**Perbandingan kebaikan Model regresi linier**

Pada tugas 3 data mining kali ini akan membandingkan 4 Model regresi linier yang didapatkan dari proses *preprocessing data* yang meliputi

1. Model 1 = Data baru setelah melalui deteksi *missing value*
2. Model 2 = Data baru setelah melalui deteksi *oulier*
3. Model 3 = Data baru setelah melalui deteksi *missing value* dan transformasi
4. Model 4 = Data baru setelah melalui deteksi *missing value*, *outlier* dan transformasi

yang menghasilkan nilai kebaikan Model regresi sebagai berikut

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Model ke- | S | R-sq | R-sq(adj) | R-sq(pred) |
| 1 | 3866.1 | 76.57% | 75.86% | 73.49% |
| 2 | 2847.04 | 61.74% | 60.40% | 57.73% |
| 3 | 0.0073345 | 94.78% | 94.63% | 94.26% |
| 4 | 0.0059909 | 96.20% | 96.06% | 95.86% |

Dari tabel diatas dapat dilihat bahwa Model ke 4 memiliki nilai *R-square* tertinggi yang menggambarkan bahwa data baru setelah melalui deteksi *missing value*, *outlier* dan transformasi memiliki Model regresi linier yang paling baik. Model ke 2 memiliki nilai *R-square* terendah, lebih rendah dari Model ke 1 karena terdapat data yang dihapus dari data 1 yang merupakan outlier yang menyebabkan nilai *R-square* Model ke 2 lebih rendah, namun dari nilai S yang dihasilkan menggambarkan Model ke 2 lebih baik dari pada Model ke 1 karena sudah terhindar dari outlier.