**Membandingkan empat model regresi dari data dengan *pre-processing* yang berbeda.**

1. *Missing Value*

|  |  |  |  |
| --- | --- | --- | --- |
| **Dep. Variable:** | y | **R-squared:** | 0.733 |
| **Model:** | OLS | **Adj. R-squared:** | 0.661 |
| **Method:** | Least Squares | **F-statistic:** | 10.16 |
| **Date:** | Sun, 10 Mar 2019 | **Prob (F-statistic):** | 6.16e-08 |
| **Time:** | 07:52:55 | **Log-Likelihood:** | -663.36 |
| **No. Observations:** | 48 | **AIC:** | 1349. |
| **Df Residuals:** | 37 | **BIC:** | 1369. |
| **Df Model:** | 10 |  |  |
| **Covariance Type:** | nonrobust |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **const** | 6.209e+05 | 6.94e+05 | 0.894 | 0.377 | -7.86e+05 | 2.03e+06 |
| **x1** | 137.2474 | 110.882 | 1.238 | 0.224 | -87.422 | 361.917 |
| **x2** | -333.7656 | 364.058 | -0.917 | 0.365 | -1071.418 | 403.887 |
| **x3** | -105.8141 | 145.921 | -0.725 | 0.473 | -401.479 | 189.851 |
| **x4** | 206.4581 | 533.269 | 0.387 | 0.701 | -874.048 | 1286.964 |
| **x5** | 733.0939 | 699.646 | 1.048 | 0.302 | -684.525 | 2150.712 |
| **x6** | 44.8635 | 62.866 | 0.714 | 0.480 | -82.516 | 172.243 |
| **x7** | -262.9602 | 138.842 | -1.894 | 0.066 | -544.281 | 18.360 |
| **x8** | -292.9801 | 82.819 | -3.538 | 0.001 | -460.787 | -125.173 |
| **x9** | -417.7147 | 492.684 | -0.848 | 0.402 | -1415.987 | 580.558 |
| **x10** | 13.8782 | 7.516 | 1.847 | 0.073 | -1.350 | 29.106 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Omnibus:** | 3.024 | **Durbin-Watson:** | 2.070 |
| **Prob(Omnibus):** | 0.220 | **Jarque-Bera (JB):** | 2.472 |
| **Skew:** | -0.431 | **Prob(JB):** | 0.291 |
| **Kurtosis:** | 2.297 | **Cond. No.** | 6.07e+05 |

1. *Missing Value, Outlier*

|  |  |  |  |
| --- | --- | --- | --- |
| **Dep. Variable:** | y | **R-squared:** | 0.913 |
| **Model:** | OLS | **Adj. R-squared:** | 0.889 |
| **Method:** | Least Squares | **F-statistic:** | 37.72 |
| **Date:** | Sun, 10 Mar 2019 | **Prob (F-statistic):** | 4.33e-16 |
| **Time:** | 07:57:40 | **Log-Likelihood:** | -623.68 |
| **No. Observations:** | 47 | **AIC:** | 1269. |
| **Df Residuals:** | 36 | **BIC:** | 1290. |
| **Df Model:** | 10 |  |  |
| **Covariance Type:** | nonrobust |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **const** | 1.279e+06 | 4.09e+05 | 3.128 | 0.003 | 4.5e+05 | 2.11e+06 |
| **x1** | -202.2296 | 75.269 | -2.687 | 0.011 | -354.882 | -49.577 |
| **x2** | 44.2832 | 215.197 | 0.206 | 0.838 | -392.157 | 480.724 |
| **x3** | 68.3333 | 86.827 | 0.787 | 0.436 | -107.759 | 244.426 |
| **x4** | 694.9607 | 313.767 | 2.215 | 0.033 | 58.612 | 1331.309 |
| **x5** | 618.4601 | 405.124 | 1.527 | 0.136 | -203.169 | 1440.089 |
| **x6** | 2.4818 | 36.713 | 0.068 | 0.946 | -71.975 | 76.938 |
| **x7** | -76.4581 | 83.207 | -0.919 | 0.364 | -245.211 | 92.295 |
| **x8** | -107.9203 | 52.508 | -2.055 | 0.047 | -214.412 | -1.429 |
| **x9** | -143.8713 | 286.891 | -0.501 | 0.619 | -725.713 | 437.970 |
| **x10** | 78.1301 | 8.623 | 9.061 | 0.000 | 60.642 | 95.618 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Omnibus:** | 5.553 | **Durbin-Watson:** | 1.609 |
| **Prob(Omnibus):** | 0.062 | **Jarque-Bera (JB):** | 2.900 |
| **Skew:** | 0.368 | **Prob(JB):** | 0.235 |
| **Kurtosis:** | 2.030 | **Cond. No.** | 5.87e+05 |

1. *Missing Value, Transformation*

|  |  |  |  |
| --- | --- | --- | --- |
| **Dep. Variable:** | y | **R-squared:** | 0.973 |
| **Model:** | OLS | **Adj. R-squared:** | 0.966 |
| **Method:** | Least Squares | **F-statistic:** | 134.7 |
| **Date:** | Sun, 10 Mar 2019 | **Prob (F-statistic):** | 5.77e-26 |
| **Time:** | 08:04:21 | **Log-Likelihood:** | 439.07 |
| **No. Observations:** | 48 | **AIC:** | -856.1 |
| **Df Residuals:** | 37 | **BIC:** | -835.6 |
| **Df Model:** | 10 |  |  |
| **Covariance Type:** | nonrobust |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **const** | 1.0004 | 2.58e-05 | 3.88e+04 | 0.000 | 1.000 | 1.000 |
| **x1** | 0.0441 | 0.016 | 2.717 | 0.010 | 0.011 | 0.077 |
| **x2** | -0.1808 | 0.026 | -7.011 | 0.000 | -0.233 | -0.129 |
| **x3** | -0.0381 | 0.027 | -1.434 | 0.160 | -0.092 | 0.016 |
| **x4** | -0.1750 | 0.087 | -2.002 | 0.053 | -0.352 | 0.002 |
| **x5** | 0.0634 | 0.100 | 0.634 | 0.530 | -0.139 | 0.266 |
| **x6** | -0.0100 | 0.009 | -1.142 | 0.261 | -0.028 | 0.008 |
| **x7** | -0.0056 | 0.029 | -0.194 | 0.847 | -0.065 | 0.053 |
| **x8** | 0.0042 | 0.015 | 0.283 | 0.779 | -0.026 | 0.035 |
| **x9** | 0.0642 | 0.073 | 0.884 | 0.383 | -0.083 | 0.211 |
| **x10** | -0.0292 | 0.001 | -22.071 | 0.000 | -0.032 | -0.026 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Omnibus:** | 36.351 | **Durbin-Watson:** | 1.844 |
| **Prob(Omnibus):** | 0.000 | **Jarque-Bera (JB):** | 108.769 |
| **Skew:** | -2.004 | **Prob(JB):** | 2.41e-24 |
| **Kurtosis:** | 9.190 | **Cond. No.** | 2.44e+04 |

1. *Missing Value, Outlier, & Transformation*

|  |  |  |  |
| --- | --- | --- | --- |
| **Dep. Variable:** | y | **R-squared:** | 0.995 |
| **Model:** | OLS | **Adj. R-squared:** | 0.993 |
| **Method:** | Least Squares | **F-statistic:** | 696.5 |
| **Date:** | Sun, 10 Mar 2019 | **Prob (F-statistic):** | 4.53e-38 |
| **Time:** | 08:09:48 | **Log-Likelihood:** | 500.30 |
| **No. Observations:** | 47 | **AIC:** | -978.6 |
| **Df Residuals:** | 36 | **BIC:** | -958.2 |
| **Df Model:** | 10 |  |  |
| **Covariance Type:** | nonrobust |  |  |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **coef** | **std err** | **t** | **P>|t|** | **[0.025** | **0.975]** |
| **const** | 1.0002 | 8.69e-06 | 1.15e+05 | 0.000 | 1.000 | 1.000 |
| **x1** | -0.0289 | 0.005 | -6.316 | 0.000 | -0.038 | -0.020 |
| **x2** | 0.0043 | 0.009 | 0.476 | 0.637 | -0.014 | 0.023 |
| **x3** | 0.0040 | 0.006 | 0.643 | 0.524 | -0.009 | 0.016 |
| **x4** | 0.0150 | 0.021 | 0.718 | 0.477 | -0.027 | 0.057 |
| **x5** | 0.0699 | 0.022 | 3.110 | 0.004 | 0.024 | 0.115 |
| **x6** | -0.0091 | 0.002 | -4.608 | 0.000 | -0.013 | -0.005 |
| **x7** | -0.0022 | 0.007 | -0.333 | 0.741 | -0.015 | 0.011 |
| **x8** | -0.0054 | 0.003 | -1.592 | 0.120 | -0.012 | 0.001 |
| **x9** | 0.0298 | 0.016 | 1.823 | 0.077 | -0.003 | 0.063 |
| **x10** | -0.0100 | 0.001 | -12.802 | 0.000 | -0.012 | -0.008 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Omnibus:** | 0.569 | **Durbin-Watson:** | 1.545 |
| **Prob(Omnibus):** | 0.752 | **Jarque-Bera (JB):** | 0.415 |
| **Skew:** | -0.225 | **Prob(JB):** | 0.813 |
| **Kurtosis:** | 2.907 | **Cond. No.** | 2.42e+04 |

Dari keempat model diatas, didapatkan nilai R-*squared* model dari data yang sama melalui *pre-processing* data yang berbeda, yaitu mengatasi *missing value*, mengatasi *missing value* dan *outlier*, mengatasi *missing value* dan melakukan transformasi, serta mengatasi *missing value*, *outlier*, dan melakukan transformasi berturut-turut adalah 0.733, 0.913, 0.973, 0.995. Apabila nilai R-*squared* semain mendekati angka satu, maka model yang dihasilkan oleh regresi tersebut semakin baik, karena R-*squared* mengindikasikan besarnya kombinasi variabel x (independen) secara bersama-sama mempengaruhi nilai variabel y (dependen). Oleh karena itu, dapat disimpulkan bahwa model regresi terbaik adalah data yang melalui *pre-processing* data berupa mengatasi *missing value*, *outlier*, dan melakukan transformasi dengan nilai R-*squared* terbesar, yaitu 0.995.