

Analisis Regresi PLS Sebagai Alternatif Dari Regresi Linear Berganda: Studi Kasus Pengaruh Luas Lahan dan Luas Panen Terhadap Produksi Padi di Kabupaten Manokwari

Esther Ria Matulesy^{1*}, Agnes Uliarta Tambunan²

Jurusan Matematika dan Statistika, FMIPA, Universitas Papua

Email: e.matulesy@unipa.ac.id

Abstrak

Analisis regresi linier berganda merupakan suatu metode yang dapat menganalisis pengaruh variabel bebas dengan variabel terikat. Dalam melakukan analisis regresi linier berganda harus dibuat asumsi berupa uji normalitas, uji autokorelasi, uji heteroskedastisitas, dan uji multikolinearitas. Jika uji multikolinearitas tidak terpenuhi, dapat diatasi dengan metode Partial Least Square Regression (PLS). Regresi Univariate Partial Least Square (PLS) adalah model yang menghubungkan variabel respon Y dengan sekumpulan variabel prediktor X. Tujuan dari PLS adalah untuk membentuk komponen yang dapat menangkap informasi dari variabel prediktor untuk memprediksi variabel respon. Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa dengan menggunakan metode PLS, luas lahan dan luas panen berpengaruh positif terhadap produksi padi. Nilai RMSE adalah 1242,6784, nilai MAE adalah 1014,8091. Nilai Rkuadrat metode Partial Least Square (PLS) memberikan nilai yang sangat besar yaitu 0,911 atau 91,1%, artinya 91,1% keragaman variabel produksi dapat dijelaskan oleh variabel luas lahan dan luas panen. Artinya metode Partial Least Square (PLS) memberikan akurasi model yang lebih baik.

Kata Kunci: *luas lahan, luas panen, produksi, analisis regresi linear berganda, PLS*

Abstract

Multiple linear regression analysis is a method that can analyze the influence of the independent variable with the dependent variable. In performing multiple linear regression analysis, assumptions must be made in the form of normality test, autocorrelation test, heteroscedasticity test, and multicollinearity test. If the multicollinearity test is not met, it can be overcome by the Partial Least Square Regression (PLS) method. Univariate Partial Least Square (PLS) regression is a model that relates a response variable Y to a set of predictor variables X. The purpose of PLS is to form a component that can capture information from predictor variables to predict the response variable. Based on the results and discussion, it can be concluded that using the PLS method, land area and harvested area have a positive effect on rice production. The RMSE value is 1242.6784, the MAE value is 1014.8091. The Rsquared value of the Partial Least Square (PLS) method gives a very large value of 0.911 or 91.1%, meaning that 91.1% of the diversity of production variables can be explained by the variables of land area and harvested area. This means that the Partial Least Square (PLS) method provides better model accuracy.

Keywords: *land area, harvested area, production, multiple linear regression analysis, PLS*

PENDAHULUAN

Salah satu komoditas pertanian yang saat ini sebagai komoditas unggulan adalah tanaman padi. Padi merupakan komoditas bahan makanan utama masyarakat Indonesia yang tingkat konsumsinya mencapai 132,98 kg/kapita/tahun. Sehingga produksi padi menjadi prioritas utama untuk mengatasi kekurangan supply, dimana peningkatan produksi padi terjadi dari beberapa faktor yaitu salah satunya luas lahan sawah di Indonesia (Saputri & Amalita, 2020).

Menurut BPS (2022), lahan sawah adalah lahan pertanian yang berpetak-petak dan dibatasi oleh pematang, saluran untuk menahan/menyalurkan air. Lahan sawah memiliki fungsi strategis, karena merupakan penyedia bahan pangan utama bagi penduduk Indonesia. Selain luas lahan, luas panen juga

merupakan salah satu faktor yang mempengaruhi peningkatan produksi padi.

Luas panen merupakan luas lahan dari hasil komoditi yang sudah siap panen. Daerah yang mempunyai lahan yang berpotensi untuk dikembangkan sebagai lahan pertanian berupa tanaman pangan salah satunya dapat dilihat dari luas panen (Saputri & Amalita, 2020). Menurut Hasan (2010), luas panen di Indonesia terus mengalami peningkatan, meski lahan sawah di Indonesia cenderung berkurang atau semakin sempit. Hal tersebut bisa saja terjadi karena luas panen merupakan hasil perkalian antara luas lahan dengan intensitas tanam. Dari data luas lahan dan luas panen padi, kita dapat melihat pengaruhnya terhadap pertumbuhan produksi padi menggunakan analisis regresi linier berganda.

Menurut Hasan (2002), analisis regresi linier berganda adalah hubungan secara linear antara dua atau lebih variabel independen. Analisis ini bertujuan untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen, apakah masing-masing variabel independen berhubungan positif atau negatif, dan untuk memprediksi nilai dari variabel dependen apabila nilai variabel independen mengalami kenaikan atau penurunan. Dalam melakukan analisis regresi linier berganda, harus melakukan uji-uji asumsi berupa uji normalitas, uji autokorelasi, uji heterokedastisitas, uji multikolinieritas. Jika uji multikolinieritas tidak terpenuhi maka dapat diatasi dengan metode Regresi *Partial Least Square* (PLS).

Regresi *Partial Least Square* (PLS) univariat adalah sebuah model yang menghubungkan antara sebuah variabel independen Y dengan sekumpulan variabel dependen. Regresi PLS merupakan salah satu metode untuk mengatasi multikolinieritas yang dapat diperoleh melalui regresi sederhana maupun berganda dengan mengambil kesimpulan dari uji signifikansi. Uji signifikansi bertujuan untuk memilih variabel dependen pembangun komponen PLS dan menentukan banyaknya komponen PLS yang terbentuk. Tujuan PLS adalah membentuk komponen yang dapat menangkap informasi dari variabel dependen untuk memprediksi variabel independen (Bastien et al., 2005). Berdasarkan latar belakang yang telah diuraikan, penulis ingin menganalisis “Analisis Regresi PLS Sebagai Alternatif Dari Regresi Linier Berganda (Studi Kasus: Pengaruh Luas Lahan dan Luas Panen Terhadap Produksi Padi Di Kabupaten Manokwari)”.

METODE

Pada penelitian ini digunakan data sekunder yang berasal dari Dinas Pertanian dan Ketahanan Pangan Kabupaten Manokwari. Objek pada penelitian ini adalah Kabupaten Manokwari. Variabel yang digunakan adalah luas lahan, luas panen dan data produksi padi.

Analisis Regresi Berganda adalah analisis yang bertujuan untuk mengetahui arah hubungan antara variabel independen dengan variabel dependen apakah masing-masing variabel independen berhubungan positif atau negatif dan untuk memprediksi nilai dari variabel dependen apabila nilai variabel independen mengalami kenaikan atau penurunan (Hasan, 2002).

Tahapan analisis yang dilakukan sebagai berikut :

1. Menyediakan data dan input data ke software Rstudio. Data yang digunakan adalah data luas lahan, luas panen dan produksi padi di kabupaten Manokwari dari Tahun 2005-2015
2. Melakukan uji analisis deskriptif
3. Melakukan uji analisis regresi berganda. Dalam analisis regresi berganda ada beberapa langkah-langkah yang harus dilakukan, diantaranya
 - a. Uji asumsi
 - b. Model regresi
 - c. Koefisien determinasi
 - d. Uji hipotesis
4. Pemodelan regresi dengan metode PLS untuk mengatasi multikolinearitas.
5. Menggunakan metode PLS dengan cara :
 - a. Mencari komponen PLS yang tepat
 - b. Menentukan model regresi.

HASIL DAN PEMBAHASAN

Dalam penelitian ini, langkah pertama yang dilakukan adalah mendeskripsikan data luas lahan, luas panen dan produksi padi. Berdasarkan Analisis Deskriptif diperoleh hasil pengukuran data masing-masing variabel sebagai berikut.

| Tabel 1. Analisis Deskriptif | | | | | |
|------------------------------|---------|---------------|--------|---------------|--------|
| Produksi | | Luas Lahan | | Luas Panen | |
| Min. | 9862 | Min. | 1240 | Min. | 2481 |
| 1st Qu | : 16562 | 1st Qu | : 2310 | 1st Qu | : 4620 |
| Median : 18024 | | Median : 2502 | | Median : 5004 | |
| Mean | : 18626 | Mean | : 2606 | Mean | : 5213 |
| 3rd | 20778 | 3rd | 2910 | 3rd | 5821 |
| Max | 25309 | Max | 3689 | Max | 7378 |

Setelah melakukan analisis deskriptif, kita dapat melakukan uji berikutnya yaitu uji analisis regresi berganda.

a. Uji Asumsi

i. Uji Normalitas

Model regresi yang baik jika distribusi normal atau mendekati normal. Hal ini dapat didukti dengan melihat nilai Shapiro-Walk. Berikut nilai Shapiro-Walk

| Tabel 2. Uji Shapiro-Walk | |
|-----------------------------|------------------|
| Shapiro-Wilk normality test | |
| data: model2\$residuals | |
| W = 0.96582, | p-value = 0.8415 |

Output nilai shapiro-walk menunjukkan 0.96582 dan *p-value* 0.8415 yang artinya berdistribusi normal.

ii. Uji Autokorelasi

Berdasarkan Tabel 3, diperoleh hasil uji autokorelasi menggunakan metode durbinwatson Test *p – valu* (0.012) < 0.05, artinya ada korelasi antara variabel dependen dengan variabel independen.

| Tabel 3. Uji Autokorelasi | | | |
|---------------------------------|-----------------|---------------|---------|
| Lag | Autocorrelation | D-W Statistic | p-value |
| 1 | 0.4789987 | 0.751355 | 0.012 |
| Alternative hypothesis: rho !=0 | | | |

iii. Uji Heterokedastisitas

Berdasarkan Tabel 4 Uji bptest, diperoleh hasil uji normalitas menggunakan metode Brusch-Pangan dengan *p-value* > 0.005 (0.9328) yang artinya tidak ada gejala heterokedastisitas.

| Tabel 4. Uji bptest | |
|--------------------------------------|--------------------------|
| studentized Breusch-Pagan test data: | |
| BP = 0.0071073, | df = 1, p-value = 0.9328 |

iv. Uji Multikolinearitas

Karena korelasi dalam variabel cukup tinggi dan kedua variabel memberikan informasi yang sama maka VIF tidak dapat dihitung.

Adapun cara untuk mengatasi terjadinya multikolineritas pada analisis regresi linier berganda adalah dengan metode Regresi Partial Least Square (PLS).

- a. Melakukan pemodelan regresi dengan metode PLS untuk mengatasi multikolineritas.
- i. Mencari komponen PLS yang tepat sehingga didapatkan model regresi yang bebas multikolinearitas. Berdasarkan Tabel 5, terdapat nilai koefisien pada Regresi PLS yang menyatakan bahwa variabel luas lahan dan luas panen memiliki nilai koefisien sebesar 2082.744 dan 2082.744, sehingga dapat

dibentuk model sebagai berikut:
 $Produksi = 2082.744 \text{ Luas Lahan} + 2082.744 \text{ Luas Panen}.$
 Berikut adalah Tabel 5

Tabel 5. Koefisien Model Regresi PLS

| |
|---------------------|
| 1 comp |
| Produksi |
| Luas.Lahan 2082.744 |
| Luas.Panen 2082.744 |

- ii. Kebaikan Model
 Pemilihan model terbaik dapat dilakukan dengan melihat nilai determinasi (R^2). Model dikatakan baik jika nilai (R^2) berkisar antara 0 dan 1. Selain memperhatikan nilai determinasi (R^2), pemilihan model terbaik juga dapat dilakukan dengan melihat nilai *Mean Absolute Error* (MAE) dan *Root Mean Square Error* (RMSE). Metode terbaik adalah metode dengan nilai MAE dan 35 nilai RMSE terkecil.

Tabel 6. Kebaikan Model

| RMSE | Rsquared | MAE |
|--------------|-----------|--------------|
| 1242.6784675 | 0.9108314 | 1014.8091257 |

Berdasarkan Tabel 6, dapat dilihat bahwa nilai RMSE adalah 1242.6784, artinya rata-rata selisih antara nilai sebenarnya dengan hasil prediksi adalah 1242.6784. Nilai MAE adalah 1014.8091, artinya rata-rata selisih (*absolute*) antara nilai sebenarnya dengan hasil prediksi adalah 1014.8091. Nilai *Rsquared* dari metode *Partial Least Square* (PLS) memberikan nilai yang sangat besar yaitu sebesar 0,911 atau 91,1%. artinya 91% keragaman variabel *Y* dapat dijelaskan oleh variabel *X*. Hal ini berarti metode *Partial Least Square* (PLS) memberikan ketepatan model yang lebih baik.

SIMPULAN

Berdasarkan hasil dan pembahasan dapat disimpulkan bahwa dengan menggunakan metode PLS, luas lahan dan luas panen berpengaruh positif terhadap produksi padi. Nilai RMSE adalah 1242.6784, artinya rata-rata selisih antara nilai sebenarnya dengan hasil prediksi adalah 1242.6784. Nilai MAE adalah 1014.8091, artinya rata-rata selisih (*absolute*) antara nilai sebenarnya dengan hasil prediksi adalah 1014.8091. Nilai *Rsquared* dari metode *Partial Least Square* (PLS) memberikan nilai yang sangat besar yaitu sebesar 0,911 atau 91,1%, artinya keragaman variabel produksi dapat dijelaskan oleh variabel luas lahan dan luas panen. Hal ini berarti metode *Partial Least Square* (PLS) memberikan ketepatan model yang lebih baik.

DAFTAR PUSTAKA

Badan Pusat Statistik. 2005. *Papua Barat Dalam Angka* <https://papuabarat.bps.go.id> [2 Februari 2022].
 Bastien, P., Vinzi, V. E. and Tenenhaus, M. 2005. *PLS generalised linear regression*. Computational Statistics & data analysis, 48(1), pp. 17–46.
 Hasan, F. 2010. *Peran luas panen dan produktivitas terhadap pertumbuhan produksi tanaman pangan di jawa timur*. Jurnal Embriyo, 7(1), pp. 15–20.
 Hasan, M. I. 2002. *Pokok-pokok materi metodologi penelitian dan aplikasinya*. Jakarta: Ghalia Indonesia.
 Saputri, W. dan Amalita, N. 2020. Analisa Tentang Luas Tanam dan Luas Panen di Bidang Komoditi Perkebunan di Provinsi Sumatera Barat dengan Menggunakan Analisis Profil. *UNIPjoMath*, 3(1), pp. 85–89.