Data Analysis Report

# Raw data summary

The underlying data has 1000 observations with 6 variables. The raw data was split into training and testing data using 0.9 partition, which resulted in a training data size of 900 and a testing data size of 100. GaussianProcess was used as the for data analysis with the training dataset.

# Data analysis summary

The goodness-of-fit measured in adj R-squared is 0.9999950829790123. Based on the best model selected, the prediction error on the testing data is 1.113575397747186. The residual plot is attached in the following section

# The performance measure table

|  |  |  |
| --- | --- | --- |
| Performance measure index | Performance measure name | Value |
| 1 | goodness-of-fit | 0.9999950829790123 |
| 2 | training rmse | 0.0018987047772700831 |
| 3 | testing rmse | 1.113575397747186 |

# The residual plot

