

Analisis Regresi Setelah *Pre-Processing* Data Iklim di Indonesia Tahun 2011-2015

Dengan menggunakan minitab didapatkan hasil analisis regresi pada data yang sebelumnya telah dilakukan *pre-processing* yaitu sebagai berikut.

a. *Missing Value*

Berikut ini merupakan hasil analisis regresi pada data iklim yang telah dilakukan deteksi *missing value* dan diatasi menggunakan imputasi *mean*.

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
603,048	47,29%	45,35%	42,48%

Regression Equation

Jumlah Curah Hujan (mm) = -4158 + 11,50 Jumlah Hari Hujan (hari)
+ 4,74 Tekanan Udara (mb)
- 3,60 Penyinaran Matahari (%) - 86,8 Suhu (°C)
+ 20,3 Kecepatan Angin (m/det) + 25,6 Kelembaban

Hasil di atas menunjukkan bahwa *R-square* atau koefisien determinasi berganda yang diperoleh sebesar 47,29%.

b. *Missing Value dan Outlier*

Berikut ini merupakan hasil analisis regresi pada data iklim yang telah dilakukan deteksi *missing value* dan *outlier* yang kemudian diatasi menggunakan imputasi *mean* dan penghapusan *outlier*.

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
575,097	51,39%	49,43%	46,33%

Regression Equation

Jumlah Curah Hujan (mm) = 10171 + 10,63 Jumlah Hari Hujan (hari)
- 12,2 Tekanan Udara (mb)
- 0,61 Penyinaran Matahari (%) - 51,3 Suhu (°C)
+ 32,8 Kecepatan Angin (m/det) + 46,8 Kelembaban

Hasil di atas menunjukkan bahwa *R-square* yang diperoleh sebesar 51,39%, dimana hasil ini memiliki koefisien determinasi berganda yang lebih besar dari analisis sebelumnya.

c. *Missing Value dan Transformasi*

1. Normalisasi

Berikut ini merupakan hasil analisis regresi pada data iklim yang telah dilakukan deteksi *missing value* dan diatasi menggunakan imputasi *mean*. Serta variabel prediktornya telah ditransformasi menggunakan metode normalisasi.

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
609,428	46,17%	44,19%	41,21%

Regression Equation

Jumlah Curah Hujan (mm) = 51970 + 2422 Jumlah Hari Hujan (hari)
- 52226 Tekanan Udara (mb)
- 9070 Penyinaran Matahari (%) - 24556 Suhu (°C)

+ 26508 Kecepatan Angin (m/det) + 26155 Kelembaban

Hasil di atas menunjukkan bahwa *R-square* atau koefisien determinasi berganda diperoleh lebih rendah dari analisis sebelumnya yaitu sebesar 46,17%.

2. Minmax

Berikut ini merupakan hasil analisis regresi pada data iklim yang telah dilakukan deteksi *missing value* dan diatasi menggunakan imputasi *mean*. Serta variabel prediktornya telah ditransformasi menggunakan metode minmax.

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
603,048	47,29%	45,35%	42,48%

Regression Equation

Jumlah Curah Hujan (mm) = 271 + 478,3 Jumlah Hari Hujan (hari)
 + 88,0 Tekanan Udara (mb)
 - 44,7 Penyinaran Matahari (%) - 93,2 Suhu (°C)
 + 77,9 Kecepatan Angin (m/det) + 223 Kelembaban

Hasil di atas menunjukkan bahwa *R-square* atau koefisien determinasi berganda yang diperoleh sebesar 47,29%.

3. 3Sigmoid

Berikut ini merupakan hasil analisis regresi pada data iklim yang telah dilakukan deteksi *missing value* dan diatasi menggunakan imputasi *mean*. Serta variabel prediktornya telah ditransformasi menggunakan metode 3 sigmoidal.

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
609,435	46,17%	44,18%	41,21%

Regression Equation

Jumlah Curah Hujan (mm) = 60035 + 5021 Jumlah Hari Hujan (hari)
 - 130488 Tekanan Udara (mb)
 - 18101 Penyinaran Matahari (%) - 48828 Suhu (°C)
 + 52998 Kecepatan Angin (m/det) + 52451 Kelembaban

Hasil di atas menunjukkan bahwa *R-square* atau koefisien determinasi berganda yang diperoleh sebesar 46,17%.

d. Missing Value, Outlier dan Transformasi

1. Normalisasi

Berikut ini merupakan hasil analisis regresi pada data iklim yang telah dilakukan deteksi *missing value* dan *outlier* yang kemudian diatasi menggunakan imputasi *mean* dan penghapusan *outlier*. Serta variabel prediktor telah ditransformasi menggunakan metode normalisasi.

Model Summary

S	R-sq	R-sq (adj)	R-sq (pred)
573,103	51,72%	49,78%	47,13%

Regression Equation

Jumlah Curah Hujan (mm) = 58148 + 1451 Jumlah Hari Hujan (hari)
 - 59571 Tekanan Udara (mb)
 - 4577 Penyinaran Matahari (%) - 47642 Suhu (°C)
 + 38056 Kecepatan Angin (m/det) + 45399 Kelembaban

Hasil di atas menunjukkan bahwa *R-square* yang diperoleh sebesar 51,72%, dimana hasil ini memiliki koefisien determinasi berganda yang lebih besar dari analisis sebelum-sebelumnya.

2. Minmax

Berikut ini merupakan hasil analisis regresi pada data iklim yang telah dilakukan deteksi *missing value* dan *outlier* yang kemudian diatasi menggunakan imputasi *mean* dan penghapusan *outlier*. Serta variabel prediktor telah ditransformasi menggunakan metode *minmax*.

Model Summary

S	R-sq	R-sq (adj)	R-sq(pred)
575,097	51,39%	49,43%	46,33%

Regression Equation

$$\begin{aligned} \text{Jumlah Curah Hujan (mm)} = & 1087 + 442,1 \text{ Jumlah Hari Hujan (hari)} \\ & - 114 \text{ Tekanan Udara (mb)} \\ & - 6,3 \text{ Penyinaran Matahari (\%)} - 37,4 \text{ Suhu (}^{\circ}\text{C)} \\ & + 67,1 \text{ Kecepatan Angin (m/det)} + 159,2 \text{ Kelembaban} \end{aligned}$$

Hasil di atas menunjukkan bahwa *R-square* yang diperoleh sebesar 51,39%.

3. 3Sigmoid

Berikut ini merupakan hasil analisis regresi pada data iklim yang telah dilakukan deteksi *missing value* dan *outlier* yang kemudian diatasi menggunakan imputasi *mean* dan penghapusan *outlier*. Serta variabel prediktor telah ditransformasi menggunakan metode 3 sigmoidal.

Model Summary

S	R-sq	R-sq (adj)	R-sq(pred)
573,094	51,73%	49,78%	47,13%

Regression Equation

$$\begin{aligned} \text{Jumlah Curah Hujan (mm)} = & 67444 + 3057 \text{ Jumlah Hari Hujan (hari)} \\ & - 149044 \text{ Tekanan Udara (mb)} \\ & - 9110 \text{ Penyinaran Matahari (\%)} - 95054 \text{ Suhu (}^{\circ}\text{C)} \\ & + 76108 \text{ Kecepatan Angin (m/det)} + 90986 \text{ Kelembaban} \end{aligned}$$

Hasil di atas menunjukkan bahwa *R-square* yang diperoleh sebesar 51,73%, yang artinya variabel jumlah curah hujan (Y) dapat dijelaskan oleh variabel-variabel prediktor (X) secara simultan sebesar 51,73%, dan sisanya dijelaskan oleh variabel lain di luar model. Nilai ini adalah yang paling besar dari analisis sebelumnya yang telah dilakukan. Sehingga dapat disimpulkan bahwa model regresi terbaik dengan nilai *R-square* paling besar adalah data dengan *pre-processing* yaitu *missing value*, *outlier* dan transformasi menggunakan metode 3 sigmoidal.