

I PENDAHULUAN

1.1 Latar Belakang

Curah hujan merupakan faktor iklim yang dapat mempengaruhi kelangsungan hidup manusia khususnya pada sektor pertanian. Ketika curah hujan ekstrim, masalah seperti banjir dapat terjadi dan menimbulkan dampak buruk terhadap hasil pertanian. Informasi sederhana yang sebenarnya mudah dipahami oleh kebanyakan orang adalah prediksi apakah akan terjadi hujan ekstrim atau tidak. Jika telah diketahui akan terjadi hujan ekstrim, maka penerima informasi akan mengambil langkah dan keputusan yang tepat untuk menghindari masalah dan kerugian yang diakibatkan oleh hujan ekstrim. Indonesia memiliki curah hujan yang cukup ekstrim di beberapa daerah seperti daerah Jawa Barat, dimana kawasan utara Jawa Barat merupakan daerah dataran rendah sedangkan bagian selatan merupakan daerah dataran tinggi. Perbedaan ketinggian mengakibatkan perbedaan intensitas curah hujan (Manik *et al.* 2014), sehingga intensitas curah hujan ekstrim di dataran rendah akan berbeda dengan intensitas curah hujan ekstrim di dataran sedang ataupun dataran tinggi. Salah satu sebaran yang sering digunakan untuk pemodelan nilai ekstrim adalah Sebaran Pareto Terapat (*Generalized Pareto Distribution*, GPD). Manurung *et al.* (2018) memanfaatkan konsep metode *measure of surprise* untuk menduga ambang batas ekstrim berdasarkan sebaran GPD pada data curah hujan Bogor di beberapa wilayah yang berbeda dan menghasilkan nilai ambang batas ekstrim yang berbeda untuk masing-masing wilayah.

Informasi tentang curah hujan biasanya diperoleh dengan memanfaatkan data luaran *Global Circulation Models* (GCM). Namun data luaran GCM masih berskala global yang memiliki dimensi besar sehingga digunakanlah teknik *statistical downscaling* untuk menghubungkan data luaran GCM tersebut dengan data curah hujan yang berskala lokal menggunakan metode statistika. Penelitian Stephenson *et al.* (1999) mengatakan bahwa ketika curah hujan ekstrim maka sebaran peluang curah hujan akan jauh dari normal. Ketika data curah hujan yang digunakan tidak simetris atau menjulur, regresi linear sederhana kurang tepat digunakan karena nilai dugaannya sensitif terhadap pencilan. Regresi kuantil merupakan metode yang dapat digunakan untuk menganalisis data yang tidak simetris (Koenker dan Basset 1978). Regresi kuantil dapat mengukur efek peubah penjelas tidak hanya di pusat sebaran data, tetapi juga pada bagian atas dan bawah ekor sebaran. Yu dan Moyeed (2001) mempopulerkan metode Bayes pada regresi kuantil yang memanfaatkan informasi contoh dan sebaran prior untuk mendapatkan sebaran posterior. Penelitian tersebut mengatakan bahwa regresi kuantil Bayes berdasarkan sebaran *Asymmetric Laplace* sehingga untuk mendapatkan pendugaan paramater dari sebaran posterior menggunakan metode *Markov Chain Monte Carlo* (MCMC). Metode MCMC yang terkenal yaitu *Metropolis-Hasting* (Yu dan Moyeed 2001) dan *Gibbs sampling* (Kozumi dan Kobayashi 2009).

Penelitian regresi kuantil Bayes yang telah dilakukan sebelumnya menggunakan respon kontinu sebagai peubah responnya. Hendri *et al.* (2019) telah melakukan pendugaan curah hujan ekstrim di Jawa Barat menggunakan regresi kuantil Bayes dengan respon kontinu, tetapi hasil yang didapatkan adalah model untuk mendapatkan nilai dugaan curah hujan ekstrim. Agar dapat menghasilkan

informasi hujan akan ekstrim atau tidak di Jawa Barat, maka pada penelitian ini akan digunakan metode regresi kuantil biner dengan tambahan penalti LASSO melalui pendekatan Bayesian. Respon biner yang digunakan pada penelitian ini didapatkan dari pengelompokan curah hujan ekstrim menjadi curah hujan ekstrim ($y_i = 1$) dan tidak ekstrim ($y_i = 0$) berdasarkan ambang batas ekstrim yang didapatkan melalui metode *Measure of Surprise*. Diharapkan penelitian ini dapat memprediksi curah hujan ekstrim dengan akurasi yang baik sehingga membantu untuk menentukan langkah selanjutnya dan menghasilkan keputusan yang lebih tepat untuk mengurangi kerugian di sektor pertanian yang diakibatkan oleh curah hujan ekstrim.

4.2 Tujuan Penelitian

Tujuan penelitian ini adalah pemodelan *statistical downscaling* untuk memprediksi curah hujan ekstrim di Jawa Barat menggunakan metode regresi kuantil biner Bayes terpenalti LASSO.

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

II TINJAUAN PUSTAKA

2.1 *Statistical Downscaling*

Global circulation model (GCM) merupakan data yang terdiri atas grid-grid yang mampu memberikan informasi tentang pergeseran iklim dari masa lampau sekarang dan di masa yang akan datang. GCM berskala besar (global) atau memiliki resolusi yang rendah sehingga belum memperhitungkan fenomena pada skala kecil (lokal). Kelemahan model GCM dalam melakukan pendugaan adalah tingkat akurasi prediksinya yang rendah. Kelemahan ini disebabkan oleh curah hujan yang ditentukan oleh suatu proses fisika yang sangat rumit, sensitif, dan nonlinier yang tidak dapat dimodelkan oleh GCM (Stockdale *et al.* 1998).

Statistical downscaling (SD) merupakan teknik yang dapat digunakan untuk menganalisis hubungan fungsional antara data berskala lokal (curah hujan) dan data berskala global (data luaran GCM) menggunakan metode statistika. Pemilihan peubah-peubah prediktor dan penentuan domain (lokasi dan jumlah grid) merupakan faktor kritis yang akan mempengaruhi kestabilan peramalan (Wilby dan Wigley 1997 dalam Wigena 2006). Menurut Busuioc *et al.* (2010), model SD dapat memberikan hasil yang baik jika memiliki: 1) hubungan yang erat antara peubah respon dengan peubah prediktor, sehingga keragaman iklim lokal dapat dijelaskan dengan baik, 2) peubah prediktor disimulasi baik oleh GCM, dan 3) hubungan antara respon dengan perubahan waktu dan tetap sama meskipun ada perubahan iklim.

Bentuk umum model SD, yaitu:

$$y_t = f(X_{t \times g})$$

dengan,

y_t : peubah iklim lokal (curah hujan)

$X_{t \times g}$: peubah luaran GCM (presipitasi)

t : banyaknya waktu (bulanan)

g : banyaknya *grid* domain GCM

Pada umumnya model SD melibatkan data deret waktu (t) dan data spasial GCM (g). Kompleksitas model ini terjadi karena X berkorelasi dan pengamatan peubah y berotokorelasi (Wigena 2006).

2.2 Regresi Kuantil

Regresi kuantil (RK) pertama kali dikemukakan oleh Koenker dan Bassett (1978). Metode ini merupakan suatu metode regresi dengan pendekatan memisahkan atau membagi data menjadi kuantil-kuantil tertentu yang kemungkinan memiliki nilai dugaan yang berbeda. Regresi kuantil dapat meminimumkan galat mutlak terboboti dan menduga model dengan menggunakan fungsi kuantil bersyarat pada suatu sebaran data. Keuntungan utama dari regresi kuantil adalah efisien jika sisaan tidak menyebar normal dan kekar terhadap adanya pencilan. Regresi kuantil tidak membutuhkan asumsi parametrik dan sangat bermanfaat untuk menganalisis bagian tertentu dari suatu sebaran bersyarat (Buhai 2004).

Peubah acak Y dalam fungsi sebaran peluang dapat disajikan dalam bentuk fungsi kumulatif sebagai berikut:

$$F(y) = P(Y \leq y) \tag{1}$$

Persamaan (1) memiliki fungsi kebalikan yang merupakan kuantil ke- τ dari

Y , yaitu

$$Q(\tau) = F^{-1}(\tau) = \inf\{y: F(y) \geq \tau\}$$

Untuk setiap $0 < \tau < 1$. Sebagai contoh untuk $\tau = \frac{1}{2}$, $Q(1/2)$ merupakan median.

Median contoh dari contoh acak Y , (y_1, \dots, y_n) , adalah penduga yang meminimumkan jumlah mutlak galat yang disajikan dalam bentuk

$$\min_{\alpha \in R} \sum_{i=1}^n |y_i - \alpha|$$

dengan α merupakan nilai duga \hat{y}_i .

Seperti halnya median contoh, metode meminimumkan jumlah mutlak galat dapat dikembangkan untuk model regresi. Berikut merupakan persamaan model regresi kuantil, yaitu

$$y = X\beta + \varepsilon \tag{2}$$

dengan y adalah vektor peubah respon berukuran $n \times 1$. X adalah matriks peubah penjelas berukuran $n \times p$. β adalah vektor parameter berukuran $p \times 1$. ε adalah vektor galat berukuran $n \times 1$. Median contoh dapat diperluas menjadi regresi median atau regresi L_1 . Penduga koefisien pada model L_1 merupakan solusi dari meminimumkan fungsi

$$\min_{\beta \in R} \sum_{i=1}^n |y_i - x'_i \beta|$$

Hubungan fungsional antara peubah respon dengan peubah penjelas pada RK merupakan hubungan fungsional antara kuantil bersyarat peubah respon dengan peubah penjelas yang membentuk fungsi linear, yaitu

$$Q(\tau|X = x_i) = x'_i \beta(\tau)$$

Secara umum menurut Koenker (2005) penduga regresi kuantil ke- τ untuk $0 < \tau < 1$ merupakan solusi dari masalah minimisasi fungsi berikut

$$\min_{\beta \in R} \left\{ \sum_{i \in \{i|y_i \geq x'_i \beta\}} \tau |y_i - x'_i \beta| + \sum_{i \in \{i|y_i < x'_i \beta\}} (1 - \tau) |y_i - x'_i \beta| \right\} \tag{3}$$

Persamaan (3) memberikan bobot τ untuk semua pengamatan yang lebih besar dari nilai optimum yang belum diketahui dan memberikan bobot $(1 - \tau)$ terhadap semua pengamatan yang lebih kecil dari nilai optimum. Persamaan (3) dapat dibentuk menjadi persamaan dengan ekspresi tunggal berikut:

$$\min_{\beta \in R} \sum_{i=1}^n \rho_\tau(y_i - x'_i \beta) \tag{4}$$

dengan $\rho_\tau(u)$ adalah fungsi kerugian yang tidak simetrik. Fungsi kerugian dapat didefinisikan sebagai berikut:

$$\rho_\tau(u) = \begin{cases} u\tau, & \text{jika } u \geq 0 \\ u(\tau-1), & \text{jika } u < 0 \end{cases} \tag{5}$$

Salah satu metode yang sering digunakan untuk pendugaan parameter regresi kuantil adalah metode simpleks pada pemrograman linear yang dilakukan oleh Chen dan Wei (2005). Menurut Koenker dan Machado (1999), Persamaan (4) akan sama dengan nilai dari *maksimum likelihood* fungsi distribusi *Asymmetric Laplace*. Yu dan Moyeed (2001) mempopulerkan metode Bayes pada regresi kuantil. Mereka memperkenalkan gagasan regresi kuantil menggunakan fungsi *likelihood* yang didasarkan pada distribusi *Asymmetric Laplace*.

Penelitian tentang *statistical downscaling* menggunakan regresi kuantil telah banyak dilakukan. Wigena dan Djuraidah (2014) melakukan pendugaan curah hujan ekstrim menggunakan regresi kuantil dengan Analisis Komponen Utama sebagai reduksi dimensinya. Kemudian, berdasarkan hasil penelitian Zaikarina *et al.* (2016) yang melakukan pendugaan terhadap curah hujan ekstrim menggunakan regresi kuantil LASSO dan gulud, didapatkan bahwa regresi kuantil LASSO lebih baik dalam memprediksi curah hujan ekstrim dibandingkan regresi kuantil gulud. Penelitian yang dilakukan oleh Hendri *et al.* (2019) adalah tentang regresi kuantil Bayesian untuk menduga curah hujan ekstrim di Jawa Barat. Hasil dari penelitian tersebut menunjukkan bahwa model regresi kuantil Bayes cukup akurat untuk memprediksi curah hujan di Jawa Barat dibandingkan dengan model regresi kuantil.

2.3 Least Absolute Shrinkage and Selection Operator (LASSO)

Tibshirani (1996) memperkenalkan metode *least absolute shrinkage and selection operator* (LASSO) yang merupakan metode alternatif dari metode kuadrat terkecil ketika terdapat masalah multikolinearitas. Multikolinearitas dapat menyebabkan ragam dari penduga menjadi besar sehingga akurasi dari prediksi mengecil. Akurasi prediksi ini dapat ditingkatkan dengan cara menyusutkan beberapa koefisien atau mengatur koefisien lainnya menuju nol. LASSO dapat menyusutkan koefisien peubah mendekati nilai nol atau bahkan tepat nol sehingga dapat menyeleksi peubah yang tidak relevan. Dengan demikian, akan ada sedikit bias untuk mengurangi variansi dari nilai prediktor sehingga dapat meningkatkan akurasi prediksi secara keseluruhan dan mengatasi permasalahan multikolinearitas.

Penduga koefisien LASSO β^{LASSO} dapat ditulis dalam bentuk *Lagrangian* berikut:

$$\beta^{LASSO} = \min_{\beta \in R} \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j \right)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

Penduga regresi kuantil ke- τ untuk $\tau \in (0, 1)$ setelah ditambahkan penalti LASSO dapat dituliskan sebagai berikut:

$$\beta_{\tau}^{LASSO} = \min_{\beta \in R} \sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{x}_i' \boldsymbol{\beta}) + \lambda \sum_{j=1}^p |\beta_j|$$

dengan koefisien λ adalah parameter penalti, n adalah banyaknya pengamatan, p adalah banyaknya parameter, dan $\sum_{j=1}^p |\beta_j| \leq \lambda$, dimana $\lambda \geq 0$.

2.4 Regresi Kuantil Biner Bayes

Regresi kuantil biner diperkenalkan oleh Benoit dan Van den Poel (2010) yang merupakan pengembangan dari metode regresi kuantil biasa dengan peubah respon pada regresi ini berskala dikotomis yang memiliki dua kategori. Penelitian tersebut mengembangkan regresi kuantil biner dengan menggunakan pendekatan Bayesian melalui algoritma *Metropolis-Hasting* untuk mencari estimasi parameternya. Alhamzawi *et al.* (2013) melakukan penyempurnaan jurnal sebelumnya dengan mengembangkan regresi kuantil biner dengan tambahan penalti LASSO menggunakan analisis Bayesian melalui algoritma *Gibbs sampling*.

Secara matematis, koefisien regresi kuantil biner dengan penalti LASSO dapat dituliskan sebagai berikut (Alhamzawi *et al.* 2013)

$$\min_{\beta \in \mathbb{R}} \sum_{i=1}^n \rho_{\tau}(y_i - \mathbf{x}'_i \boldsymbol{\beta}) + \lambda \|\boldsymbol{\beta}\|_1$$

dengan y_i adalah peubah respon biner ke- i yang berdasarkan peubah y_i^* , dengan

$$y_i^* = \mathbf{x}'_i \boldsymbol{\beta}_{(\tau)} + \varepsilon_{i(\tau)}$$

dan

$$y_i = \begin{cases} 1, & \text{jika } y_i^* \geq 0 \\ 0, & \text{lainnya} \end{cases}$$

Koefisien $\lambda \geq 0$ adalah *Lagrange multiplier*, serta $\|\boldsymbol{\beta}\|_1 = \sum_{j=1}^p |\beta_j|$ adalah penalti l_1 LASSO untuk regresi kuantil biner. Penambahan penalti l_1 LASSO pada regresi kuantil biner dapat menyusutkan koefisien regresi mendekati nol bahkan tepat nol sehingga mampu menyeleksi peubah dan mengatasi permasalahan multikolinearitas.

Parameter populasi dipandang sebagai besaran yang tidak diketahui dalam metode klasik. Namun dari sudut pandang Bayes, populasi dipandang sebagai peubah yang memiliki sebaran yang disebut dengan sebaran prior. Pada metode Bayes, untuk menentukan sebaran posterior dari data diperlukan sebaran prior dan fungsi *likelihood* bersyarat dengan formula sebagai berikut:

$$\pi(\theta|x) \propto L(x|\theta)\pi(\theta).$$

Yu dan Moyeed (2001) memperkenalkan regresi kuantil Bayes yang mengikuti sebaran *Asymmetric Laplace* dengan fungsi kepekatan

$$f_{\tau}(r) = \tau(1 - \tau) \exp\{-\rho_{\tau}(r)\} \quad (6)$$

dengan $0 < \tau < 1$ dan $\rho_{\tau}(r)$ didefinisikan pada persamaan (5). Persamaan (6) memiliki rata-rata dan ragam sebagai berikut:

$$E(r) = \frac{1-2\tau}{\tau(1-\tau)} \text{ dan } Var(r) = \frac{1-2\tau+2\tau^2}{\tau^2(1-\tau)^2}$$

Jika terdapat parameter skala (σ) dan parameter lokasi (μ), maka fungsi kepekatan peluang dari sebaran *Asymmetric Laplace* sebagai berikut:

$$f_{\tau}(r; \mu, \sigma) = \frac{\tau(1-\tau)}{\sigma} \exp\left\{-\rho_{\tau}\left(\frac{r-\mu}{\sigma}\right)\right\}$$

dengan $\sigma > 0$ dan $0 < \tau < 1$.

Diasumsikan bahwa ε_i menyebar *Asymmetric Laplace* dengan fungsi kepekatan peluang sebagai berikut:

$$f_{\tau}(\varepsilon_i) = \tau(1 - \tau) \exp\{\rho_{\tau}(\varepsilon_i)\}$$

dengan $0 < \tau < 1$ dan $\rho_{\tau}(u)$ didefinisikan pada persamaan (5). Kozumi dan Kobayashi (2009) mengasumsikan bahwa $z \sim \exp(1)$ dan $u \sim N(0,1)$. Jika peubah acak ε menyebar *Asymmetric Laplace*, maka ε dapat direpresentasikan sebagai *location-scale mixture* dari sebaran normal yang dapat ditulis menjadi

$$\varepsilon = \omega z + \psi\sqrt{z}u \quad (7)$$

dengan

$$\omega = \frac{1-2\tau}{\tau(1-\tau)} \text{ dan } \psi^2 = \frac{2}{\tau(1-\tau)}$$

Substitusi persamaan (7) ke dalam persamaan (2) sehingga y_i^* dapat ditulis sebagai berikut:

$$y_i^* = \mathbf{x}'_i \boldsymbol{\beta} + \omega z_i + \psi\sqrt{z_i}u_i \quad (8)$$

dengan $z_i \sim \exp(1)$ dan $u_i \sim N(0,1)$ yang masing-masing saling bebas stokastik.

Jika terdapat parameter skala $\sigma > 0$ di dalam model, maka persamaan (8) menjadi

$$y_i^* = \mathbf{x}'_i \boldsymbol{\beta} + \omega \sigma z_i + \psi \sigma \sqrt{z_i} u_i \quad (9)$$

Persamaan (9) belum sesuai untuk menghasilkan *Gibbs Sampling* sebagai parameter skala yang muncul pada kondisi rata-rata dari y_i^* . Oleh karena itu, dimisalkan $v_i = \sigma z_i$ sehingga persamaan (9) menjadi

$$y_i^* = \mathbf{x}'_i \boldsymbol{\beta} + \omega v_i + \psi \sqrt{\sigma v_i} u_i$$

dengan $\mathbf{v} = [v_1 \dots v_n]'$ dan $\mathbf{v} \sim \exp(\sigma)$.

Fungsi sebaran kumulatif dari u_i adalah

$$F(u_i) = \int_{-\infty}^{u_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u_i^2}{2}\right) du_i \quad (10)$$

dengan $-\infty < u_i < \infty$ dan $i = 1, 2, \dots, n$. Jika

$$u_i = \frac{y_i^* - \mathbf{x}'_i \boldsymbol{\beta} - \omega v_i}{\psi \sqrt{\sigma v_i}} \text{ dan } du_i = \frac{1}{\psi \sqrt{\sigma v_i}} dy_i^*$$

maka persamaan (10) menjadi

$$F(u_i) = \int_{-\infty}^{y_i^*} \frac{1}{\psi \sqrt{\sigma v_i} \sqrt{2\pi}} \exp\left(-\frac{(y_i^* - \mathbf{x}'_i \boldsymbol{\beta} - \omega v_i)^2}{2\psi^2 \sigma v_i}\right) dy_i^*$$

sehingga diperoleh fungsi kepekatan peluang dari y_i^* bersyarat $\boldsymbol{\beta}, \sigma, v_i$ sebagai berikut

$$f(y_i^* | \boldsymbol{\beta}, \sigma, v_i) = \frac{1}{\psi \sqrt{\sigma v_i} \sqrt{2\pi}} \exp\left(-\frac{(y_i^* - (\mathbf{x}'_i \boldsymbol{\beta} + \omega v_i))^2}{2\psi^2 \sigma v_i}\right) \quad (11)$$

Persamaan (11) merupakan fungsi kepekatan peluang yang menyebar normal dengan rata-rata $\mathbf{x}'_i \boldsymbol{\beta} - \omega v_i$ dan ragam $\psi^2 \sigma v_i$.

Fungsi *likelihood* dari persamaan (11), yaitu

$$L(\mathbf{y}^* | \boldsymbol{\beta}, \mathbf{v}, \sigma) = \left(\prod_{i=1}^n \frac{1}{\psi \sqrt{\sigma v_i} \sqrt{2\pi}} \right) \left(\exp \left(- \sum_{i=1}^n \frac{(y_i^* - (x_i' \boldsymbol{\beta} + \omega v_i))^2}{2\psi^2 \sigma v_i} \right) \right) \\ \propto \left(\prod_{i=1}^n (\sigma v_i)^{-1/2} \right) \left(\exp \left(- \sum_{i=1}^n \frac{(y_i^* - (x_i' \boldsymbol{\beta} + \omega v_i))^2}{2\psi^2 \sigma v_i} \right) \right)$$

Sebaran prior untuk $\beta_j, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta$ yang digunakan dalam regresi kuantil biner telah dijelaskan pada Alhamzawi *et al.* (2013). Selanjutnya, setelah didapatkan sebaran prior maka sebaran posterior dapat dicari melalui sebaran prior yang ada. Sebaran posterior bersama dapat dicari melalui perkalian antara fungsi *likelihood* yang diperoleh dari data amatan dengan sebaran prior. Akan dicari sebaran posterior bersama $\pi(\mathbf{y}^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta | \mathbf{y})$ dengan metode Box-Tiao yaitu

$$\pi(\mathbf{y}^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta | \mathbf{y}) \propto \pi(\mathbf{y}^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta) \pi(\mathbf{y} | \mathbf{y}^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta) \\ \propto \pi(\mathbf{y}^* | \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta) \pi(\boldsymbol{\beta} | \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta) \\ \pi(\eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta) L(\mathbf{y} | \mathbf{y}^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta) \quad (12)$$

Persamaan (12) apabila diselesaikan secara analitik akan sulit dilakukan karena memiliki bentuk yang kompleks. Oleh karena itu, akan dilakukan pendekatan *Gibbs sampling* untuk mencari distribusi posterior bersama. Dalam *Gibbs sampling* diperlukan sebaran bersyarat lengkap dari setiap parameter yang akan diduga. Sebaran bersyarat lengkap dari setiap parameter dijelaskan pada Rahayu (2015).

2.4 Markov Chain Monte Carlo (MCMC) algoritma Gibbs Sampling

Tujuan dasar dari algoritma MCMC adalah mensimulasi contoh-contoh dari sebaran posterior untuk suatu parameter. Setiap contoh yang terambil bergantung pada contoh yang terambil sebelumnya sehingga membentuk suatu rantai *Markov Monte Carlo* digunakan untuk membuat suatu pendekatan pada suatu nilai harapan dengan menggunakan contoh dari rantai *Markov*. Algoritma yang menerapkan MCMC yaitu *Gibbs Sampling* yang dipopulerkan oleh Geman dan Geman (1984). *Gibbs sampling* dapat diterapkan apabila sebaran peluang bersama tidak diketahui secara eksplisit, tetapi sebaran bersyarat dari tiap parameter diketahui (Hoff 2009). Menurut Alhamzawi *et al.* (2013) algoritma *Gibbs sampling* lebih efisien digunakan daripada algoritma *Metropolis-Hasting* yang lebih kompleks dan menghabiskan waktu yang lama. Kerangka kerja *Gibbs sampling* adalah mendekomposisikan sebaran posterior bersama ke dalam sebaran bersyarat untuk setiap parameter dalam model, kemudian contoh diambil dari proses tersebut. Rata-rata dari parameter digunakan sebagai penduga parameter model.

Proses *Gibbs sampling* untuk mencari pendugaan regresi kuantil biner Bayes dengan penalti LASSO dilakukan dengan membangkitkan contoh sebaran posterior bersyarat. Sebaran bersyarat lengkap untuk setiap parameter yang telah dijelaskan dalam Rahayu (2015) digunakan dalam proses *Gibbs sampling* untuk mencari sebaran posterior bersama. Dugaan parameter untuk $y_i^*, \beta_j, \eta^2, \zeta, \sigma, s_j, v_i, \delta$ pada regresi kuantil biner Bayes dengan penalti LASSO adalah rata-rata dari sebaran posterior bersama $\pi(y_i^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta | \mathbf{y})$ yang didapatkan dari proses *Gibbs sampling* tersebut.

Alhamzawi *et al.* (2013) telah melakukan simulasi MCMC dengan tiga parameter berbeda dan membandingkan performa beberapa metode yaitu regresi kuantil biner Bayes dengan penalti LASSO, regresi kuantil biner, dan regresi kuantil biner dengan pemulusan. Hasil dari penelitian tersebut menunjukkan bahwa regresi kuantil biner Bayes dengan penalti LASSO memiliki performa paling baik dibandingkan metode lainnya. Mereka juga mengaplikasikan data aktual ke dalam metode regresi kuantil biner Bayes terpenalti LASSO dan didapatkan hasil yang sama baiknya seperti simulasi yang telah dilakukan.

@Hak cipta milik IPB University

IPB University





III METODE PENELITIAN

3.1 Data

Data yang digunakan pada penelitian ini adalah data sekunder dari tahun 1981 hingga 2009 dengan data curah hujan bulanan di Jawa Barat merupakan peubah respon yang diperoleh dari BMKG. Data curah hujan dapat dikelompokkan berdasarkan ketinggian dataran yaitu dataran rendah (0-200 mdpl), dataran sedang (200-500 mdpl) dan dataran tinggi (>500 mdpl). Data curah hujan untuk dataran rendah diperoleh dari 12 stasiun yang terletak di dataran rendah dan untuk data curah hujan di dataran sedang dan tinggi diperoleh dari 3 stasiun yang masing-masing terletak di dataran sedang dan tinggi.

Peubah penjelas pada penelitian ini adalah data luaran GCM yang merupakan data presipitasi bulanan *Climate Forecast System Reanalysis* (CFSR) dengan ukuran grid $2.5^{\circ} \times 2.5^{\circ}$ pada domain 5×8 grid atau 40 grid selama 29 tahun (1981 hingga 2009) berdasarkan pemilihan domain terbaik (Fadli *et al.* 2019). Data presipitasi bulanan CFSR adalah sebuah hasil model matematika yang menggambarkan interaksi global antara dataran, lautan, dan udara bumi yang diperoleh dari *The National Centers for Environmental Prediction* (NCEP) (<https://rda.ucar.edu>).

3.2 Metode Analisis

Adapun langkah-langkah penelitian ini adalah sebagai berikut:

1. Pengelompokkan daerah di Jawa Barat menjadi tiga bagian, yaitu dataran rendah (0-200 mdpl), dataran sedang (200-500 mdpl) dan dataran tinggi (>500 mdpl).
2. Pengelompokkan curah hujan bulanan di Jawa Barat sehingga menjadi dua tipe yaitu curah hujan ekstrim $y_i = 1$ dan tidak ekstrim $y_i = 0$ dengan menggunakan *measure of surprise* yang telah dilakukan oleh Manurung *et al.* (2018) untuk menentukan ambang batas.
3. Eksplorasi data curah hujan biner.
4. Pembagian data curah hujan bulanan di Jawa Barat menjadi dua bagian, yaitu data curah hujan tahun 1981-2008 sebagai data pemodelan dan data curah hujan tahun 2009 sebagai data validasi.
5. Memprediksi curah hujan ekstrim bulanan dengan metode Regresi Kuantil Biner Bayes pada kuantil 0.70, 0.75, 0.80, 0.85, 0.90, 0.925, 0.95, 0.975, dan 0.99 dengan model berikut:

$$y_i^* = \mathbf{x}_i' \boldsymbol{\beta} + \omega v_i + \psi \sqrt{\sigma v_i} u_i$$

dimana $y_i = \begin{cases} 1, & \text{jika } y_i^* \geq \text{ambang batas} \\ 0, & \text{lainnya} \end{cases}$

Menurut Benoit dan Van den Poel (2017) paling tidak hasil model dari 9 kuantil dibutuhkan untuk mendapatkan interpretasi yang bermakna dari peluang prediksi berupa *confusion matrices*. Selanjutnya, pendugaan parameter yang dicari digunakan metode Bayes dengan menggunakan bantuan algoritma MCMC (*Markov Chain Monte Carlo*) yaitu *Gibbs sampling* dengan algoritma sebagai berikut:

a) Ambil sebuah nilai inisiasi untuk $y_i^*, \beta_j, \eta^2, \zeta, \sigma, s_j, v_i, \delta$ yaitu $y_i^{*(0)}, \beta_j^{(0)}, \eta^{2(0)}, \zeta^{(0)}, \sigma^{(0)}, s_j^{(0)}, v_i^{(0)}, \delta^{(0)}$.

b) Untuk iterasi ke-1

Bangkitkan contoh $y_i^{*(1)}, y_i^{*(1)} \sim \pi(y_i^* | \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta, \mathbf{y})$

Bangkitkan contoh $\beta_j^{(1)}, \beta_j^{(1)} \sim \pi(\beta_j | y^*, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta, \mathbf{y})$

Bangkitkan contoh $\sigma^{(1)}, \sigma^{(1)} \sim \pi(\sigma | y^*, \boldsymbol{\beta}, \eta^2, \zeta, \mathbf{s}, \mathbf{v}, \delta, \mathbf{y})$

Bangkitkan contoh $\eta^{2(1)}, \eta^{2(1)} \sim \pi(\eta^2 | y^*, \boldsymbol{\beta}, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta, \mathbf{y})$

Bangkitkan contoh $s_j^{(1)}, s_j^{(1)} \sim \pi(s_j | y^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{v}, \delta, \mathbf{y})$

Bangkitkan contoh $v_i^{(1)}, v_i^{(1)} \sim \pi(v_i | y^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \delta, \mathbf{y})$

Bangkitkan contoh $\zeta^{(1)}, \zeta^{(1)} \sim \pi(\zeta | y^*, \boldsymbol{\beta}, \eta^2, \sigma, \mathbf{s}, \mathbf{v}, \delta, \mathbf{y})$

Bangkitkan contoh $\delta^{(1)}, \delta^{(1)} \sim \pi(\delta | y^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \mathbf{y})$

c) Ulangi langkah 5(b) sebanyak $m=5000$ iterasi. Proses *burn-in* dilakukan terhadap $p=1000$ iterasi pertama. Setelah proses *burn-in*, iterasi berikutnya menunjukkan kekonvergenan rantai *Markov*.

d) Didapatkan contoh yang memiliki sebaran posterior bersama $\pi(y_i^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta | \mathbf{y})$.

Penduga-penduga regresi kuantil biner Bayes terpenalti LASSO yang dihasilkan dari tahap *gibbs sampling* di atas adalah rata-rata dari sebaran posterior bersama $\pi(y_i^*, \boldsymbol{\beta}, \eta^2, \zeta, \sigma, \mathbf{s}, \mathbf{v}, \delta | \mathbf{y})$.

6. Evaluasi model berdasarkan nilai akurasi.

Tabel 1 *Confusion matrices* antara kelas sebenarnya dan kelas prediksi

<i>Predicted Values</i>	<i>Actual Values</i>	
	Ekstrim (1)	Tidak Ekstrim (0)
Ekstrim (1)	A	B
Tidak Ekstrim (0)	C	D

Akurasi merupakan tingkat ketepatan prediksi secara keseluruhan. Nilai akurasi dapat dihitung dengan rumus sebagai berikut (Zhou et al. 2017)

$$\text{Akurasi} = \frac{(A+D)}{(A+B+C+D)}$$

dengan,

A : Jumlah amatan curah hujan ekstrim diprediksi ekstrim

B : Jumlah amatan curah hujan tidak ekstrim diprediksi ekstrim

C : Jumlah amatan curah hujan ekstrim diprediksi tidak ekstrim

D : Jumlah amatan curah hujan tidak ekstrim diprediksi tidak ekstrim

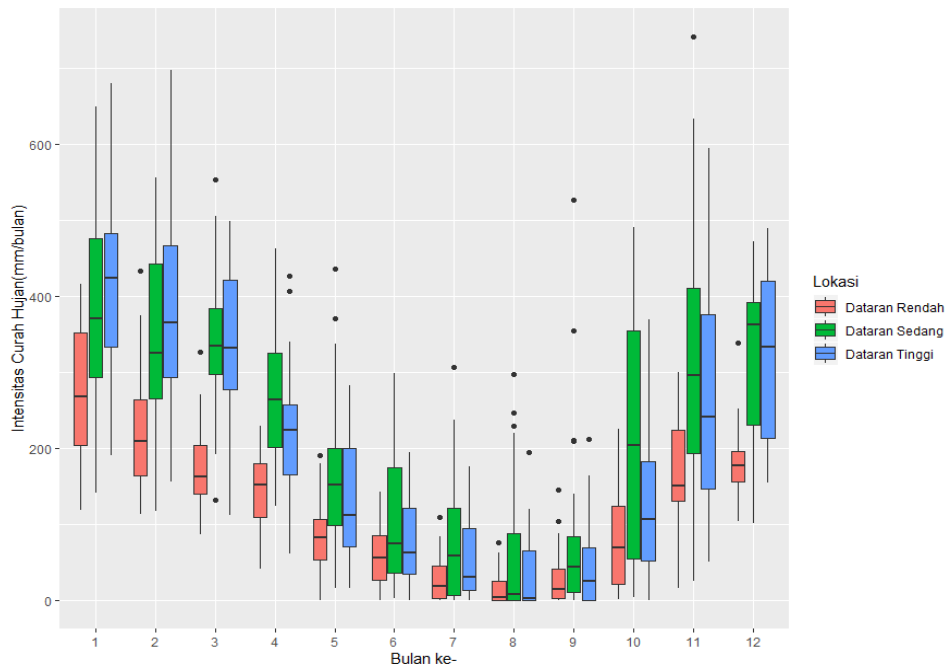


IV HASIL DAN PEMBAHASAN

4.1 Eksplorasi Data

4.1.1 Data Curah Hujan

Data yang digunakan dalam penelitian ini terlampir pada Lampiran 1. Analisis statistika deskriptif untuk data curah hujan di masing-masing dataran disajikan pada Lampiran 2-4. Deskripsi data curah hujan di Jawa Barat setiap dataran disajikan pada Gambar 1.



Gambar 1 Diagram kotak garis data curah hujan di Jawa Barat

Berdasarkan Gambar 1, terlihat pola dari curah hujan setiap dataran berbentuk U. Musim hujan memiliki rata-rata intensitas curah hujan lebih besar dari 150 mm/bulan sehingga musim hujan di dataran rendah Jawa Barat terjadi antara bulan Oktober - Mei, sedangkan di dataran sedang dan tinggi musim hujan terjadi sepanjang tahun. Intensitas curah hujan pada dataran sedang dan tinggi lebih tinggi daripada intensitas curah hujan di dataran rendah. Curah hujan ekstrim menurut BMKG memiliki intensitas curah hujan bulanan lebih besar dari 400 mm/bulan. Berdasarkan nilai tersebut, curah hujan ekstrim terjadi pada bulan Januari dan Februari di dataran rendah. Di dataran sedang dan tinggi, curah hujan ekstrim terjadi pada bulan Oktober - April. Namun, sebenarnya intensitas curah hujan ekstrim tidak dapat disamaratakan untuk setiap dataran. Ambang batas ekstrim untuk intensitas curah hujan ekstrim di masing-masing dataran dapat berbeda dikarenakan perbedaan ketinggian dataran. Data curah hujan ini kemudian akan digunakan untuk dikelompokkan berdasarkan nilai ambang batas yang didapatkan dari metode MOS pada tahap selanjutnya.

4.1.2 Data Luaran GCM

Data luaran GCM memiliki dimensi besar yang memungkinkan terjadinya multikolinieritas sehingga perlu diperiksa dengan melihat nilai *variance inflation factor* (VIF). Nilai VIF yang lebih dari 10 menunjukkan nilai korelasi antara peubah penjelas sangat tinggi yang menyebabkan masalah multikolinieritas menjadi serius. Masalah tersebut mengakibatkan tingkat akurasi dari prediksi berkurang dan dapat menjadi keliru dalam mengambil keputusan. Berdasarkan Tabel 2, terdapat grid-grid atau peubah yang memiliki nilai VIF lebih besar dari 10. Hal tersebut menunjukkan bahwa terdapat multikolinieritas pada grid atau peubah tersebut. Masalah multikolinieritas pada peubah-peubah ini dapat diatasi dengan penambahan penalti LASSO pada metode regresi kuantil biner Bayes yang mampu menyusutkan koefisien peubah mendekati nol dan menyeleksi peubah yang kurang berpengaruh dalam model yang akan dipakai pada tahap selanjutnya.

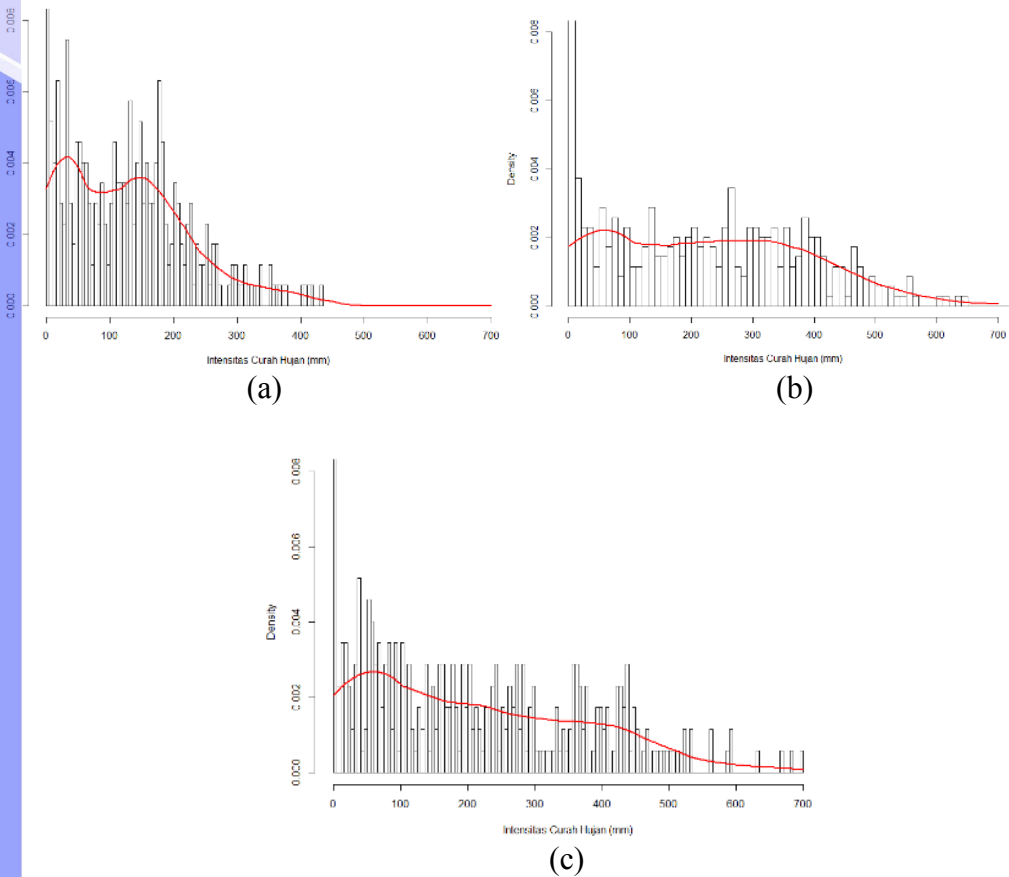
Tabel 2 Nilai VIF data luaran GCM

Peubah	VIF	Peubah	VIF	Peubah	VIF	Peubah	VIF
X1	2.22	X11	5.04	X21	13.99	X31	14.14
X2	3.10	X12	5.20	X22	12.33	X32	12.54
X3	3.09	X13	5.70	X23	8.32	X33	4.61
X4	2.58	X14	4.02	X24	5.52	X34	7.66
X5	2.65	X15	4.37	X25	5.41	X35	8.73
X6	3.29	X16	4.66	X26	8.29	X36	7.45
X7	2.98	X17	4.20	X27	8.00	X37	8.18
X8	3.44	X18	4.48	X28	10.20	X38	14.81
X9	3.80	X19	5.75	X29	10.27	X39	18.34
X10	4.63	X20	8.06	X30	7.68	X40	11.12

4.2 Measure of Surprise

Sebaran peluang curah hujan akan jauh dari normal ketika terjadi curah hujan ekstrim. Bentuk sebaran di bawah ambang batas sulit ditentukan dari eksplorasi grafik. Sebelum dilakukan pendugaan nilai ambang batas dengan plot *measure of surprise* perlu dilihat terlebih dahulu bentuk sebaran data curah hujan di masing-masing dataran dengan histogram dan plot densitasnya. Histogram dan plot densitas data curah hujan untuk rata-rata curah hujan di beberapa stasiun di dataran rendah, sedang, dan tinggi di Jawa Barat disajikan pada Gambar 2.

Ketiga histogram curah hujan di masing-masing dataran memiliki persamaan yaitu puncak global di sekitar nilai 50-70 mm. Perbedaannya terdapat pada puncak lokal pada masing-masing dataran. Curah hujan di dataran rendah memiliki puncak lokal setelah puncak global yaitu sekitar 110 mm dan 150 mm. Curah hujan di dataran sedang memiliki puncak lokal setelah puncak global yaitu sekitar 250 mm dan 330 mm. Curah hujan di dataran tinggi memiliki puncak lokal setelah puncak global yaitu sekitar 210, 350 dan 410 mm. Meskipun dilihat secara subjektif, tetapi nilai-nilai tersebut dapat berpotensi sebagai nilai ambang batas yang akan digunakan untuk mengkategorikan curah hujan ke dalam kelompok ekstrim dan tidak ekstrim di setiap dataran di Jawa Barat.

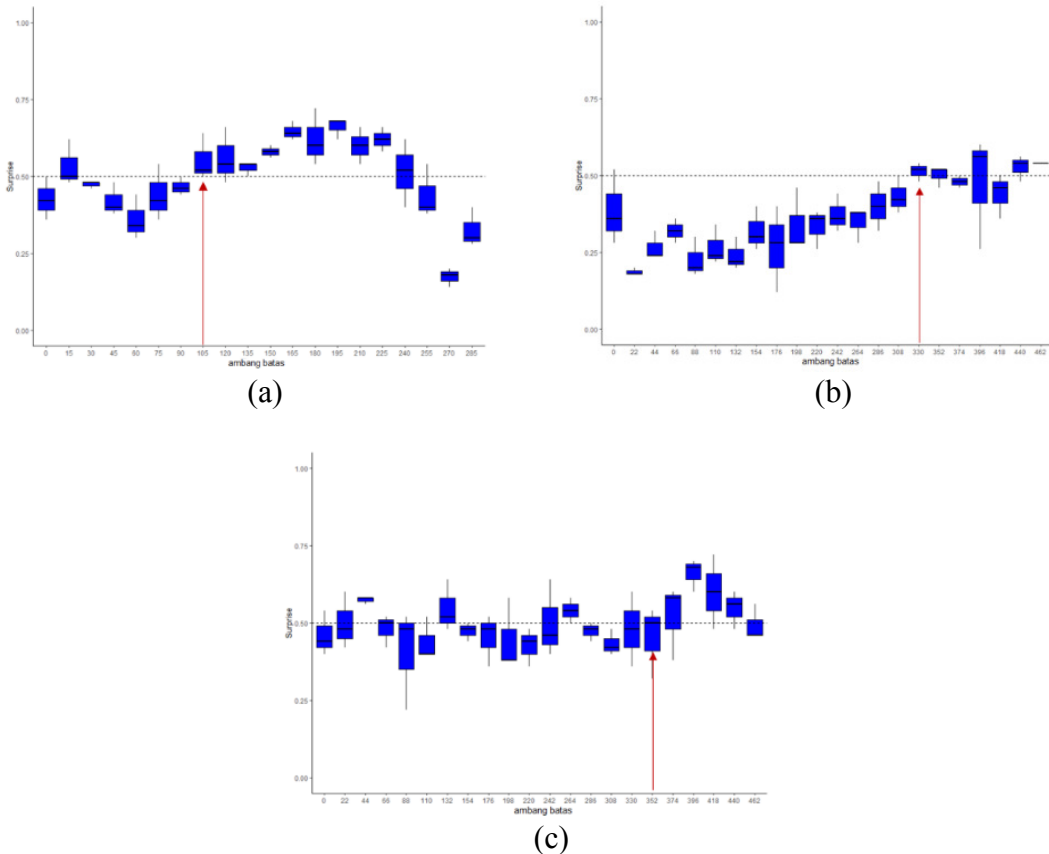


Gambar 2 Histogram curah hujan (a) dataran rendah, (b) dataran sedang, (c) dataran tinggi

Setelah didapatkan nilai-nilai yang berpotensi menjadi nilai ambang batas dari histogram di atas, selanjutnya dapat dilihat juga diagram MOS berdasarkan sebaran *Generalized Pareto Distribution* (GPD) dari data curah hujan di masing-masing dataran. Diagram MOS curah hujan untuk masing-masing dataran di Jawa Barat terdapat pada Gambar 3. Diagram MOS terdiri dari beberapa diagram kotak garis yang mewakili jumlah kandidat ambang batas pada masing-masing dataran. Garis putus-putus menunjukkan batas terendah saat *surprise* mulai stabil pada nilai di 0.5.

Calon kandidat ambang batas pada diagram MOS yang digunakan pada penelitian ini ditentukan secara otomatis seperti pendekatan membuat histogram agar lebih praktis. Ambang batas dugaan mencapai 285 mm pada dataran rendah, sedangkan di dataran sedang dan tinggi mencapai 462 mm. Hal tersebut menunjukkan bahwa intensitas curah hujan di dataran rendah lebih kecil dibandingkan intensitas curah hujan di dataran sedang dan dataran tinggi Jawa Barat. Sebaran GPD di dataran rendah dimulai pada 105 mm, di dataran sedang 330 mm, dan di dataran tinggi 352 mm. Titik-titik yang masuk dalam sebaran GPD tersebut selanjutnya akan menjadi kandidat dugaan ambang batas untuk curah hujan ekstrim pada masing-masing dataran.

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mempublikasikan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



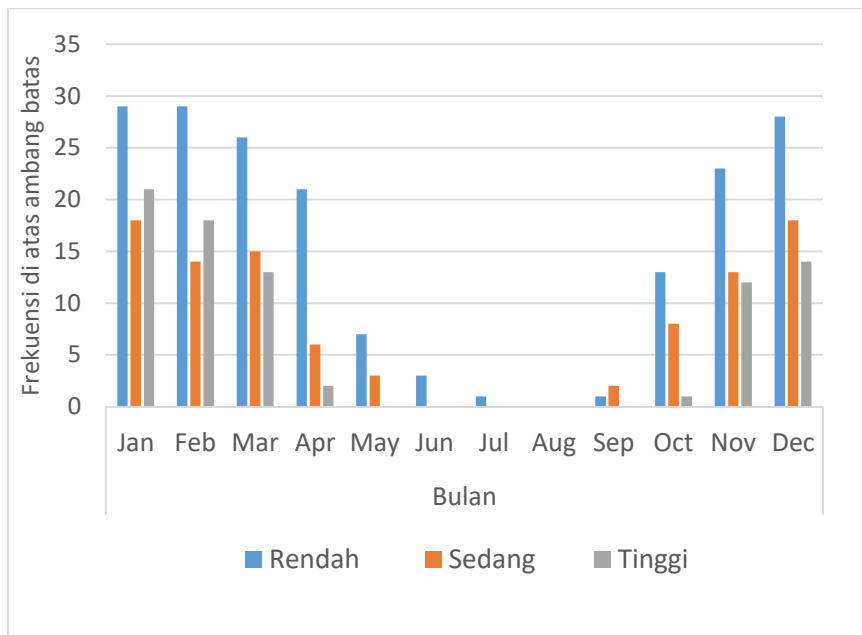
Gambar 3 Diagram *measure of surprise* (a) data dataran rendah, (b) data dataran sedang, (c) data dataran tinggi

Berdasarkan eksplorasi grafik dan diagram MOS yang telah dilakukan sebelumnya, didapatkan ambang batas untuk dataran rendah adalah sekitar titik 105 (± 5) mm, dataran sedang 330 (± 5) mm, dataran tinggi 352 (± 5) mm. Selanjutnya nilai-nilai tersebut digunakan untuk mengkategorikan curah hujan di setiap dataran di Jawa Barat ke dalam kelompok curah hujan ekstrim $y_i = 1$ dan tidak ekstrim $y_i = 0$. Tabel 3 menunjukkan analisis diagram MOS di setiap dataran dan kaitannya dengan nilai yang dicurigai dari eksplorasi grafik.

Tabel 3 Analisis *measure of surprise* data curah hujan

Dataran	Jenis Puncak	Kandidat titik dugaan	Mencapai 0.5	Stabil di sekitar dugaan
Rendah	Global	50	Belum	Tidak
	Lokal	110	Sudah	Ya
	Lokal	150	Sudah	Ya
Sedang	Global	50	Belum	Tidak
	Lokal	250	Belum	Ya
	Lokal	330	Sudah	Ya
Tinggi	Global	70	Belum	Tidak
	Lokal	210	Belum	Ya
	Lokal	350	Sudah	Ya
	Lokal	410	Sudah	Ya

Visualisasi untuk data curah hujan setelah dikelompokkan ke dalam tipe curah hujan ekstrim dan tidak ekstrim berdasarkan nilai ambang batas sebaran GDP disajikan pada Gambar 4. Frekuensi amatan curah hujan yang telah diubah menjadi data biner di atas ambang batas adalah 181 amatan (52%) di dataran Rendah, 97 amatan (28%) di dataran Sedang, dan 81 amatan (23%) di dataran Tinggi. Data biner ini kemudian digunakan untuk memprediksi curah hujan ekstrim di Jawa Barat dengan metode regresi kuantil biner Bayes terpenalti LASSO.



Gambar 4 Diagram batang curah hujan di atas ambang batas pada masing-masing dataran di Jawa Barat

4.3 Regresi Kuantil Biner Bayes

Pembentukan model regresi kuantil biner Bayes dilakukan pada kuantil 0.70, 0.75, 0.80, 0.85, 0.90, 0.925, 0.95, 0.975, dan 0.99 di masing-masing dataran. Pembentukan model regresi kuantil biner Bayes di 9 titik kuantil tersebut ditujukan untuk mendapatkan nilai akurasi. Data biner yang didapatkan dari tahap sebelumnya digunakan sebagai peubah respon, sedangkan peubah penjelasnya adalah data luaran GCM pada domain berukuran 5×8 grid atau 40 grid yang masih mengandung multikolinearitas.

Secara umum, koefisien dari peubah penjelas yang masuk ke dalam model regresi kuantil biner Bayes dengan penalti LASSO disusutkan mendekati nol dan bahkan tepat nol sehingga terdapat beberapa peubah penjelas yang mengandung multikolinearitas terseleksi dari model. Tabel 4 merupakan contoh koefisien dan peubah LASSO yang terpilih untuk dataran rendah pada kuantil 0.75. Peubah penjelas pada model regresi kuantil biner Bayes Q (0.75) yang semulanya terdapat 40 peubah, telah terseleksi oleh LASSO dan menyisakan 17 peubah yang masuk ke dalam model. Beberapa peubah yang telah terseleksi dan dikeluarkan dari model tersebut, ternyata terdapat peubah yang memiliki nilai VIF lebih dari 10 yang menyebabkan multikolinearitas yaitu X22, X28, X31, X32, dan X39. Oleh karena telah dikeluarkannya beberapa peubah yang mengandung multikolinearitas

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :

a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.

2. Dilarang mengumunkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

tersebut, maka prediksi dari model yang terbentuk menjadi lebih akurat. Koefisien dari peubah penjelas yang masuk ke dalam model regresi kuantil biner Bayes juga telah signifikan dalam model regresi kuantil biner Bayes karena memiliki *credible interval* yang tidak memuat nilai nol di dalamnya. Peubah LASSO dan koefisien dari model regresi kuantil biner Bayes di masing-masing dataran disajikan dalam Lampiran 5-11.

Tabel 4 Koefisien dan peubah LASSO model regresi kuantil biner Bayes pada kuantil 0.75 di dataran rendah

Peubah	Koefisien	<i>Credible Interval</i>	
		Batas Bawah	Batas Atas
X1	-11.45	-15.70	-2.18
X5	0.46	0.03	0.89
X9	7.81	2.16	9.68
X10	3.27	1.00	5.06
X12	2.48	1.49	3.11
X14	4.56	1.27	5.70
X16	2.70	1.52	3.72
X17	3.10	1.03	4.15
X19	3.61	1.70	4.37
X21	3.87	1.47	4.95
X26	3.06	1.26	3.60
X27	1.73	0.78	2.39
X29	3.51	1.57	3.96
X35	3.44	1.79	4.10
X37	2.96	0.94	5.44
X38	3.00	1.50	4.06
X40	4.68	1.50	6.18

4.4 Evaluasi Model Regresi Kuantil Biner Bayes

Koefisien dan peubah dari model regresi kuantil biner Bayes telah diperoleh sebelumnya, untuk selengkapnya terlampir pada Lampiran 5-11. Selanjutnya, dilakukan evaluasi model pada masing-masing dataran yaitu dengan melihat nilai akurasi prediksi curah hujan 1, 2, dan 3 tahun terakhir dari setiap model yang terbentuk pada beberapa titik ekstrim. Nilai akurasi prediksi untuk dataran rendah disajikan pada Tabel 5.

Tabel 5 Nilai akurasi model regresi kuantil biner Bayes pada dataran rendah

Titik Ekstrim	Akurasi Dataran Rendah		
	1 tahun	2 tahun	3 tahun
100mm	91.67%	95.83%	91.67%
105mm	91.67%	95.83%	91.67%
110mm	91.67%	95.83%	94.44%
400mm (BMKG 2008)	91.67%	95.83%	94.44%

Berdasarkan Tabel 5, akurasi prediksi model regresi kuantil biner Bayes pada dataran rendah untuk prediksi 1 tahun terakhir adalah sebesar 91.67%, untuk

prediksi 2 tahun terakhir sebesar 95.83%, sedangkan untuk 3 tahun terakhir pada titik ekstrim 100 dan 105 mm sebesar 91.67%, serta titik ekstrim 110 mm dan titik ekstrim 400 mm (BMKG 2008) sebesar 94.44%. Berdasarkan nilai tersebut, dipilih titik ekstrim yang memiliki nilai akurasi prediksi yang tertinggi untuk digunakan sebagai ambang batas dari MOS yaitu 110 mm.

Akurasi prediksi model regresi kuantil biner Bayes pada dataran sedang dan tinggi disajikan pada Tabel 6 dan Tabel 7. Berdasarkan nilai akurasi prediksi pada Tabel 6 yaitu dataran sedang, terlihat bahwa nilai akurasinya sedikit naik turun jika dilihat sampai prediksi 3 tahun terakhir. Hal tersebut dimungkinkan karena terdapat beberapa data yang kosong pada stasiun di dataran sedang pada beberapa tahun di awal, tetapi hal tersebut masih dapat ditoleransi. Sehingga, dipilih titik ekstrim yang memiliki nilai akurasi yang tinggi dan stabil pada 1, 2, dan 3 tahun terakhir yaitu titik 330mm, tetapi jika dilihat nilai akurasi prediksi sampai 2 tahun terakhir maka titik 335mm dapat dipilih menjadi titik ekstrim. Berdasarkan Tabel 7 yaitu dataran tinggi, dipilih titik ekstrim yang memiliki nilai akurasi prediksi yang tertinggi yaitu titik 347 mm.

Tabel 6 Nilai akurasi model regresi kuantil biner Bayes pada dataran sedang

Titik Ekstrim	Akurasi Dataran Sedang		
	1 tahun	2 tahun	3 tahun
325mm	91.67%	79.17%	86.11%
330mm	91.67%	87.50%	88.89%
335mm	91.67%	91.67%	83.33%
400mm (BMKG 2008)	91.67%	91.67%	88.89%

Tabel 7 Nilai akurasi model regresi kuantil biner Bayes pada dataran tinggi

Titik Ekstrim	Akurasi Dataran Tinggi		
	1 tahun	2 tahun	3 tahun
347mm	91.67%	95.83%	83.33%
352mm	91.67%	87.50%	83.33%
357mm	91.67%	87.50%	83.33%
400mm (BMKG 2008)	91.67%	87.50%	86.11%

Selanjutnya dilakukan validasi model yang telah didapatkan. Validasi merupakan tahapan yang penting karena mencerminkan keakuratan hasil prediksi dari model yang terbentuk. Kekonsistenan dari model regresi kuantil biner Bayes di masing-masing dataran dapat diketahui dari hasil pendugaan yang konsisten pada berbagai waktu yang berbeda. Akurasi prediksi dari model regresi kuantil biner Bayes untuk validasi 1 tahun dengan waktu yang berbeda disajikan dalam Tabel 8.

Tabel 8 Nilai akurasi model regresi kuantil biner Bayes untuk validasi 1 tahun

Data Pemodelan	Data Validasi	Akurasi		
		Dat. Rendah	Dat. Sedang	Dat. Tinggi
1981-2008	2009	91.67%	91.67%	91.67%
1981-2007	2008	100%	83.33%	83.33%
1981-2006	2007	91.67%	75%	83.33%
Rata-rata		94.45%	83.33%	86.11%
Simpangan Baku		0.048	0.083	0.048

Berdasarkan evaluasi dan validasi yang telah dilakukan, secara keseluruhan model regresi kuantil biner Bayes untuk semua dataran memiliki akurasi prediksi yang baik. Nilai akurasi prediksi di dataran rendah lebih tinggi daripada di dataran sedang dan tinggi. Nilai simpangan baku dari akurasi model regresi kuantil biner Bayes cukup kecil yaitu 0.048 untuk dataran rendah dan tinggi, dan 0.083 untuk dataran sedang. Hal tersebut menunjukkan bahwa model regresi kuantil biner Bayes di semua dataran cukup konsisten dalam menghasilkan prediksi 1 tahun kedepan.

@Hak cipta milik IPB University

IPB University



- Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
 2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.



V SIMPULAN DAN SARAN

5.1 Simpulan

Model regresi kuantil biner Bayes cukup akurat untuk memprediksi curah hujan ekstrim di Jawa Barat. Nilai ambang batas yang didapatkan dari metode MOS cukup baik untuk menentukan kategori curah hujan ekstrim. Model regresi kuantil biner Bayes di dataran rendah lebih baik dibandingkan dataran sedang dan tinggi. Model regresi kuantil biner Bayes lebih konsisten untuk prediksi 1 tahun terakhir di semua dataran.

5.2 Saran

Pada penelitian ini model *statistical downscaling* dengan regresi kuantil biner Bayes digunakan untuk prediksi curah hujan ekstrim dengan LASSO sebagai penaltinya. Peneliti menyarankan untuk membandingkan regresi kuantil biner Bayes terpenalti LASSO dengan fungsi penalti lain seperti *elastic net* atau *grouped LASSO*. Perbandingan tersebut diharapkan akan diperoleh akurasi yang lebih baik untuk pendugaan curah hujan ekstrim.

@Hak Cipta milik IPB University

DAFTAR PUSTAKA

- Alhamzawi R, Benoit DF, Yu K. 2013. Bayesian Lasso Binary Quantile Regression. *Journal of Computational Statistics*. 28(6): 2861-2873.
- Benoit DF, Van den Poel D. 2010. Binary Quantile Regression: A Bayesian Approach Based on the Asymmetric Laplace Distribution. *Journal of Applied Econometrics*. 27(7): 1174-1188.
- Benoit DF, Van den Poel D. 2017. BayesQR: A Bayesian Approach to Quantile Regression. *Journal of Statistical Software*, 76(7): 1–32.
- [BMKG] Badan Meteorologi dan Geofisika. 2008. Laporan Meterologi dan Geofisika. Jakarta (ID).
- Buhai S. 2004. *Quantile Regression: Overview and Selected Applications*. Rotterdam (NL): Tinbergen Institute and Erasmus University.
- Busuioc A, Chen D, Hellstrom C. 2001. Performance of Statistical Downscaling Models in GCM Validation and Regional Climate Change Estimates: Application for Swedish precipitation. *Int J Climatol*. 21(5): 557-578.
- Chen C, Wei Y. 2005. Computation issue for quantile regression. *The Indian Journal Statistics*. 67(2): 399-417.
- Fadli N, Wigena AH, Djuraidah A. 2019. Determination of General Circulation Model Grid Resolution to Improve Accuracy of Rainfall Prediction in West Java. *International Journal of Scientific and Research Publication*. 9(7): 835–838.
- Geman S, Geman D. 1984. Stochastic Relation, Gibbs Distribution, and The Bayesian Restoration of Image, *IEEE Transaction on Pattern Analysis and Machine Intelligence*. 6(PAMI-6): 721-741.
- Hendri EP, Wigena AH, Djuraidah A. 2019. Statistical Downscaling with Bayesian Quantile Regression to Estimate Extreme Rainfall in West Java. *International Journal of Science: Basic and Applied Research (IJSBAR)*. 47(2): 142–151.
- Hoff PD. 2009. *A First Course in Bayesian Statistical Methods*. New York (US): Springer.
- Koenker R. 2005. *Quantile Regression*. New York (US): Cambridge University Press.
- Koenker R, Bassett G. 1978. Regression Quantiles. *Econometrica*. 46(1): 33-50.
- Koenker R, Machado J. 1999. Goodness of Fit and Related Inference Processes for Quantile Regression. *Journal of the American Statistical Association*. 94(448): 1296-1310.
- Kozumi H, Kobayashi G. 2009. Gibbs Sampling Methods for Bayesian Quantile Regression. *Journal of Statistical Computation and Simulation*. 81(11): 1565-1578.
- Manik TK, Rosadi B, Nurhayati E. 2014. Mengkaji Dampak Perubahan Iklim Terhadap Distribusi Curah Hujan Lokal di Provinsi Lampung. *Forum Geografi*. 28(1): 73-86.
- Manurung AM, Wigena AH, Djuraidah A. 2018. GDP Threshold Estimation Using Measure of Surprise. *International Journal of Science: Basic and Applied Research (IJSBAR)*. 42(3): 16–25.

- Rahayu DN. 2015. Analisis Bayesian untuk Regresi Binari Kuantil Terpenalti dengan Menggunakan Algoritma *Gibbs Sampling* [skripsi]. Yogyakarta [ID]: Universitas Gadjah Mada.
- Stephenson DB, Kumar KR, Doblas-Reyes FJ, Royer JF, Hauvin FC. 1999. Extreme Daily Rainfall Events and Their Impact on Ensemble Forecasts of the Indian Monsoon. *Monthly Weather Review*. 127(9): 1954–1966.
- Stockdale TN, Busalacchi AJ, Harrison DE, Seager R. 1998. Ocean modeling for ENSO. *Journal of Geophysical research*. 103(C7):14325-14355.
- Tibshirani R. 1996. Regression Shrinkage and Selection via LASSO. *Journal of The Royal Statistical Society. Series B (Methodological)*. 58(1): 267-288.
- Wigena AH. 2006. Pemodelan *Statistical Downscaling* dengan Regresi Projection Pursuit untuk Prediksi Curah Hujan Bulanan [disertasi]. Bogor (ID): Institut Pertanian Bogor.
- Wigena AH, Djuraidah A. 2014. Quantile Regression in Statistical Downscaling to Estimate Extreme Monthly Rainfall. *Science Journal of Applied Mathematics and Statistics*. 2(3): 66-70.
- Yu K, Moyeed RA. 2001. Bayesian Quantile Regression. *Statistics & Probability Letter*. 54(4): 437-447.
- Zaikarina H, Djuraidah A, Wigena AH. 2016. Lasso and Ridge Quantile Regression using Cross Validation to Estimate Extreme Rainfall. *Global Journal of Pure and Applied Mathematics*. 12(3): 3305–3314.
- Zhou L, Wang Q, Fujita H. 2017. One Versus One Multi-class Classification Fusion Using Optimizing Decision Directed Acyclic Graph for Predicting Listing Status of Companies. *Inf. Fusion*. 36(1): 80–89

RIWAYAT HIDUP

Penulis dilahirkan di Bandar Lampung pada 05 Agustus 1996. Penulis merupakan anak kedua dari Bapak Hi. Syaifullah dan Ibu Hj. Sukartini serta memiliki 1 kakak laki-laki dan 1 adik perempuan. Pendidikan sekolah menengah atas ditempuh di SMAN 9 Bandar Lampung dan lulus pada tahun 2014. Penulis melanjutkan studi S1 di jurusan Matematika FMIPA Universitas Lampung dan lulus tahun 2018. Pada tahun yang sama, Penulis berkesempatan melanjutkan studi S2 di Institut Pertanian Bogor pada program studi Statistika Terapan.

Selama mengikuti program S2, Penulis aktif menjadi anggota Himpunan Mahasiswa Profesi (HIMPRO) Pascasarjana Statistika IPB sebagai anggota divisi Kewirausahaan periode 2018/2019 dan sebagai sekretaris divisi Kewirausahaan periode 2019/2020.

Hak Cipta Dilindungi Undang-undang

1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
 - a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
 - b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.