

PENINGKATAN PRESISI DUGAAN BERAT GABAH MELALUI PROSES SELEKSI PEUBAH DALAM PEMBELAJARAN MESIN STATISTIKA

Muhlis Ardiansyah^{1,2}, Khairil Anwar Notodiputro^{2*}, Bagus Sartono²

¹BPS Kabupaten Kotawaringin Timur, Kalimantan Tengah, Indonesia

² Department Statistika, IPB University, Bogor, Indonesia

khairil@apps.ipb.ac.id

Abstrak. Pandemi COVID-19 berpotensi menyebabkan turunnya tingkat respon beberapa survei yang dilakukan oleh lembaga penyedia statistik resmi di berbagai negara termasuk salah satunya Survei Ubinan oleh Badan Pusat Statistik Indonesia. Pengukuran berat gabah dari Survei Ubinan pada masa pandemi tidak selalu berhasil dilakukan sehingga terjadi nonrespon. Dalam penelitian ini, dicari solusi baru untuk mengatasi masalah nonrespon yaitu dengan menduga data nonrespon berat gabah berdasarkan beberapa peubah yang diperoleh dari hasil wawancara dengan petani setelah kebijakan pembatasan sosial berakhir. Dalam penelitian dicoba berbagai metode seleksi peubah yaitu Stepwise, LASSO, Elastic Net, Adaptive LASSO, dan Relaxed LASSO guna menemukan model terbaik dalam pendugaan berat gabah berdasarkan data Survei Ubinan tahun 2019 di Kalimantan Tengah. Gugus data dibagi menjadi dua, yaitu data latih dan data uji. Pengambilan 75 persen data latih dilakukan secara acak dan diulang sebanyak 100 kali dan setiap kali divalidasi menggunakan data uji. Selanjutnya peubah yang terseleksi digunakan dalam model *Quantile Regression Forest* (QRF). Hasilnya menunjukkan bahwa Adaptive-QRF memberikan kinerja terbaik dengan nilai RMSE tekecil. Peubah terpilih yang berpengaruh nyata terhadap berat gabah hasil penyeleksian Adaptive LASSO adalah varietas benih, cara penanaman (monokultur atau tumpangsari), sistem penanaman (jajar legowo atau bukan), terkena serangan hama atau tidak, lokasi penanaman, dan bulan panen.

Kata Kunci: *Adaptive LASSO, Produktivitas Padi, Survei Ubinan, Quantile Regression Forest.*

1. Pendahuluan

Pandemi COVID-19 membawa dampak perubahan di segala aspek kehidupan baik individu, masyarakat, maupun para pelaku usaha. Lembaga penyedia statistik resmi (*Official Statistics*) di berbagai negara pun ikut terkena dampak penyebaran virus ini termasuk Badan Pusat Statistik (BPS) selaku lembaga penyedia statistik resmi di Indonesia. Adanya kebijakan Pembatasan Sosial Berskala Besar (PSBB) sejak Maret 2020 menyebabkan sulitnya untuk melakukan pengumpulan data di lapangan. Kebijakan PSBB berpotensi menyebabkan turunnya tingkat respon pada beberapa survei.

BPS berusaha mengoptimalkan penggunaan teknologi seperti pemanfaatan Big Data dan pengumpulan data secara online. Permasalahannya, tidak semua data

dapat dikumpulkan secara *online* melainkan tetap harus berkunjung ke lapangan. Salah satu survei BPS yang tidak dapat diganti dengan pengumpulan data secara *online* adalah Survei Ubinan. Survei ini bertujuan untuk mendapatkan angka produktivitas padi. Data ini dibutuhkan untuk mendukung tersedianya informasi tentang capaian kedua *Sustainable Development Goals* (SDGs). Hingga sekarang, angka produktivitas padi tidak dapat dihitung menggunakan teknologi penginderaan jauh maupun secara *online*. Data produktivitas padi juga tidak dapat diperoleh melalui wawancara pasca panen. Berdasarkan Ardiansyah dan Tofri (2019), data produktivitas padi hasil wawancara (pengakuan petani) lebih rendah (*underestimate*) dibanding hasil pengukuran plot ubinan.

Dalam penelitian ini, dicari solusi untuk mengatasi masalah berkurangnya data akibat nonrespon. Hilangnya data akibat nonrespon mengakibatkan potensi bias dalam pendugaan parameter dan memperlemah generalisasi dari hasil penelitian. Metode imputasi diajukan untuk mengatasi permasalahan ini yaitu dengan mengganti nilai data nonrespon berdasarkan informasi peubah-peubah prediktor. Setelah seluruh data nonrespon diimputasi, maka gugus data dianalisis dengan menggunakan teknik baku sebagaimana data lengkap. Metode imputasi menggunakan pembelajaran mesin yaitu *Quantile Regression Forest* (QRF).

Dalam penelitian dicoba berbagai metode seleksi peubah untuk meningkatkan kinerja *Quantile Regression Forest* (QRF) dalam menduga berat gabah Survei Ubinan. Seleksi peubah merupakan tahapan penting untuk mendapatkan model yang terbaik. Metode seleksi peubah yang diperbandingkan adalah Stepwise, LASSO, Elastic Net, Adaptive LASSO, dan Relaxed LASSO. Metode *stepwise* menggabungkan langkah yang dilakukan pada metode *forward selection* dan *backward elimination*. Proses memasukkan dan mengeluarkan peubah prediktor dilakukan bergantian hingga diperoleh model dengan nilai Akaike's Information Criterion (AIC) minimum (Morozova *et. al.* 2015). Semakin kecil nilai AIC maka model yang diperoleh semakin baik. $AIC(k) = -2 (\text{Log Likelihood dari } \hat{\beta}_k) + 2p$.

Metode LASSO (*Least Absolute Shrinkage and Selection Operator*) mampu menyusutkan koefisien regresi tepat nol sehingga dapat digunakan untuk menyeleksi peubah. Metode ini dikembangkan pertama kali oleh Tibshirani pada 1996 (Kim *et. al.* 2019). Penduga koefisien regresi LASSO disusutkan ke arah nol seiring dengan peningkatan nilai λ . Parameternya diduga dengan $\min_{\beta} \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j|$ dimana λ adalah hyper-parameter non negatif. Hyper-parameter λ dioptimasi menggunakan *Cross Validation* (Hastie *et. al.* 2015).

Zou dan Hastie (2005) memperkenalkan regularisasi baru yaitu Elastic-Net yang merupakan penalti gabungan antara regresi ridge dan LASSO. Pendugaan koefisien regresi ridge dilakukan dengan penalti yaitu $\sum_{j=1}^p \beta_j^2 \leq t, t \geq 0$ sedangkan pada LASSO digunakan penalti $\sum_{j=1}^p |\beta_j| \leq t, t \geq 0$ dan penalti Elastic-Net $\lambda \sum_{j=1}^p [(1 - \alpha) \beta_j^2 + \alpha |\beta_j|]$ dimana β_j adalah koefisien penduga regresi, dengan t merupakan suatu besaran yang mengendalikan besarnya penyusutan dengan nilai $t \geq 0$ (Zou dan Hastie 2005). Besaran α (alpha) akan menjadi penalti ridge apabila nilai $\alpha = 0$ dan menjadi penalti LASSO apabila nilai $\alpha = 1$. Penduga parameter pada Elastic-Net diperoleh dengan: $\min_{\beta} \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p [(1 - \alpha) \beta_j^2 + \alpha |\beta_j|]$ dengan λ adalah hyper-parameter non negatif (Zou dan Hastie 2005).

Adaptive Lasso adalah metode yang mampu menyusutkan beberapa koefisien mendekati nol hingga tepat nol dengan menggunakan bobot adaptive. Pendugaan parameter pada Adaptif Lasso yaitu: $\min_{\beta} \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p w_j |\beta_j|$ dengan $w_j = \frac{1}{|\hat{\beta}|^r}$ dimana r adalah bilangan positif dan $\hat{\beta}$ diduga dengan regresi ridge yaitu dengan $\min_{\beta} \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p \beta_j^2$ (Zou 2006). Sedangkan, Relaxed LASSO melakukan pemilihan model dan penyusutan koefisien dengan dua hyper-parameter λ dan ϕ . Pendugaan parameter pada Relaxed Lasso yaitu: $\min_{\beta} \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \{\beta_j \times M_j\})^2 +$

$\phi \lambda \sum_{j=1}^p w_j |\beta_j|$ dimana M_j mengindikasikan bahwa peubah prediktor j diduga sebagai non-zero atau zero dengan LASSO (Meinshausen 2007). Notasi Tabel 1 digunakan pada makalah ini.

Tabel 1. Notasi yang digunakan

Notasi	Deskripsi
n	banyaknya observasi
p	banyaknya peubah prediktor (fitur)
$i = 1, 2, \dots, n$	urutan observasi
$j = 1, 2, \dots, p$	urutan peubah prediktor
y_i	peubah respon pada observasi ke- i
x_{ij}	nilai peubah prediktor j pada observasi ke- i
β_j	nilai koefisien parameter yang melekat pada prediktor j

Jadi, dari uraian tentang metode seleksi pengembangan LASSO nampak bahwa perbedaannya terletak pada penalti yang ditambahkan dalam model yaitu

$$\min_{\beta} \sum_{i=1}^n (y_i - \sum_{j=1}^p x_{ij} \beta_j)^2 + \text{penalti.}$$

2. Metode Penelitian

2.1 Data



Gambar 1. Tahapan pelaksanaan Survei Ubinan di Provinsi Kalimantan Tengah
2020

Data diperoleh dari hasil Survei Ubinan tahun 2019 di Provinsi Kalimantan Tengah. Ubinan adalah cara untuk memperkirakan hasil panen tanaman padi melalui pengukuran dan penimbangan hasil panen. Gambaran pelaksanaan Survei Ubinan dapat dilihat pada Gambar 1.

Ukuran satu plot ubinan adalah $2.5 \times 2.5 \text{ m}^2$ atau jika dihitung luas satu plot ubinan adalah 6.25 m^2 . Pada pelaksanaan Survei Ubinan, petugas mengambil secara acak satu plot padi di petak lahan sawah maupun bukan sawah dengan metode *simple random sampling*. Selanjutnya, petugas memanen tanaman padi untuk ditimbang dalam bentuk gabah kering panen (GKP). Berdasarkan hasil ubinan maka dapat diperkirakan produktivitas tanaman padi suatu daerah. Selanjutnya, petugas melakukan wawancara dengan petani untuk memperoleh informasi tambahan yang menghasilkan peubah-peubah yang akan digunakan untuk pemodelan yang disajikan pada Tabel 2.

Tabel 2. Peubah yang digunakan untuk pemodelan

Peubah	Nama Peubah	Keterangan	Sumber berasal dari Survei Ubinan 2019
y_i	Berat gabah kering panen	Satuan kg per $2.5 \times 2.5 \text{ m}^2$	Rincian 701
x_{11i}, x_{12i}	Subround	3 Kategori dikurangi 1	Subround
$x_{21i}, x_{22i}, x_{23i}, x_{24i}, x_{25i}, x_{26i}, x_{27i}, x_{28i}, x_{29i}, x_{210i}, x_{211i}$	Bulan	12 Kategori dikurangi 1	Bulan
$x_{31i}, x_{32i}, x_{33i}, x_{34i}, x_{35i}, x_{36i}, x_{37i}, x_{38i}, x_{39i}, x_{310i}, x_{311i}, x_{312i}, x_{313i}$	Kabupaten	14 Kategori dikurangi 1	R102
x_{4i}	Jenis tanaman pangan	1. Padi Sawah 0. Padi Ladang	R113
x_{5i}	Jenis lahan	1. Lahan irigasi 0. Bukan lahan irigasi	R601
x_{6i}	Luas petak	Numerik	R604
x_{7i}	Cara penanaman	1. Monokultur 0. Tumpangsari	R605

x_{8i}	Sistem penanaman	1. Jajar legowo 0. Bukan jajar legowo	R606a
x_{9i}	Program bantuan pemerintah atau bukan	1. Bantuan pemerintah 0. Non-bantuan pemerintah	R607
x_{10i}	Banyak benih perpetak	Numerik	R608 per R604
x_{11i}	Varietas benih	1. Hirida 0. Inbrida	R609
x_{12i}	Banyaknya pupuk urea kg/ha	Numerik	R610_1 per R604
x_{13i}	Banyaknya TSP/SP36 kg/ha	Numerik	R610_2 per R604
x_{14i}	Banyaknya pupuk KCL kg/ha	Numerik	R610_3 per R604
x_{15i}	Banyaknya pupuk NPK kg/ha	Numerik	R610_4 per R604
x_{16i}	Banyaknya pupuk kompos kg/ha	Numerik	R610_5 per R604
x_{17i}	Banyaknya pupuk organik cair kg/ha	Numerik	R610_6 per R604
x_{18i}	Banyak rumpun per plot	Numerik	R702
x_{19i}	Anggota kelompok tani atau bukan	1. Ya 0. Tidak	R803a
x_{20i}	Serangan OPT	1. Terserang 0. Tidak	R804b
x_{21i}	Dampak perubahan iklim	1. Terdampak 0. Tidak terdampak	R805b
x_{22i}	Kecukupan air	1. Cukup 0. Berlebih/kurang	R806b
x_{23i}	Titik Longitude	Numerik	R303
x_{24i}	Titik Latitude	Numerik	R303
$x_{12i} * x_{13i}$	Interaksi urea* TSP/SP36	Numerik	R610_1* R610_2
$x_{12i} * x_{14i}$	Interaksi urea* KCL	Numerik	R610_1* R610_3
$x_{12i} * x_{15i}$	Interaksi urea* NPK	Numerik	R610_1* R610_4
$x_{12i} * x_{16i}$	Interaksi urea* kompos	Numerik	R610_1* R610_5
$x_{12i} * x_{17i}$	Interaksi urea* organik cair	Numerik	R610_1* R610_6
$x_{13i} * x_{14i}$	Interaksi TSP/SP36* KCL	Numerik	R610_2* R610_3

$x_{13i} * x_{15i}$	Interaksi TSP/SP36* NPK	Numerik	R610_2* R610_4
$x_{13i} * x_{16i}$	Interaksi TSP/SP36* kompos	Numerik	R610_2* R610_5
$x_{13i} * x_{17i}$	Interaksi TSP/SP36* organik cair	Numerik	R610_2* R610_6
$x_{14i} * x_{15i}$	Interaksi KCL* NPK	Numerik	R610_3* R610_4
$x_{14i} * x_{17i}$	Interaksi KCL* organik cair	Numerik	R610_3* R610_6
$x_{15i} * x_{16i}$	Interaksi NPK* kompos	Numerik	R610_4* R610_5
$x_{15i} * x_{17i}$	Interaksi NPK* organik cair	Numerik	R610_4* R610_6
$x_{16i} * x_{17i}$	Interaksi kompos* organik cair	Numerik	R610_5* R610_6

Peubah yang ingin diduga adalah y_i dengan memanfaatkan informasi tambahan sebanyak 61 peubah prediktor (p) yang terdiri dari peubah dummy, peubah numerik, dan beberapa interaksi antarpeubah. Berbagai peubah prediktor tersebut dimanfaatkan untuk pendugaan berat gabah plot ubinan $2.5 \times 2.5 m^2$.

2.2 Model Pembelajaran Mesin Statistika

Peubah hasil seleksi digunakan untuk pemodelan pada pembelajaran mesin statistika. Dalam penelitian ini ditetapkan menggunakan *Quantile Regression Forest* (QRF). Model QRF adalah penggabungan antara model regresi kuantil dan random forest yang merupakan model non-parametrik dengan algoritma pembelajaran mesin statistika. Model QRF pertama kali diusulkan oleh Meinshausen (2006). Misalkan $q_\tau(Y|X = x)$ adalah kuantil ke- τ dari peubah Y bersyarat $X = x$ dengan $\tau \in (0,1)$. Untuk $X = x$, fungsi sebaran bersyarat $F(y|X = x)$ adalah peluang bahwa Y lebih kecil atau sama dengan $y \in R$, yaitu: $F(y|X = x) = P(Y \leq y|X = x)$. Kuantil ke- τ yaitu $q_\tau(Y|X = x)$ didefinisikan sebagai y sedemikian rupa sehingga $q_\tau(Y|X = x) = \inf\{y: F(y|X = x) \geq \tau\}$.

Model QRF menduga $q_\tau(Y|X = x)$ dengan membangun random forest yang dibangun dari contoh bootstrap (Fang *et al.* 2018). Kemudian sebaran bersyarat diduga oleh distribusi tertimbang dari peubah respon. Misalkan sebaran

Y bersyarat $X = x$ berdasarkan pohon ψ dan daun yang berisi x dilambangkan dengan $L_n(x, \psi)$, maka bobot $\omega_i(x, \psi)$ diberikan oleh $\omega_i(x, \psi) = \frac{I(X_i \in L_n(x, \psi))}{|\{j: X_j \in L_n(x, \psi)\}|}$. Banyaknya pohon T adalah $\psi_1, \dots, \psi_T$ dan $\omega_i(x)$ adalah rata-rata dari $\omega_i(x, \psi)$ untuk semua pohon, maka $\omega_i(x) = \frac{1}{T} \sum_{t=1}^T \omega_i(x, \psi_t)$. Penduga $\hat{F}(y|X=x)$ diberikan oleh $\hat{F}(y|X=x) = \sum_{i=1}^n \omega_i(x) I(Y_i \leq y)$. Kemudian kuantil ke- τ $q_\tau(Y|X=x)$ diprediksi dengan $\hat{q}_\tau(Y|X=x) = \inf\{y: \hat{F}(y|X=x) \geq \tau\}$ (Fang *et al.* 2018). Dalam makalah ini, digunakan algoritma Meinshausen (2017) melalui Package ‘quantregForest’ versi 1.3-7 pada program R.

Pendugaan parameter dilakukan dengan terlebih dahulu membagi data menjadi dua bagian yaitu data latih dan data uji yang diambil secara acak. Parameter yang dimaksud dalam penelitian ini adalah nilai berat gabah pada plot ubinan. Pendugaan nilai parameter y_i pada data uji menggunakan model QRF dengan seluruh peubah dan QRF dengan peubah hasil seleksi. Perbandingan model dilakukan dengan melihat tingkat presisi prediksi antara keenam model berdasarkan

$$\text{Root Mean Square Error} \left(RMSE = \sqrt{\frac{\sum_{i=1}^{n_{data\ uji}} [y_i - \hat{y}_i]^2}{n_{data\ uji}}} \right).$$

Teknik statistik *paired samples t-test* dipilih untuk membandingkan objek yang sama karena *set.seed* pengambilan contoh data latih adalah sama untuk berbagai metode yang dicobakan. Masing-masing data latih mendapatkan perlakuan yang berbeda yaitu metode *stepwise-QRF*, *LASSO-QRF*, *Elastic Net-QRF*, *Adaptive-QRF*, dan *Relaxed-QRF*. Teknik ini dipilih dengan alasan himpunan data latih dianggap sebagai objek yang diberi enam perlakuan yang berbeda. Statistik uji pada *paired samples t-test* yaitu: $t_{hitung} = \frac{\bar{y}_d - \mu_0}{S_d / \sqrt{n}}$ dengan

$y_{di} = y_i^1 - y_i^2$, $\bar{y}_d = \sum_{i=1}^n y_{di} / n$, $S_d^2 = \frac{\sum_{i=1}^n (y_{di} - \bar{y}_d)^2}{n-1}$, $v = n - 1$, dan n adalah banyaknya pasangan objek yang diperbandingkan. Uji signifikansi dengan kaidah: tolak H_0 jika $t^* < -t(\alpha; v)$. Statistik uji *Paired samples t-test* mengikuti distribusi *student-t*.

2.3 Analisis Data

Analisis data pada penelitian ini dibagi menjadi dua tahap, yaitu: tahap persiapan data dan tahap pemodelan. Tahap persiapan data (*preprocessing*) yaitu membuat vektor peubah respon berukuran $n \times 1$ dan matriks peubah prediktor berukuran $n \times p$ yang berisi peubah dummy, peubah numerik, dan beberapa interaksi antar peubah. Kemudian, data set dibagi menjadi data latih dan data uji. Data latih diambil secara acak dan diulang sebanyak 100 kali dan setiap ulangan dilakukan validasi menggunakan data uji.

Tahap pemodelan yaitu melakukan seleksi peubah menggunakan metode Stepwise, LASSO, Elastic Net, Adaptive LASSO, dan Relaxed LASSO. Titik koordinat longitude dan latitude tidak disertakan dalam proses penyeleksian tetapi langsung dimasukkan dalam model QRF. Titik koordinat umumnya akan menyusut (*shrinkage*) ke nol padahal saat diterapkan pada mesin pembelajaran, titik koordinat mampu meningkatkan akurasi dan presisi pendugaan. Hal ini berdasarkan Ardiansyah *et al* (2018) bahwa titik koordinat mampu meningkatkan akurasi dan presisi pendugaan data produktivitas padi. Kemudian, membangun model QRF dengan seluruh peubah dan model QRF menggunakan peubah hasil seleksi ditambah titik koordinat longitude dan latitude. Kemudian, menduga nilai y_i berdasarkan peubah prediktor pada data uji. Selanjutnya dilakukan pengujian beda rata-rata RMSE (*paired samples t-test*) antar model dari 100 ulangan untuk menyimpulkan metode seleksi dengan presisi terbaik.

3. Hasil dan Pembahasan

Data lengkap yang terkumpul pada pelaksanaan Survei Ubinan tanaman padi di Kalimantan Tengah Tahun 2019 adalah sebanyak 585 titik plot contoh yang tersebar di 14 kabupaten/kota se-Kalimantan Tengah. Selanjutnya dilakukan perbandingan RMSE antarmodel berdasarkan pada seluruh peubah yang dapat dilihat pada Tabel 3.

Tabel 3. Perbandingan RMSE antarmodel berdasarkan pada seluruh peubah

Model	RMSE			
	15%Testing Data	20%Testing Data	25%Testing Data	Rata-rata
Regresi LASSO	0.8207	0.8343	0.8400	0.8316
Regresi Elastic Net	0.8211	0.7948	0.8320	0.8160
Regresi Adaptive LASSO	0.8495	0.8170	0.8515	0.8394
Regresi Relaxed LASSO	0.8089	0.8418	0.7958	0.8155
Quantile Regression Forest	0.7311	0.7198	0.7542	0.7350

Dapat dilihat pada Tabel 3, presisi terbaik untuk memodelkan data produktivitas padi di Kalimantan Tengah adalah menggunakan model Quantile Regressions Forest dibandingkan Regresi LASSO, Regresi Elastic Net, Regresi Adaptive LASSO, dan Regresi Relaxed LASSO. Dalam pendugaan nilai peubah respon menggunakan mesin pembelajaran statistika, seringkali jumlah peubah banyak justru berpotensi menurunkan tingkat presisi dugaan. Untuk meningkatkan presisi dugaannya maka dilakukan penyeleksian peubah menggunakan metode Stepwise, LASSO, Elastic Net, Adaptive, dan Relaxed. Banyaknya peubah yang terseleksi dapat dilihat pada Tabel 4.

Tabel 4. Peubah hasil seleksi menggunakan metode stepwise, LASSO, elastic net, adaptive, dan relaxed LASSO

Metode Seleksi	Jumlah Peubah Prediktor (p)	Nama Prediktor
Stepwise	30	$x_{11i}, x_{21i}, x_{26i},$ $x_{31i}, x_{32i}, x_{33i}, x_{34i}, x_{35i}, x_{36i},$ $x_{37i}, x_{38i}, x_{39i}, x_{310i}, x_{312i}, x_{313i}, x_{4i}, x_{6i},$ $x_{7i}, x_{11i}, x_{13i}, x_{15i}, x_{16i}, x_{18i}, x_{20i}, x_{21i},$ $x_{23i}, x_{24i}, x_{12i} * x_{16i}, x_{15i} * x_{16i}, x_{16i} * x_{17i}$
LASSO	23	$x_{21i}, x_{23i}, x_{26i}, x_{33i}, x_{37i}, x_{38i}, x_{311i}, x_{312i},$ $x_{313i}, x_{4i}, x_{6i}, x_{7i}, x_{8i}, x_{9i}, x_{11i}, x_{13i}, x_{15i},$ $x_{16i}, x_{18i}, x_{20i}, x_{23i}, x_{24i}, x_{15i} * x_{16i}$
Elastic Net	15	$x_{21i}, x_{33i}, x_{37i}, x_{311i}, x_{4i}, x_{6i}, x_{7i}, x_{8i},$ $x_{11i}, x_{13i}, x_{15i}, x_{20i}, x_{23i}, x_{24i}, x_{15i} * x_{16i}$
Adaptive LASSO	11	$x_{23i}, x_{33i}, x_{37i}, x_{38i}, x_{311i}, x_{7i}, x_{8i}, x_{11i},$ $x_{20i}, x_{23i}, x_{24i}$

Relaxed LASSO	16	$x_{23i}, x_{33i}, x_{37i}, x_{38i}, x_{311i}, x_{4i}, x_{6i}, x_{7i}, x_{8i}, x_{11i}, x_{13i}, x_{15i}, x_{18i}, x_{20i}, x_{23i}, x_{24i}$
---------------	----	---

Dapat dilihat dari Tabel 4 bahwa hasil seleksi peubah menggunakan metode stepwise terdapat 30 peubah yang signifikan, metode LASSO memilih 23 peubah prediktor, Relaxed-LASSO memilih 16 peubah prediktor, Elastic Net memilih 15 peubah prediktor dan Adaptive-LASSO memilih 11 peubah prediktor. Jumlah peubah hasil seleksi menggunakan metode stepwise adalah yang terbanyak dan metode Adaptive-LASSO adalah paling sedikit. Setelah melakukan seleksi peubah, dilakukan pengulangan penarikan contoh data latih sebanyak 100 kali. Kemudian dilakukan pemodelan menggunakan data latih dan dilakukan prediksi berdasarkan peubah prediktor pada data uji. Hasil prediksi pada setiap ulangan divalidasi menggunakan data aktual y_i yang ada pada gugus data uji hingga diperoleh nilai RMSE setiap ulangan dan perlakuan yang ditampilkan pada Tabel 5.

Tabel 5. Perbandingan nilai RMSE antara model QRF menggunakan seluruh peubah dan model QRF menggunakan hasil seleksi peubah dengan 100 ulangan

Dataset	QRF seluruh peubah	Perlakuan				
		Stepwise-QRF	Lasso-QRF	Elastic Net-QRF	Adaptive-QRF	Relaxed-QRF
1	0.7370	0.7313	0.7162	0.7116	0.7040	0.7182
2	0.7908	0.7763	0.7481	0.7254	0.6907	0.7265
3	0.7956	0.7843	0.7636	0.7823	0.7793	0.7495
4	0.6635	0.6673	0.6334	0.6520	0.6496	0.6494
⋮	⋮	⋮	⋮	⋮	⋮	⋮
98	0.7120	0.6969	0.6895	0.6732	0.6975	0.6792
99	0.7553	0.7384	0.7088	0.6861	0.7127	0.7123
100	0.6537	0.6410	0.6052	0.6043	0.5791	0.6222
Rata-rata	0.7169	0.6951	0.6728	0.6653	0.6583	0.6740

Berdasarkan Tabel 5, terlihat bahwa seluruh metode seleksi mampu menurunkan rata-rata nilai RMSE. Hal ini menunjukkan bahwa metode seleksi mampu memperbaiki kinerja pemodelan QRF. RMSE terendah dihasilkan oleh Adaptive-QRF yang mampu menurunkan rata-rata RMSE dari 0.7169 menjadi 0.6583. Untuk

melihat apakah penurunan nilai RMSE adalah nyata secara statistik maka dilakukan uji t-berpasangan. Hasilnya ditampilkan pada Tabel 6.

Tabel 6. Hasil uji t-berpasangan nilai RMSE antara model QRF menggunakan seluruh peubah dan model QRF menggunakan hasil seleksi peubah

Perbandingan Model	Selisih rata-rata	SK 95%	t-hitung	p-value
QRF seluruh peubah vs Stepwise-QRF	0.0218	(0.01935; 0.02430)	17.49	0.000
QRF seluruh peubah vs Lasso-QRF	0.0441	(0.04107; 0.04709)	29.07	0.000
QRF seluruh peubah vs Elastic Net-QRF	0.0516	(0.04576; 0.05751)	17.45	0.000
QRF seluruh peubah vs Adaptive-QRF	0.0586	(0.05110; 0.06611)	15.50	0.000
QRF seluruh peubah vs Relaxed-QRF	0.04285	(0.03927; 0.04644)	23.74	0.000
Adaptive-QRF vs Stepwise-QRF	-0.0368	(-0.04348; -0.03009)	-10.91	0.000
Adaptive-QRF vs Lasso-QRF	-0.0145	(-0.02098; -0.00808)	-4.47	0.000
Adaptive-QRF vs Elastic Net-QRF	-0.0070	(-0.01320; -0.00074)	-2.22	0.029
Adaptive-QRF vs Relaxed-QRF	-0.01576	(-0.02201; -0.00950)	-5.00	0.000

Berdasarkan Tabel 6, seluruh metode seleksi yang diajukan memberikan pengaruh nyata terhadap penurunan rata-rata nilai RMSE dengan tingkat kepercayaan 95%. Model Adaptive-QRF adalah model dengan kinerja terbaik dibanding model lain yang diperbandingkan dalam penelitian ini. Model Adaptive-QRF memiliki nilai RMSE terkecil secara nyata.

4. Kesimpulan

Metode Adaptive-LASSO adalah metode seleksi peubah dengan kinerja terbaik untuk menduga berat gabah plot Survei Ubinan dibanding metode Stepwise, LASSO, Elastic Net, dan Relaxed LASSO. Peubah terseleksi oleh Adaptive-LASSO digunakan dalam pemodelan Adaptive-QRF yaitu varietas benih, cara penanaman (monokultur atau tumpangsari), sistem penanaman (jajar legowo atau

bukan), terkena serangan hama atau tidak, lokasi penanaman, bulan panen, dan titik koordinat longitude dan latitude. Model Adaptive-QRF memberikan tingkat presisi terbaik dengan nilai RMSE terkecil secara nyata dengan tingkat kepercayaan 95%.

5. Ucapan Terima Kasih

Penulis mengucapkan terima kasih kepada BPS, IPB, dan semua pihak yang berkontribusi dalam perbaikan makalah ini (Rima, Cynthia, dan Kak Lani BPS Provinsi Kalimantan Tengah).

Daftar Pustaka

- Ardiansyah M, Djuraidah A, Kurnia A. (2018). Pendugaan Produktivitas Padi di Tingkat Kecamatan Menggunakan Geoadaptive Small Area Model. *Jurnal Penelitian Pertanian Tanaman Pangan* 2 (2). pp.101-110.
- Ardiansyah M dan Tofri Y. (2019). Perbandingan Data Produktivitas Padi Antara Hasil Wawancara Pascapanen dengan Data Survei Ubinan di Kalimantan Tengah. *Jurnal Penelitian Pertanian Tanaman Pangan* 3 (1). pp.17-22.
- Fang Y, Xu P, Yang J, Qin Y. (2018). A quantile regression forest based method to predict drug response and assess prediction reliability. *PLOS ONE* 13 (10). pp.1-16.
- Hastie T, Tibshirani R, dan Wainwright M. (2015). *Statistical Learning with Sparsity The Lasso and Generalizations*. New York: Taylor & Francis Group.
- Kim Y, Hao J, Mallavarapu T, Park J and Kang M. (2019). Hi-LASSO: High-Dimensional LASSO." *IEEE Access*. 7. pp.44562-44573.
- Meinshausen N. (2006). Quantile Regression Forests. *Journal of Machine Learning Research* 7. pp.983-999.
- Meinshausen N. (2007). Relaxed lasso. *Computational Statistics and Data Analysis*. 52 (1). pp. 374-393.
- Meinshausen N. (2017). An R Package 'quantregForest' (Quantile Regression Forests). Version: 1.3-7. Published: 19 December 2017. URL <https://CRAN.R-project.org/package=quantregForest>. [9 Juni 2020].
- Zou H and Hastie T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*. 67 (2). pp.301-320.
- Zou H. (2006). The adaptive lasso and its oracle properties," *American Statistical Association Journal of the American Statistical Association*. 101 (476). pp. 1418-142.
- Morozova O, Levina O, Uusküla A, dan Heimer R. 2015. Comparison of subset selection methods in linear regression in the context of health-related quality of life and substance abuse in Russia. *BMC Medical Research Methodology*. Pp. 15-71.