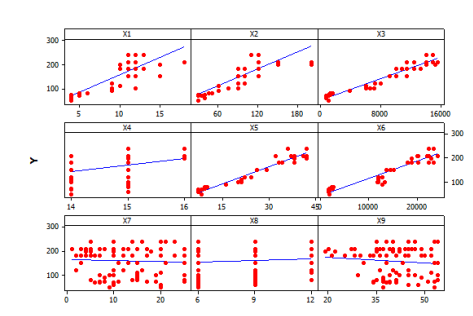
- 9) Pendugaan produktivitas bagan perahu berdasarkan nilai koefisien masing-masing analisis regresi serta identifikasi peubah yang berpengaruh terhadap hasil tangkapan nelayan.

IV. HASIL DAN PEMBAHASAN

A. Eksplorasi Data

Eksplorasi data dilakukan dengan membuat plot antara peubah tak bebas (Y) dengan setiap peubah bebasnya (X). Berdasarkan plot tebaran peubah, tidak semua X yang memiliki hubungan linier dengan Y (Gambar 2). Secara eksploratif X yang memiliki hubungan linear dengan Y adalah jumlah tenaga kerja (X_1), jumlah bahan bakar (X_2), volume jaring/waring bagan perahu (X_3), ukuran perahu (X_5) dan jumlah watt lampu (X_6) sedangkan X yang tidak memiliki hubungan linear dengan Y adalah jumlah jam tangkapan (X_4), pengalaman melaut (X_7), pendidikan (X_8) dan umur (X_9). Data pengalaman melaut (X_7), pendidikan (X_8) dan umur (X_9) merupakan peubah bebas yang ditujukan

kepada pemimpin dalam bagan perahu yang digunakan.



Gambar 2. Plot peubah tak bebas (Y) dengan masing-masing peubah bebas (X)

B. Regresi Linier Berganda

Analisis regresi linear berganda dengan menggunakan MKT menghasilkan model dengan F-hitung sebesar 266.06 dan nilai-p sebesar 0.000. Hal ini menunjukkan bahwa minimal terdapat satu peubah bebas yang memengaruhi hasil tangkapan nelayan. Model regresi yang dihasilkan adalah sebagai berikut:

$$Y = 6.4 + 1.75X_1 - 0.245X_2 + 0.0074X_3 + 2.84X_5 + 33.2X_6$$

Pemeriksaan multikolinearitas pada hasil MKT dapat dilakukan dengan melihat nilai VIF dari masing-masing peubah bebas (Tabel II) . Peubah dengan nilai VIF >10 mengindikasikan adanya multikolinieritas. Multikolinieritas terjadi pada peubah bebas X_3 dan X_5 .

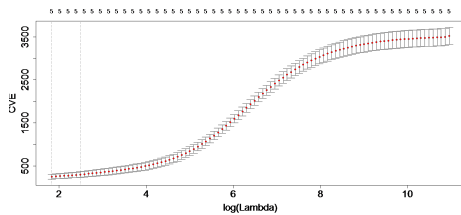
Tabel II
KOEFSISIEN REGRESI HASIL MKT

Peubah	koefisien MKT	nilai-p	VIF
β_0	6.400	0.668	-
X_1	1.750	0.206	7.815
X_2	-0.245	0.017	7.229
X_3	0.007	0.002	57.304
X_5	2.840	0.002	46.451
X_6	33.200	0.000	4.887

C. Regresi dengan penalti

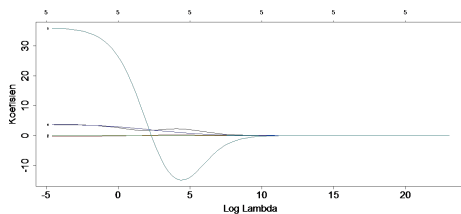
1) *Regresi Gulud*: Parameter gulud optimum diperoleh dengan metode validasi silang (cross validation-CV), yang menghasilkan CVE minimum.

CV dilakukan dengan membagi data menjadi 2 bagian yaitu data pemodelan dan data validasi. Pembagian data pemodelan sebanyak 64 amatan dan data validasi sebanyak 28 amatan dilakukan secara acak pada data yang tersedia, sebanyak 20 kali ulangan. Pemilihan model untuk gulud menggunakan CVE minimum yang pertama. Gambar 3 menunjukkan nilai CVE untuk setiap log (λ).



Gambar 3. Validasi silang parameter gulud (λ)

Dua garis vertikal pada CV menggambarkan nilai λ optimum dengan CVE minimum. Garis vertikal pertama menyatakan nilai log λ minimum (λ_{min}) dan garis vertikal kedua menyatakan nilai log λ terbesar dalam satu kesalahan standar λ minimum (λ_{1se}). Namun, nilai λ_{1se} lebih direkomendasikan dijadikan λ optimum karena lindung nilai terhadap *overfitting* dengan memilih nilai λ lebih besar. λ_{1se} adalah sebesar 12.011 dan log λ_{1se} adalah sebesar 2.468. RMSEP pemodelan gulud sebesar 16.758 dan RMSEP validasi sebesar 22.188.



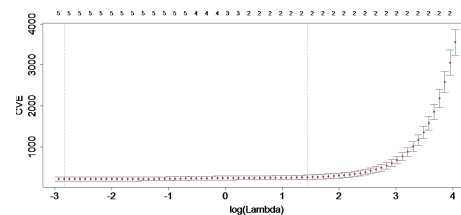
Gambar 4. Plot koefisien gulud

Plot koefisien gulud untuk berbagai nilai λ terdapat pada Gambar 4. Koefisien-koefisien gulud tersebut sudah sangat mendekati nol saat nilai log λ mendekati 12 tetapi memiliki nilai CVE yang tinggi sehingga perlu adanya pengurangan nilai λ hingga diperoleh CVE minimum (λ optimal). Angka 5 pada setiap nilai λ menyatakan bahwa nilai koefisien bukan nol pada nilai λ tersebut sebanyak 5 koefisien. Hal ini sesuai dengan prinsip dari regresi gulud yang menyusutkan koefisien hanya mendekati nol.

Model regresi gulud yang dihasilkan menggunakan λ optimum adalah sebagai berikut:

$$Y = 40.254 + 1.138X_1 + 0.125X_2 + 0.004X_3 + 2.84X_5 + 33.2X_6$$

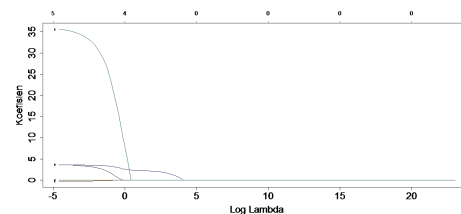
2) *LASSO*: Parameter penyusutan (λ) *LASSO* diperoleh menggunakan validasi silang (*cross validation-CV*). (λ) optimal dipilih berdasarkan nilai CVE minimum. Proses CV pada metode ini sama dengan CV pada regresi gulud. hasil CV dapat dilihat pada Gambar 5. Nilai λ optimal yang dipilih adalah λ_{1se} sebesar 4.217 dan log λ_{1se} sebesar 1.639. Nilai RMSEP pemodelan *LASSO* sebesar 15.451 dan RMSEP validasi sebesar 19.739. Pengulangan sebanyak 20 kali menghasilkan nilai RMSEP yang hampir sama baik pemodelan maupun validasi sehingga model yang dibangun bisa dikatakan stabil.



Gambar 5. Validasi silang parameter LASSO (λ)

Pemodelan regresi *LASSO* menggunakan λ_{1se} sebagai λ optimal. Model *LASSO* yang dihasilkan adalah sebagai berikut:

$$Y = 47.318 + 0.005X_3 + 2.179X_5$$

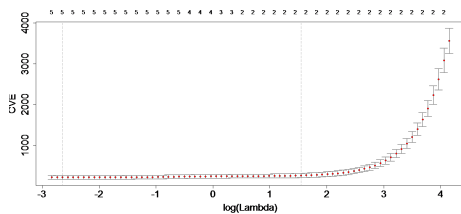


Gambar 6. Plot koefisien LASSO

Koefisien regresi pada model regresi *LASSO* cenderung lebih kecil dibandingkan dengan dugaan koefisien regresi pada model *MKT*. *LASSO* melakukan seleksi peubah bebas dengan menyusutkan koefisien regresi menjadi tepat nol. Plot koefisien *LASSO* terdapat pada Gambar 6.

Peubah bebas yang mengalami penyusutan dari 5 peubah bebas yang dianalisis adalah sebanyak 3 peubah bebas yaitu X_1, X_2 dan X_6 .

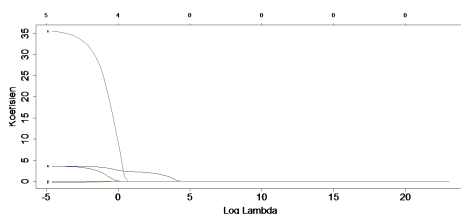
3) *Elastic-net*: Regresi *elastic-net* menggunakan dua penalti penyusutan, yaitu penalti LASSO $\sum |\beta_j|$ dan penalti gulud $\sum \beta_j^2$. Parameter untuk penalti LASSO adalah $\lambda \times \alpha$ sedangkan parameter untuk penalti gulud adalah $\lambda \times (1 - \alpha)$, dengan $\alpha \in (0, 1)$. Nilai α yang digunakan adalah $0 \leq \alpha \leq 1$. Pemilihan nilai α dan λ optimal dengan CV yang menghasilkan CVE minimum. Nilai α optimal adalah sebesar 0.9 dengan λ_{1se} sebesar 5.644 serta $\log \lambda_{1se}$ sebesar 1.731. RMSEP pemodelan *elastic-net* sebesar 15.792 dan RMSEP validasi sebesar 20.178. Hasil CV terdapat pada Gambar 7.



Gambar 7. Validasi silang parameter *elastic-net*

Hasil CV menunjukkan bahwa nilai parameter untuk penalti LASSO lebih besar daripada nilai parameter untuk penalti gulud. Nilai RMSEP yang diperoleh dari model LASSO juga lebih kecil daripada RMSEP pemodelan regresi gulud. Koefisien peubah-peubah bebas *elastic-net* yang dihasilkan dapat dilihat pada Gambar 8. Serupa dengan LASSO, peubah bebas yang terseleksi dengan metode *elastic-net* ini adalah peubah bebas X_1, X_2 dan X_6 . Model *elastic-net* menggunakan pasangan α dan λ optimum adalah sebagai berikut:

$$Y = 49.735 + 0.005X_3 + 2.129X_5$$



Gambar 8. Plot koefisien *elastic-net*

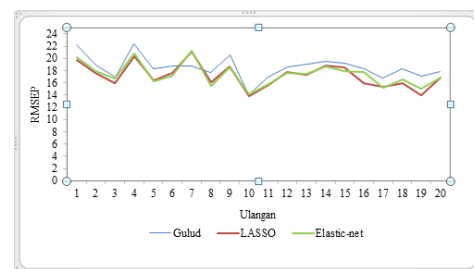
D. Pemilihan Model Terbaik

Metode penyusutan koefisien regresi yaitu regresi gulud, LASSO dan *elastic-net* memiliki nilai-nilai koefisien yang lebih kecil dibandingkan nilai koefisien MKT. Regresi gulud hanya menyusutkan koefisien hasil MKT kearah nol, sedangkan LASSO dan *elastic-net* dapat menyusutkan koefisien hasil MKT sampai tepat nol sehingga secara otomatis peubah-peubah bebas tersebut akan terseleksi dari model yang terpilih. Peubah-peubah bebas yang terseleksi dari model LASSO dan *elastic-net* merupakan peubah bebas yang tidak berpengaruh nyata terhadap peubah tak bebas (Y). Perbandingan dugaan koefisien regresi hasil analisis dengan MKT, regresi gulud, LASSO dan *elastic-net* dapat terlihat pada Tabel III.

Tabel III
KOEFSIEN REGRESI HASIL MKT, GULUD, LASSO DAN *elastic-net*

Peubah	MKT	Gulud	LASSO	<i>Elastic-net</i>
β_0	6.400	40.254	47.318	49.375
X_1	1.750	1.138	0.000	0.000
X_2	-0.245	0.125	0.000	0.000
X_3	0.007	0.004	0.005	0.005
X_5	2.840	1.892	2.179	2.129
X_6	33.200	0.0021	0.000	0.000

Pemilihan model terbaik dari model regresi gulud, LASSO dan *elastic-net* berdasarkan nilai RMSEP pada data validasi. Perbandingan RMSEP validasi dari ketiga model disajikan pada Gambar 9. Pengulangan sebanyak 20 kali menghasilkan rata-rata nilai RMSEP validasi dengan metode LASSO lebih kecil dibandingkan nilai RMSEP validasi regresi gulud dan *elastic-net* sehingga LASSO lebih baik dibandingkan dua metode lainnya dalam menyusutkan koefisien regresi.



Gambar 9. Nilai RMSEP validasi model regresi

Model LASSO menyatakan bahwa dengan penambahan volume jaring/waring bagan perahu sebesar 1 m^3 akan meningkatkan hasil tangkapan nelayan sebesar 0.005 kg dan dengan penambahan ukuran perahu sebesar 1 GT akan meningkatkan hasil tangkapan nelayan sebesar 2.179 kg.

E. Simpulan dan saran

1) *Simpulan*: Pendugaan produktivitas bagan perahu di Pantai Carocok Tarusan Sumatera Barat menggunakan regresi gulud, LASSO dan *elastic-net* memberikan nilai-nilai koefisien yang lebih kecil dibandingkan koefisien pada metode kuadrat terkecil. Model terbaik diberikan oleh LASSO yang memiliki nilai RMSEP validasi terkecil dibandingkan regresi gulud dan *elastic-net*. Disimpulkan bahwa terdapat dua peubah bebas (faktor produktivitas) yang mempengaruhi hasil tangkapan nelayan (peubahan produktivitas) yaitu jumlah jaring/waring bagan perahu dan ukuran perahu yang digunakan oleh nelayan.

2) *Saran*: Saran untuk penelitian selanjutnya diharapkan dapat membandingkan metode gulud, LASSO dan *elastic-net* dengan menggunakan simulasi untuk mengetahui sifat-sifat dari ketiga metode tersebut.

DAFTAR PUSTAKA

- Aunuddin (2005). *Statistika: Rancangan dan Analisis Data*. Bogor: IPB Press.
- Greene, W. H. (2012). *Econometric Analysis Ed ke-7*. New York: Pearson.
- Hastie, T., R. Tibshirani, and J. Friedman (2008). *The Elements of Statistical Learning. Data Mining, Inference, and Prediction*. New York: Springer.
- James, G., D. Witten, and R. Tibshirani (2013). *An Introduction Learning with Applications in R*. New York: Springer.
- Mela, C. F. and P. K. Kopalle (2002). The impact of collinearity on regression analysis: the asymmetric effect of negative and positive correlation. *Applied Economics* 34, 667–677.
- Montgomery, D. C. and E. A. Peck (1992). *Introduction to Linear Regression Analysis 2nd Ed*. New York: Wiley.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society Series B* 58, 267–288.

Zou, H. and T. Hastie (2005). Regularization and variable selection via the *elastic-net*. *Journal of the Royal Statistical Society Series B* 67, 301–320.