

PENDUGAAN CURAH HUJAN DENGAN TEKNIK *STATISTICAL DOWNSCALING* MENGGUNAKAN *CLUSTERWISE REGRESSION* SEBARAN TWEEDIE*

R Indriani Rakhmalia¹, Agus M Soleh^{2‡}, and Bagus Sartono³

¹Department of Statistics, IPB University, Indonesia, riza_indriani@apps.ipb.ac.id

²Department of Statistics, IPB University, Indonesia, agusms@apps.ipb.ac.id

³Department of Statistics, IPB University, Indonesia, bagusco@apps.ipb.ac.id

‡corresponding author

Indonesian Journal of Statistics and Its Applications (eISSN:2599-0802)

Vol 4 No 3 (2020), 473 - 483

Copyright © 2020 R Indriani Rakhmalia, Agus M Soleh, and Bagus Sartono. This is an open-access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Rainfall prediction is one of the most challenging problems of the last century. The statistical downscaling technique is one of the rainfall estimation techniques that can be used for that. This paper aims to predict the rainfall at several rain posts in West Java Province, Indonesia using clusterwise regression modeling with the Tweedie distribution. The data used in this paper are precipitation from Climate Forecast System Reanalysis (CFSR) version 2 and rainfall from Meteorological, Climatological, and Geophysical Agency Indonesia (BMKG). Data collection was carried out from January 2010 to December 2019 at the Bogor, Citeko, Jatiwangi, and Bandung rain posts. The best model is evaluated with the Root Mean Square Error Prediction. The RMSEP value of the Bogor rain post is 17.11 (three clusters), the Citeko rain post is 14.85 (two clusters), the Jatiwangi rain post is 15.26 (three clusters), and the Bandung rain post is 14.33 (two clusters). This model is able to make models and clusters well in daily rainfall applications.

Keywords: cluster-wise regression, statistical downscaling, tweedie distribution.

* Received May 2020; Accepted Oct 2020; Published online on Nov 2020

1. Pendahuluan

Curah hujan merupakan salah satu faktor yang berpengaruh langsung terhadap perubahan cuaca. Pendugaan curah hujan telah menjadi salah satu masalah yang paling ilmiah dan teknologi menantang seluruh dunia pada abad terakhir. Salah satu metode yang dapat digunakan untuk prediksi curah hujan adalah *Statistical Downscaling* (SD). Teknik *Statistical Downscaling* menggunakan data yang berskala lokal atau regional sebagai peubah tak bebas dan berskala global sebagai peubah bebas. Data yang berskala lokal atau regional adalah seperti data curah hujan dari stasiun BMKG, sedangkan data yang berskala global adalah data luaran *global climate model* (GCM). Data luaran GCM memiliki dimensi yang tinggi, sehingga pada umumnya terdapat multikolinearitas. Beberapa metode dapat mengatasi masalah multikolinearitas ini di antaranya seleksi peubah, *least absolute shrinkage and selection operator* (LASSO), analisis komponen utama, dan lain-lain.

Pemodelan curah hujan telah banyak dilakukan menggunakan *Statistical Downscaling* seperti Soleh et al. (2015) menggunakan metode LASSO dan *Principal Component Regression* (PCR) yang menunjukkan metode LASSO memberikan hasil yang lebih baik. Penelitian lain dengan menggunakan metode LASSO menghasilkan nilai RMSEP lebih kecil dari pada metode Gulud (Zaikarina, 2016). Pada tahun 2016 telah dilakukan penelitian dengan menggunakan teknik regularisasi *Elastic Net* dengan sebaran gamma. Nilai RMSEP yang dihasilkan lebih kecil jika ditambahkan peubah *dummy* pada gerombol curah hujan (Soleh, 2015; Soleh et al., 2016; Permatasari et al., 2016). Pada tahun 2020 penelitian *Statistical Downscaling* dengan analisis *clusterwise* atau regresi gerombol dilakukan oleh Syafruddin et al. (2019) Pada penelitian tersebut diperoleh jumlah gerombol 4 dengan sebaran peubah respon Gamma dan Normal dengan rata-rata RMSE 5 persen. Pada tahun 2020 penelitian *Statistical Downscaling* juga dilakukan oleh Hadiana, dengan menggunakan Regresi Kuadrat Terkecil Parsial (RKTP) dan *Multiclass Random Forest* menghasilkan rata-rata RMSEP sebesar 16,67 persen (Hadiana et al., 2020).

Berdasarkan penjelasan di atas data curah hujan yang memiliki nilai minimal nol dapat didekati dengan sebaran Tweedie. Pada sebaran Tweedie (Poisson dan Gamma) memiliki batas rentang ≥ 0 . Sehingga aplikasi sebaran Tweedie dapat digunakan pada data curah hujan, oleh karena itu perlu dilakukan metode *Clusterwise* dengan peubah respon mengikuti sebaran Tweedie. Curah hujan yang memiliki nilai minimal nol dapat didekati dengan sebaran Tweedie.

Penelitian ini bertujuan mengembangkan serta mengaplikasikan metode *Clusterwise* dengan sebaran Tweedie (Poisson dan Gamma) dengan Regularisasi *Elastic Net*. Ukuran kebaikan model yang digunakan adalah RMSEP. Semakin kecil nilai RMSEP maka semakin baik model yang diperoleh, begitu pula sebaliknya. Nilai RMSEP yang semakin besar dihasilkan akan semakin buruk model yang dihasilkan.

2. Metodologi

2.1 Data

Pada penelitian ini data yang digunakan merupakan data sekunder yaitu data curah hujan harian dari situs resmi Badan Meteorologi dan Klimatologi Geofisika (BMKG) di

Jawa Barat serta data presipitasi luaran GCM. Pos pengamatan hujan yang akan digunakan adalah pos pengamatan hujan Bogor, Citeko, Jatiwangi, serta Bandung. Data curah hujan BMKG sebagai peubah respon, serta data presipitasi luaran *Climate Forecast System Reanalysis* (CFSR) versi 2 sebagai peubah prediktor. Data presipitasi luaran CFSR ini adalah data dalam grid, dengan luas wilayah yang termasuk dalam pengamatan adalah 6 x 6 grid, dengan satu grid berukuran $0.5^\circ \times 0.5^\circ$ untuk setiap pos pengamatan. Berdasarkan ukuran grid tersebut maka akan ada 36 peubah prediktor yang digunakan. Data luaran CFSR maupun data BMKG yang digunakan mulai dari tahun 2010 – 2019. Data curah hujan lokal harian dapat diunduh dari <http://dataonline.bmkg.go.id>, sedangkan data luaran CFSR dapat diunduh dari <https://rda.ucar.edu>.

2.2 Metode Penelitian

Metode Penelitian dilakukan melalui tahapan sebagai berikut:

1. Mengembangkan regresi gerombol dengan sebaran Tweedie.

Regresi gerombol merupakan pengembangan dari analisis cluster dengan menggunakan model regresi sebagai dasar penggerombolan. Pada regresi gerombol data dibagi kedalam beberapa subpopulasi, sehingga terdapat beberapa model dengan parameter berbeda untuk setiap subpopulasi (Spath, 1979). Gerombol terbaik diperoleh dengan cara mengontrol masing-masing gerombol sehingga diperoleh model regresi terbaik (de Carvalho et al., 2010). Pendugaan model regresi dengan sebaran Tweedie dilakukan pada setiap gerombol yang terbentuk dengan menggunakan model linear terampat (GLM). Parameter pada model linear terampat diduga dengan memaksimalkan fungsi kepekatan. Fungsi hubung untuk distribusi Tweedie ada penelitian ini menggunakan fungsi log (Rahmawati & Darti, 2019). Berikut merupakan tahapan serta algoritma clusterwise dengan sebaran Tweedie.

- a. Inialisasi banyaknya gerombol (k), dan membagi masing-masing amatan ke dalam k gerombol secara acak.
- b. Memberikan label untuk masing-masing amatan, contoh dari 1 sampai n .
- c. Berikan inisial untuk masing-masing gerombol dengan amatan acak didalamnya
- d. Melakukan pemodelan awal dengan menggunakan data yang sudah di bentuk gerombol dengan sebaran peubah respon Tweedie, serta mereduksi dimensi pada peubah bebas. Teknik reduksi dimensi yang digunakan adalah elastic-net.

Sebaran *tweedie* salah satu sebaran dari keluarga sebaran eksponensial. Keluarga sebaran eksponensial merupakan sebaran respon dari model linear terampat. Jika Y mengikuti sebaran keluarga eksponensial memiliki rataan μ , fungsi ragam $V(\cdot)$, dan parameter dispersi ϕ dengan nilai $\phi > 0$, maka fungsi ragam dari Y adalah :

$$V(Y) = \phi V(\mu) \quad (1)$$

Sebaran *tweedie* merupakan kasus khusus dari keluarga sebaran eksponensial dengan $V(\mu) = \mu^p$ dan $V(Y) = \phi \mu^p$ (Dunn & Smyth, 2004), terdapat berbagai macam sebaran *tweedie* seperti sebaran Normal ($p = 0$), Gamma ($p = 1$), *Inverse Gaussian* ($p = 3$), dan sebaran compound Poisson-Gamma ($1 < p < 2$), dimana p adalah *power parameter*.

Proses elastic-net yang ada pada penelitian ini dilakukan secara bersamaan dengan proses regresi gerombol. Tujuan pada proses elastic net adalah mencari nilai λ yang optimal. *Elastic net* merupakan teknik regularisasi yang digunakan untuk mengatasi masalah-masalah multikolinearitas. *Elastic net* merupakan kombinasi regresi gulud dan LASSO (Zou & Hastie, 2005). *Elastic-Net* dapat melengkapi kekurangan pada regresi gulud dan LASSO. *Elastic net* mengestimasi koefisien dengan cara meminimumkan *loss function*.

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

Pada persamaan diatas disebut pinalti gulud ketika $\alpha = 0$ ($\sum_{j=1}^p \beta_j^2 \leq t$), sedangkan ketika $\alpha = 1$ ($\sum_{j=1}^p |\beta_j|$) maka akan disebut sebagai pinalti LASSO (Cahyani et al., 2016). Penalti *Elastic net* memberikan estimasi parameter sebagai berikut:

$$\hat{\beta} = \underset{\beta \in R}{\operatorname{argmin}} \left[\sum_{i=1}^t (y_i - x_i' \beta) + \lambda \sum_{j=1}^p [\alpha |\beta_j| + (1 - \alpha) \beta_j^2] \leq t \right] \quad (3)$$

- e. Menghitung galat untuk masing-masing model regresi yang terbentuk.
 - f. Melakukan perpindahan amatan gerombol ke dalam gerombol lainnya, jika galat yang diperoleh lebih kecil dari pada galat sebelum amatan tersebut berpindah maka amatan tersebut akan ikut berpindah ke dalam gerombol baru, sehingga akan terbentuk model regresi baru. Tetapi, jika amatan gerombol dipindahkan dan galat yang diperoleh lebih besar dari gerombol senelumnnya, maka amatan tersebut tidak dipindahkan. Langkah ini dilakukan untuk semua amatan.
 - g. Menduga parameter dari model regresi akhir yang diperoleh.
2. Membagi data menjadi dua bagian: data training dan data testing. Proporsi data training adalah 80% dan data testing adalah 20%.
 3. Melakukan pendugaan model untuk setiap gerombol menggunakan data training, dan melakukan pendugaan respon menggunakan data testing. Sebelum melakukan pendugaan data respon terlebih dahulu mencari jarak antara data observasi dengan centroid data masing-masing cluster. Metode jarak yang digunakan adalah jarak terdekat.

$$d_i = \sqrt{\sum_{j=1}^p (y_{ij} - \bar{x}_{ij})^2} \quad (4)$$

y_{ij} = amatan pada data *testing* peubah ke-j pada gerombol ke-i

\bar{x}_{ij} = rata-rata pada data *training* peubah ke-j gerombol ke-i.

4. Melakukan proses ulangan langkah 2-3 sebanyak 100 kali.
5. Menghitung ukuran kebaikan model dengan RMSEP
Metode evaluasi yang digunakan adalah *Root Mean Square Error Prediction* (RMSEP) dengan rumus RMSEP sebagai berikut:

$$RMSEP = \sqrt{\frac{\sum (y_{(-i)} - \hat{y}_{(-i)})^2}{n}} ; i = 1, 2, 3, \dots, n \quad (5)$$

Untuk $y_{(-i)}$ adalah data ke-i dari data *testing*, dan $\hat{y}_{(-i)}$ adalah data dugaan ke-i dari data *testing*.

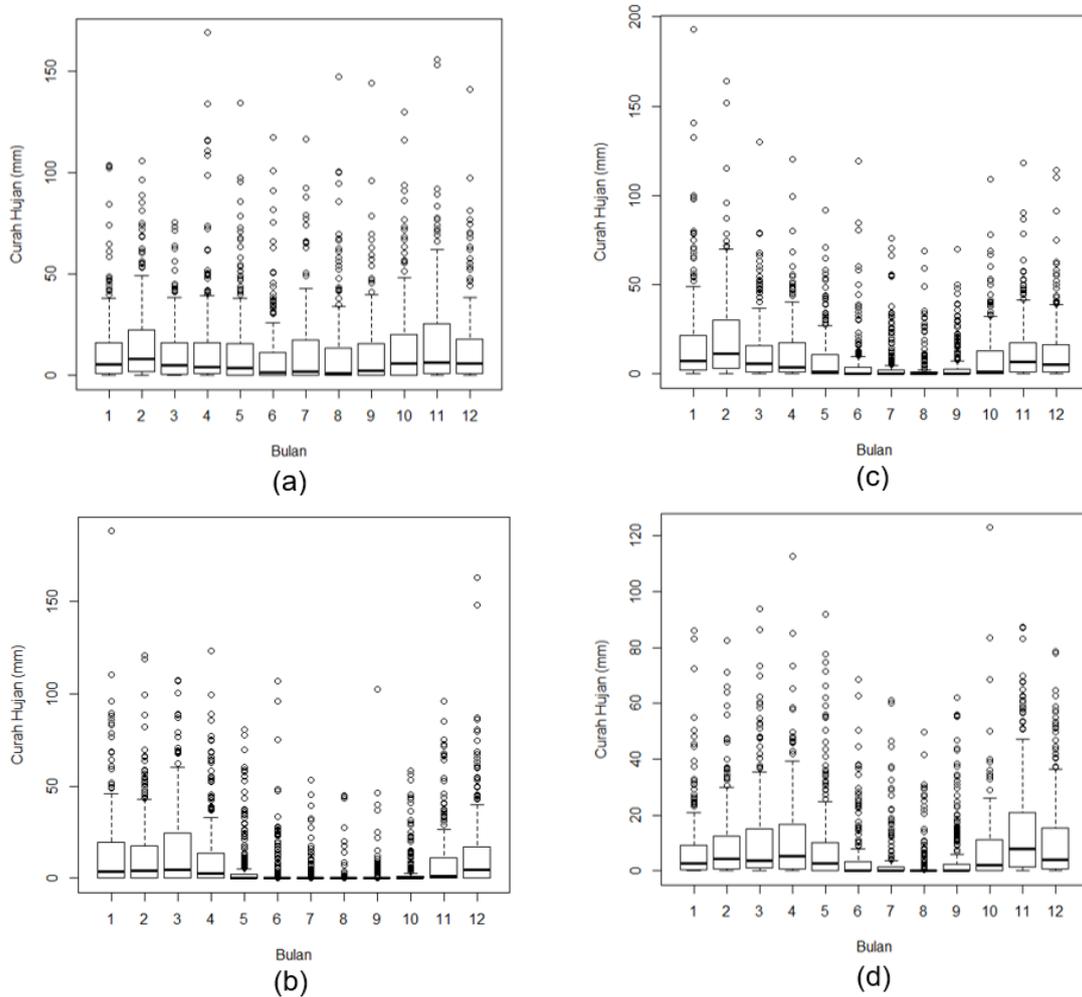
3. Hasil dan Pembahasan

3.1 Deskripsi Data

Pada penelitian ini data curah hujan harian yang diperoleh pada stasiun BMKG tidak lengkap, oleh karena itu perlu dilakukan pembersihan data. Pada data stasiun BMKG terdapat kode 8888 yang artinya data tidak terukur, dan kode 9999 yang artinya tidak terdapat data pada hari tersebut. Pembersihan data dilakukan dengan menghilangkan data missing, data dengan kode 8888, serta data dengan kode 9999. Oleh karena itu, banyaknya data pada setiap pos hujan berbeda-beda. Setelah data dibersihkan, maka dilakukan analisis secara deskriptif. Untuk analisis deskriptif data curah hujan lokal yang digunakan adalah pos hujan di Jawa Barat, seperti pos hujan Bogor, Citeko, Bandung, dan Jatiwangi. Masing-masing pos hujan tersebut memiliki tingkat curah hujan dan jumlah hari hujan yang berbeda-beda. Data yang digunakan pada penelitian ini adalah data harian diambil mulai dari 1 Januari 2010 hingga 31 Desember 2019. Berikut ini Gambar 1 merupakan gambaran secara umum curah hujan pada masing-masing pos pengamatan hujan.

Berdasarkan Gambar 1, pada pos hujan Bogor rata-rata curah hujan harian tidak banyak disekitar angka 0 mm/hari. Artinya, di Bogor terjadinya hujan lebih sering jika dibandingkan dengan pos hujan lainnya. Sebaran curah hujan harian untuk pos hujan Jatiwangi memiliki sebaran curah hujan disekitar 0 mm/hr yaitu mulai pada bulan Mei hingga Oktober. Pada bulan Mei hingga Oktober terjadi curah hujan yang rendah, sedangkan pada bulan Januari hingga April kemudian November dan Desember pada pos hujan Jatiwangi terjadi curah hujan yang cukup tinggi. Pada pos hujan Jatiwangi jumlah bulan dengan tingkat curah hujan rendah adalah paling banyak dibandingkan dengan pos hujan lainnya. Pada pos hujan Citeko sebaran curah hujan rendah terjadi pada bulan Juni hingga September. Sebaran hujan tertinggi pada bulan Februari hampir mencapai 200 mm/hr. Untuk pos hujan Bandung curah hujan rendah terjadi pada bulan Juni hingga September. Sepanjang tahun 2010 hingga 2019 curah hujan paling tinggi terjadi pada bulan Oktober di pos hujan Bandung. Jika dilihat untuk

sebaran curah hujan disemua pos hujan memiliki sebaran yang menjulur ke kanan, yang artinya lebih banyak curah hujan yang bernilai 0 mm/hr. Pada Gambar 1 di atas terlihat bahwa untuk pos hujan Bogor memiliki tingkat curah hujan yang berbeda jika dibandingkan dengan pos hujan lainnya.



Gambar 1: Sebaran Curah Hujan Pos Hujan (a) Bogor, (b) Jatiwangi, (c) Citeko, (d) Bandung.

3.2 Pengembangan Regresi Gerombol dengan Sebaran Tweedie

Terdapat dua kajian pada penelitian ini, yaitu simulasi pengembangan regresi gerombol dan aplikasinya. Pada kajian simulasi data dibangkitkan dengan sebaran tertentu yang kemudian dilakukan pengelompokkan. Tujuan dari kajian simulasi adalah untuk melihat kemampuan regresi gerombol dalam mengelompokkan data pada data bangkitan dengan sebaran tertentu. Pada bagian kajian simulasi dilakukan tiga skenario, yaitu skenario pertama data respon dengan sebaran normal, skenario kedua data respon dengan sebaran tweedie, dan skenario yang ketiga dengan sebaran campuran.

Setiap simulasi akan didapatkan rata-rata akurasi dari 100 kali pengulangan. Terdapat beberapa nilai parameter yang digunakan diantaranya tampak pada Tabel 1. Nilai rata-rata akurasi data respon bangkitan untuk gerombol satu normal dan gerombol dua normal kemudian diduga dengan model normal untuk gerombol satu dan dua (NN) sebesar 99.95% dengan nilai rata-rata RMSEP paling rendah yaitu sebesar 0.76. Nilai rata-rata akurasi dugaan untuk gerombol satu normal dan gerombol dua normal adalah yang terbesar jika dibandingkan dengan nilai rata-rata akurasi dugaan tweedie (TT) yang hanya berkisar 46.95% dan 50.05%, dan akurasi dugaan model tweedie gerombol satu normal gerombol dua (TN) sebesar 76.42% dan 73.97%. Pada Tabel 1 menunjukkan jika rata-rata akurasi dan juga rata-rata nilai RMSEP pada model normal paling besar, jika dibandingkan dengan model tweedie dan campuran. NN untuk sebaran normal gerombol satu dan dua, TT untuk sebaran tweedie gerombol satu dan dua, sedangkan TN untuk sebaran tweedie gerombol satu dan normal gerombol dua, yang selanjutnya akan disebut sebagai NN, TT, dan TN.

Tabel 1: Hasil akurasi pada data bangkitan sebaran normal

Normal - Normal	Akurasi (%)			RMSEP		
	TT	NN	TN	TT	NN	TN
$\beta_{0,1,2} = \{1,2,1\}$	46.95	99.95	76.42	26.67	0.76	1.01
$\beta_{0,1,2,3} = \{1,1,2,3\}$						
$\beta_{0,1,2,3} = \{1,1,1,1\}$	50.05	96.1	73.97	17.14	1.02	8.33
$\beta_{0,1,2} = \{1,3,2\}$						

Keterangan: NN untuk sebaran normal gerombol satu dan dua, TT untuk sebaran tweedie gerombol satu dan dua, TN untuk sebaran tweedie gerombol satu dan normal gerombol dua.

Data respon yang dibangkitkan mengikuti sebaran tweedie untuk gerombol satu dan sebaran normal untuk gerombol dua, dan kemudian diduga dengan sebaran normal untuk gerombol satu dan gerombol dua memiliki rata-rata akurasi sebesar 87.17% dan 84.12% (Tabel 2), sedangkan nilai rata-rata RMSEP sebesar 2.62 dan 2.71.

Data respon bangkitan tweedie untuk gerombol satu dan normal untuk gerombol dua, kemudian diduga dengan sebaran tweedie untuk gerombol satu dan dua memiliki rata-rata akurasi sebesar 65.19% dan 65.56%, sedangkan untuk nilai rata-rata RMSEP sebesar 6.30 dan 5.53. Dugaan data respon dengan sebaran normal untuk gerombol satu dan dua memiliki nilai akurasi sebesar 46.25% dan 46.24%, sedangkan nilai rata-rata RMSEP sebesar 2.86 dan 2.85. Data respon yang dibangkitkan dari sebaran tweedie normal akan cocok jika diduga dengan model tweedie normal karena memiliki nilai rata-rata akurasi paling besar.

Tabel 2: Hasil akurasi pada data bangkitan sebaran campuran

Tweedie - Normal	Akurasi (%)			RMSEP		
	TT	NN	TN	TT	NN	TN
$\beta_{1,2,3} = \{1,1,1\}, \phi = 1.3$	65.19	46.25	87.17	6.30	2.86	2.62
$\beta_{0,1,2,3} = \{1,3,3,2\}$						
$\beta_{1,2,3} = \{1,2,3\}, \phi = 1.3$	65.56	46.24	84.12	5.53	2.85	2.71
$\beta_{0,1,2,3} = \{1,3,3,2\}$						

Keterangan: NN untuk sebaran normal gerombol satu dan dua, TT untuk sebaran tweedie gerombol satu dan dua, TN untuk sebaran tweedie gerombol satu dan normal gerombol dua.

Terdapat dua skenario, yaitu skenario pertama dengan nilai $\phi=1$ dan skenario kedua dengan nilai $\phi=5$. Nilai rata-rata akurasi untuk dugaan dengan sebaran tweedie pada kedua gerombol yaitu sebesar 97.93% dan 96.20% (Tabel 3). Rataan akurasi dengan data bangkitan tweedie untuk gerombol satu dan dua, kemudian diduga dengan tweedie (Gerombol 1) dan normal (Gerombol 2) memiliki nilai sebesar 53.93% dan 53.65%. Sedangkan untuk dugaan dengan sebaran normal untuk gerombol satu dan dua memiliki rata-rata akurasi sebesar 83.33 dan 50.42%. Nilai rata-rata RMSEP untuk dugaan tweedie lebih kecil jika dibandingkan dengan nilai rata-rata RMSEP untuk dugaan normal dan campuran yaitu sekitar 1.36 dan 1.41. Oleh karena itu jika data respon diduga dengan asal sebaran awal, maka akan menghasilkan nilai akurasi tertinggi dibandingkan lainnya.

Tabel 3: Hasil akurasi pada data bangkitan sebaran tweedie

Tweedie - Tweedie	Akurasi (%)			RMSEP		
	TT	NN	TN	TT	NN	TN
$\phi=1, \beta_{1,2,3} = \{1,1,0\}$	97.93	83.33	53.93	1.36	3.52	3.53
$\phi=5, \beta_{1,2,3} = \{1,1,0\}$	96.20	50.42	53.65	1.41	3.53	3.45

3.3 Aplikasi Pemodelan Regresi Gerombol dengan Sebaran Tweedie

Curah hujan lokal harian yang berdistribusi Tweedie dimodelkan dengan menggunakan Clusterwise. Peubah penjelas yang digunakan adalah laju presipitasi data luaran GCM dengan 36 grid yang berada di atas pos pengamatan penelitian. Pemodelan dengan regresi gerombol merupakan pemodelan yang terdiri dari pembentukan gerombol, dan menentukan model antar gerombol secara bersama-sama. Jumlah gerombol yang digunakan adalah menurut kehendak peneliti. Pemodelan regresi gerombol dengan sebaran tweedie menghasilkan nilai RMSE yang lebih kecil dari rata-rata curah hujan di masing-masing pos hujan.

Tabel 4: Nilai RSME dengan Sebaran Tweedie.

Pos Hujan	Tanpa Penggerombolan	2 Gerombol	3 Gerombol	4 Gerombol
Bogor	20.39	18.12	17.11	19.72
Citeko	17.14	14.86	15.79	16.69
Jatiwangi	17.6	17.09	15.26	16.44
Bandung	19.09	14.33	14.43	15.80

Pada kasus *Statistical Downscaling* memiliki jumlah peubah prediktor yang banyak dan memiliki korelasi yang tinggi antar prediktor. Korelasi yang tinggi yang ada antar peubah prediktor perlu dilakukan penanganan. Teknik regularisasi elastic net merupakan salah satu cara dalam menangani korelasi tinggi antar peubah prediktor. Tabel 4 di atas menyajikan hasil nilai rata-rata RMSEP untuk tanpa penggerombolan dan dengan penggerombolan dengan dilakukan proses elastic net didalamnya. Tanpa penggerombolan menghasilkan nilai rata-rata RMSEP besar. Nilai rata-rata RMSEP tanpa penggerombolan untuk pos hujan Bogor mencapai 20.39, dan untuk pos hujan Bandung mencapai 19.09, untuk pos hujan Citeko 17.14, dan untuk pos hujan Jatiwangi 17.6. Untuk jumlah 4 gerombol nilai rata-rata RMSEP lebih besar jika dibandingkan dengan jumlah 2 gerombol dan 3 gerombol. Semakin kecil nilai RMSEP yang diperoleh maka akan semakin baik pula model tersebut. Semakin banyak jumlah gerombol yang terbentuk akan semakin banyak pula kombinasi model yang terbentuk, sehingga tidak efisien. Nilai rata-rata RMSEP paling kecil dengan jumlah gerombol dua yang terbaik adalah pos hujan Citeko dan Bandung yaitu sebesar 14.86, dan 14.33. Sedangkan nilai rata-rata RMSEP paling kecil dengan jumlah gerombol 3 terbaik adalah pos hujan Bogor 17.11 dan pos hujan Jatiwangi 15.79. Sebaran yang terbentuk pada model terbaik adalah sebaran Tweedie untuk masing-masing gerombol. Nilai rata-rata RMSEP diperoleh dari nilai rata-rata RMSEP dengan 100 kali perulangan.

4. Simpulan dan Saran

Pada penelitian ini dilakukan dua kajian yaitu kajian simulasi dengan tujuan melihat kemampuan regresi gerombol dalam mengelompokkan data, dan kajian yang kedua yaitu aplikasi regresi gerombol dengan data curah hujan harian. Pada kajian simulasi terlihat bahwa regresi gerombol mampu menggerombolkan data dengan baik,

sehingga regresi gerombol dengan model tweedie dapat diaplikasikan kedalam data *real*. Pendugaan curah hujan menggunakan data GCM yang memiliki korelasi tinggi antar peubah prediktor dengan metode elastic net. Nilai rata-ran RMSEP yang diperoleh dengan penggerombolan baik 2 gerombol, 3 gerombol, maupun 4 gerombol lebih baik jika dibandingkan dengan tanpa penggerombolan. Jumlah gerombol terbaik untuk pos hujan Bogor dan Jatiwangi adalah 3 gerombol dengan nilai RMSEP sebesar 17.11 dan 15.26. Untuk pos hujan Citeko dan Bandung jumlah gerombol terbaik yang terbentuk adalah dua gerombol dengan nilai rata-ran RMSEP masing-masing adalah 14.86 dan 14.33. Sehingga, pada penelitian ini pemodelan regresi gerombol dengan sebaran tweedie mampu menghasilkan gerombol dengan baik.

Daftar Pustaka

- Cahyani, T. B. N., Wigena, A. H., & Djuraidah, A. (2016). Quantile regression with elastic-net in statistical downscaling to predict extreme rainfall. *Global Journal of Pure and Applied Mathematics*, 12(4): 3517–3524.
- de Carvalho, F. D. A., Saporta, G., & Queiroz, D. N. (2010). A clusterwise center and range regression model for interval-valued data. In *Proceedings of COMPSTAT'2010* (pp. 461-468). Physica-Verlag HD.
- Dunn, P. K., & Smyth, G. K. (2004). Series evaluation of Tweedie exponential dispersion model densities. *Statistics and Computing*, 15: 267–280.
- Hadiana, R., Soleh, A. M., & Sartono, B. (2020). Two-Stage Statistical Downscaling Modeling with Multi-Class Random Forest on Rainfall Prediction. *Proceedings of the 1st International Conference on Statistics and Analytics*. Bogor (ID): Departemen Statistika IPB.
- Spath, H. (1979). Algorithm 39: Clusterwise Linear Regression. *Computing*, 22: 367–373.
- Permatasari, S., Djuraidah, A., & Soleh, A. M. (2016). *Statistical Downscaling with Gamma Distribution and Elastic Net Regularization*. 121–129.
- Rahmawati, M., & Darti, I. (2019). Pencadangan Klaim IBNR dengan Pendekatan Distribusi Keluarga Tweedie pada Generalized Linear Model. *Journal of Mathematics and Its Applications*, 16(1): 11–25.
- Soleh, A. M., Wigena, A. H., Djuraidah, A., & Saefuddin, A. (2015). Statistical downscaling to predict monthly rainfall using linear regression with L1 regularization (LASSO). *Applied Mathematical Sciences*, 9: 5361–5369.
- Soleh, A. M., Wigena, A. H., Djuraidah, A., & Saefuddin, A. (2016). Gamma distribution linear modeling with statistical downscaling to predict extreme monthly rainfall in Indramayu. In *2016 12th International Conference on Mathematics, Statistics, and Their Applications (ICMSA)* (pp. 134-138). IEEE.
- Soleh, A. M. (2015). Pemodelan linier sebaran gamma dan pareto terampat dengan regularisasi L1 pada statistical down scaling untuk pendugaan curah hujan bulanan [Disertasi]. *Institut Pertanian Bogor, Bogor (ID)*.

Syafruddin, R., Soleh, A. M., & Wigena, A. H. (2019). Clusterwise Regression Model Development with Gamma Distribution. In *Proceedings of the 1st International Conference on Statistics and Analytics*. Bogor (ID): Departemen Statistika IPB.

Zaikarina, H. (2016). *Regresi kuantil lasso dan gulud dengan validasi silang untuk prediksi curah hujan ekstrim* (Thesis). Bogor (UI): IPB University.

Zou, H., & Hastie, T. (2005). Regression Shrinkage and Selection via the Elastic Net, with Applications to Microarrays. *Journal of the Royal Statistical Society: Series B*, 67(2).