Perbandingan Metode untuk Model Terbaik pada Regresi Linear

Data solar-radiation merupakan data meteorologi dari stasiun cuaca HI-SEAS yang berlangsung selama 4 bulan (Sempember-Desember). Pengukuran tiap variabel dilakukan secara otomatis tiap 5 meni setiap harinya. Beberapa variabel yang diukur antara lain *Solar Radiation, Temperature, Humidity, Barometric Pressure, Wind Direction, Wind Speed, Sunrise And Sunset Time*. Pada pratikum ini melihat ada tidaknya pengaruh dan besar pengaruh dari *Temperature, Humidity, Barometric Pressure, Wind Direction, Wind Speed, Duration* (lama matahari bersinar) terhadap *Humidity* di suatu daerah. Dikarenakan data dicatat tiap 5 menit, maka pada praktikum ini diambil data dengan waktu pengamatan yang sama yaitu pukul 08.05 AM.Berikut analisis regresi setelah penanganan *missing value*.

Hasil analisis regresi setelah penanganan missing value dengan imputasi mean

**Regression Analysis: Humidity versus Radiation; Temperature; ...**

The regression equation is

Humidity = - 1408 - 0,0535 Radiation - 4,26 Temperature + 42,9 Pressure

+ 0,078 Speed + 0,594 duration

Predictor Coef SE Coef T P

Constant -1408 1387 -1,02 0,312

Radiation -0,05346 0,02131 -2,51 0,014

Temperature -4,2631 0,6544 -6,51 0,000

Pressure 42,88 45,75 0,94 0,351

Speed 0,0777 0,5770 0,13 0,893

duration 0,59439 0,08900 6,68 0,000

S = 20,5719 R-Sq = 55,3% R-Sq(adj) = 53,2%

Analysis of Variance

Source DF SS MS F P

Regression 5 56063 11213 26,49 0,000

Residual Error 107 45283 423

Total 112 101346

Berdasarkan pengujian dengan minitab menggunakan data *solar prediction* didapatkan bahwa variabel radiasi, temperature, duration secara individu berpengaruh signifikan terhadap model karena p-value>0,05. R-square dari model sebesar 55,3%, dapat dikatakan bahwa model memnggambarkan lebih dari setengah data. Variabel radiasi dan temperature berpengaruh negatif terhadap kelembaban, sedangkan variabel lainnya berpengaruh positif. Selanjutnya dilakukan analisis regresi pada data dengan menghilangkan pengamatan *outlier* dan transformasi. Berikut hasil analisis regresi data setelah menghilangkan pengamatan *outlier*.

Hasil analisis regresi setelah penanganan *missing value* dan *outlier*

**Regression Analysis: Humidity versus Radiation; Temperature; ...**

The regression equation is

Humidity = - 924 - 0,0548 Radiation - 4,23 Temperature + 27,2 Pressure

- 0,0385 WindDirection(Degrees) + 0,102 Speed

+ 0,593 duration(minute)

Predictor Coef SE Coef T P

Constant -924 1632 -0,57 0,572

Radiation -0,05475 0,02215 -2,47 0,015

Temperature -4,2324 0,6727 -6,29 0,000

Pressure 27,16 53,68 0,51 0,614

WindDirection(Degrees) -0,03854 0,02725 -1,41 0,160

Speed 0,1020 0,6502 0,16 0,876

duration(minute) 0,59294 0,09008 6,58 0,000

S = 20,7652 R-Sq = 55,2% R-Sq(adj) = 52,6%

Analysis of Variance

Source DF SS MS F P

Regression 6 54649,1 9108,2 21,12 0,000

Residual Error 103 44412,9 431,2

Total 109 99062,0

Jika dilihat dari output minitab dari analisis regresi data yang telah ditangani missing value dan outlier didapatkan bahwa variabel yang secara individu berpengaruh signifikan terhadap model karena p-value>0,05 sama seperti sebelumnya. R-square dari model sebesar 55,2%, lebih kecil dibandingkan dengan data. Variabel radiasi, *temperature,* dan *wind direction* berpengaruh negatif terhadap kelembaban, sedangkan variabel lainnya berpengaruh positif.

Hasil analisis regresi setelah penanganan *missing value* dan transformasi

**Regression Analysis: Humidity\_1 versus Radiation\_1; Temperature\_1; ...**

The regression equation is

Humidity\_1 = 0,054 - 0,111 Radiation\_1 - 3,82 Temperature\_1 - 2,65 Pressure\_1

- 0,0831 WindDirection(Degrees)\_1 + 0,269 Speed\_1

+ 0,475 duration(minute)\_1

Predictor Coef SE Coef T P

Constant 0,0539 0,1012 0,53 0,596

Radiation\_1 -0,11115 0,05673 -1,96 0,053

Temperature\_1 -3,8201 0,5776 -6,61 0,000

Pressure\_1 -2,654 1,861 -1,43 0,157

WindDirection(Degrees)\_1 -0,08307 0,04605 -1,80 0,074

Speed\_1 0,2694 0,5624 0,48 0,633

duration(minute)\_1 0,4747 0,2572 1,85 0,068

S = 0,0130419 R-Sq = 64,0% R-Sq(adj) = 62,0%

Analysis of Variance

Source DF SS MS F P

Regression 6 0,0320671 0,0053445 31,42 0,000

Residual Error 106 0,0180296 0,0001701

Total 112 0,0500967

Hasil pengujian dengan minitab menggunakan data yang telah mengalami proses imputasi missing value dan transformasi didapatkan bahwa R-square dari model sebesar 64%, lebih besar jika dibandingkan dengan hasil R-square dua perlakuan sebelumnya. Variabel temperature saja yang secara individu berpengaruh signifikan terhadap model karena p-value>0,05. Variabel speed dan duration berpengaruh positif terhadap kelembaban, sedangkan variabel lainnya berpengaruh negatif.

Hasil analisis regresi setelah penanganan *missing value, outlier,* dan transformasi

**Regression Analysis: Humidity\_2 versus Radiation\_2; Temperature\_2; ...**

The regression equation is

Humidity\_2 = 0,061 - 0,115 Radiation\_2 - 3,87 Temperature\_2 - 2,54 Pressure\_2

- 0,0864 WindDirection(Degrees)\_2 + 0,266 Speed\_2

+ 0,460 duration(minute)\_2

Predictor Coef SE Coef T P

Constant 0,0605 0,1045 0,58 0,564

Radiation\_2 -0,11480 0,05879 -1,95 0,054

Temperature\_2 -3,8683 0,6217 -6,22 0,000

Pressure\_2 -2,541 1,910 -1,33 0,186

WindDirection(Degrees)\_2 -0,08644 0,04726 -1,83 0,070

Speed\_2 0,2661 0,6466 0,41 0,681

duration(minute)\_2 0,4597 0,2639 1,74 0,085

S = 0,0132133 R-Sq = 62,9% R-Sq(adj) = 60,8%

Analysis of Variance

Source DF SS MS F P

Regression 6 0,0305327 0,0050888 29,15 0,000

Residual Error 103 0,0179829 0,0001746

Total 109 0,0485156

Hasil pengolahan terakhir dengan imputasi missing value, menghilangkaan pengamatan outlier, dan transformasi didapatkan nilai R-square dari model sebesar 62,9%, lebih kecil jika dibandingkan dengan R-square data sebelumnya. Variabel temperature secara individu berpengaruh signifikan terhadap model karena p-value>0,05. Variabel radiasi dan temperature berpengaruh negatif terhadap kelembaban, sedangkan variabel lainnya berpengaruh positif sama seperti sebelumnya.

Dari empat macam pengujian, perlakuan penganganan terhadap missing value dan transformasi data yang menghasilkan nilai R-square tertinggi yaitu 64%. Maka dapat dikatakan untuk data regresi ini, model terbaik dihasilkan dari penanganan missing value dan transformasi dari data.