**Perbandingan Model Terbaik Regresi Linier**

Pada permasalahan pada kasus ini terdapat 4 model yang digunakan dalam data Australia Open 2013 yang diambil dari UCI Regression Datasets. Berikut merupakan penjelasan dari keempat model tersebut.

1. Model 1 : *missing value* sudah diatasi.
2. Model 2 : *missing value* dan *outlier* sudah diatasi.
3. Model 3 : transformasi data tanpa *missing value.*
4. Model 4 : transformasi data tanpa *missing value* dan *outlier.*

Setelah dilakukan permodelan dengan keempat model di atas maka dilakukan perbandingan untuk mencari model terbaik untuk data Australia Open 2013 menggunakan *software* Minitab. Model dikatakan baik jika nilai R-Square besar (mendekati 1) karena R-Square menjelaskan seberapa besar varians data dapat dijelaskan oleh model. Berikut merupakan hasil *output* Minitab regresi linier dari keempat model di atas.

1. **Regresi Linier setelah *Missing Value* Diatasi**

**Regression Analysis: TPW versus FSP; DBF; WNR; NPW**

The regression equation is

TPW = 54,7 + 0,168 FSP + 2,30 DBF + 0,884 WNR + 0,327 NPW

Predictor Coef SE Coef T P

Constant 54,66 23,22 2,35 0,020

FSP 0,1676 0,3457 0,48 0,629

DBF 2,2960 0,8031 2,86 0,005

WNR 0,8842 0,1311 6,75 0,000

NPW 0,3275 0,2358 1,39 0,167

S = 26,5425 R-Sq = 40,0% R-Sq(adj) = 38,0%

1. **Regresi Linier setelah *Missing Value* dan *Outlier* Diatasi**

**Regression Analysis: TPW\_1 versus FSP\_1; DBF\_1; WNR\_1; NPW\_1**

The regression equation is

TPW\_1 = 54,8 + 0,183 FSP\_1 + 2,49 DBF\_1 + 0,787 WNR\_1 + 0,386 NPW\_1

Predictor Coef SE Coef T P

Constant 54,80 23,74 2,31 0,023

FSP\_1 0,1826 0,3485 0,52 0,601

DBF\_1 2,4877 0,8779 2,83 0,005

WNR\_1 0,7867 0,1403 5,61 0,000

NPW\_1 0,3863 0,2685 1,44 0,153

S = 26,5113 R-Sq = 34,5% R-Sq(adj) = 32,2%

1. **Transformasi Data setelah *Missing Value* Diatasi**

**Regression Analysis: 4 versus 0; 1; 2; 3**

The regression equation is

4 = 0,567 - 0,599 0 - 0,131 1 - 0,230 2 - 0,196 3

Predictor Coef SE Coef T P

Constant 0,566894 0,006270 90,42 0,000

0 -0,59867 0,01860 -32,18 0,000

1 -0,13062 0,09196 -1,42 0,158

2 -0,23022 0,02032 -11,33 0,000

3 -0,19585 0,02634 -7,44 0,000

S = 0,0110045 R-Sq = 90,2% R-Sq(adj) = 89,9%

1. **Transformasi Data setelah *Missing Value* dan *Outlier* Diatasi**

**Regression Analysis: 4\_1 versus 0\_1; 1\_1; 2\_1; 3\_1**

The regression equation is

4\_1 = 0,568 - 0,606 0\_1 - 0,159 1\_1 - 0,227 2\_1 - 0,180 3\_1

Predictor Coef SE Coef T P

Constant 0,567783 0,006229 91,16 0,000

0\_1 -0,60602 0,01859 -32,59 0,000

1\_1 -0,15950 0,09525 -1,67 0,097

2\_1 -0,22704 0,02015 -11,27 0,000

3\_1 -0,18032 0,02882 -6,26 0,000

S = 0,0108315 R-Sq = 90,6% R-Sq(adj) = 90,3%

Berikut merupakan perbandingan R-Square dari *output* Minitab.

Tabel 1. Perbandingan Nilai R-Sq untuk Setiap Model

|  |  |
| --- | --- |
| **Model** | **Nilai R-Sq** |
| 1 | 40,0% |
| 2 | 34,5% |
| 3 | 90,2% |
| 4 | 90,6% |

Dari Tabel 1 dapat diketahui bahwa model 4 di mana data sudah ditransformasi, tanpa *missing value* dan *outlier,* memiliki nilai R-Square paling besar yaitu 90,6% dibandingkan dengan model lain maka dapat disimpulkan bahwa model 4 merupakan model terbaik untuk diterapkan pada data Australia Open 2013. Sedangkan model 2 memiliki nilai R-Square terendah yaitu sebesar 34,5% karena pada model ini masih terdapat nilai *outlier* yang menyebabkan nilai R-Square rendah daripada model lainnya.