1. Model 1 = Dengan deteksi missing value dan inputasi dengan *Mean*

The regression equation is

Compressive strength = - 2633113 - 641 Cement + 17187 Blast Furnace Slag

+ 4641 Fly Ash + 8314 Water - 31228 Superplastic

- 19 Coarse Aggregate + 265 Fine Aggregate - 3392 Age

Predictor Coef SE Coef T P

Constant -2633113 2365079 -1.11 0.266

Cement -641 3379 -0.19 0.850

Blast Furnace Slag 17187 7335 2.34 0.019

Fly Ash 4641 13012 0.36 0.721

Water 8314 7026 1.18 0.237

Superplastic -31228 105378 -0.30 0.767

Coarse Aggregate -19 1038 -0.02 0.985

Fine Aggregate 265 1377 0.19 0.848

Age -3392 9674 -0.35 0.726

S = 13371037 R-Sq = 0.8% R-Sq(adj) = 0.0%

Analysis of Variance

Source DF SS MS F P

Regression 8 1.43826E+15 1.79783E+14 1.01 0.430

Residual Error 1021 1.82539E+17 1.78785E+14

Total 1029 1.83977E+17

1. Model 2 = Dengan menghapus outlier

The regression equation is

Compressive strength\_1 = 102616 + 83 Cement\_1 + 21 Blast Furnace Slag\_1

+ 248 Fly Ash\_1 + 106 Water\_1 - 13209 Superplastic\_1

- 68.8 Coarse Aggregate\_1 + 64 Fine Aggregate\_1

- 654 Age\_1

Predictor Coef SE Coef T P

Constant 102616 209240 0.49 0.624

Cement\_1 83.4 247.4 0.34 0.736

Blast Furnace Slag\_1 20.7 575.2 0.04 0.971

Fly Ash\_1 248 1127 0.22 0.826

Water\_1 105.9 506.5 0.21 0.835

Superplastic\_1 -13209 9688 -1.36 0.173

Coarse Aggregate\_1 -68.75 75.64 -0.91 0.364

Fine Aggregate\_1 63.7 102.7 0.62 0.535

Age\_1 -654 1720 -0.38 0.704

S = 938713 R-Sq = 0.4% R-Sq(adj) = 0.0%

Analysis of Variance

Source DF SS MS F P

Regression 8 3.30782E+12 4.13478E+11 0.47 0.878

Residual Error 939 8.27430E+14 8.81182E+11

Total 947 8.30737E+14

1. Model 3 = Data dideteksi dengan missing value kemudian dilakukan transformasi

The regression equation is

y = 0.231 - 0.151 x1 - 0.193 x2 - 0.197 x3 - 0.139 x4 + 0.431 x5 - 0.295 x6

- 0.295 x7 + 0.0431 x8

Predictor Coef SE Coef T P

Constant 0.230795 0.007509 30.74 0.000

x1 -0.15134 0.01343 -11.27 0.000

x2 -0.19303 0.02664 -7.24 0.000

x3 -0.19692 0.02902 -6.79 0.000

x4 -0.13881 0.02303 -6.03 0.000

x5 0.4313 0.3934 1.10 0.273

x6 -0.29501 0.01056 -27.95 0.000

x7 -0.29505 0.01066 -27.67 0.000

x8 0.04309 0.04421 0.97 0.330

S = 0.0287902 R-Sq = 47.6% R-Sq(adj) = 47.2%

Analysis of Variance

Source DF SS MS F P

Regression 8 0.768397 0.096050 115.88 0.000

Residual Error 1021 0.846283 0.000829

Total 1029 1.614680

1. Model 4 = Data dideteksi dengan missing value, menghapus oulier kemudian dilakukan transformasi

The regression equation is

y\_1 = 0.184 - 0.111 x1\_1 - 0.156 x2\_1 - 0.221 x3\_1 - 0.0830 x4\_1 + 0.679 x5\_1

- 0.233 x6\_1 - 0.240 x7\_1 + 0.349 x8\_1

Predictor Coef SE Coef T P

Constant 0.183804 0.007763 23.68 0.000

x1\_1 -0.11140 0.01267 -8.79 0.000

x2\_1 -0.15553 0.02561 -6.07 0.000

x3\_1 -0.22061 0.03474 -6.35 0.000

x4\_1 -0.08301 0.02226 -3.73 0.000

x5\_1 0.6786 0.4214 1.61 0.108

x6\_1 -0.23334 0.01075 -21.71 0.000

x7\_1 -0.24011 0.01102 -21.79 0.000

x8\_1 0.34910 0.06992 4.99 0.000

S = 0.0252681 R-Sq = 40.2% R-Sq(adj) = 39.7%

Analysis of Variance

Source DF SS MS F P

Regression 8 0.403857 0.050482 79.07 0.000

Residual Error 939 0.599531 0.000638

Total 947 1.003388

1. Kesimpulan

|  |  |  |  |
| --- | --- | --- | --- |
| Model ke | S | R-sq | R-sq(adj) |
| 1 | 13371037 | 0.80% | 0.00% |
| 2 | 938713 | 0.40% | 0.00% |
| 3 | 0.0287902 | 47.60% | 47.20% |
| 4 | 0.0252681 | 40.20% | 39.70% |

Berdasarkan tabel diatas dapat diketahui bahwa model yang paling baik adalah model ke 3 dengan data yang dideteksi dengan missing value kemudian dilakukan transformasi, dengan nilai *R-square* paling tinggi sebesar 47.6% dengan nilai s sebesar 0,0287902