Data Analysis Report

# Raw data summary

The underlying data has 53822 observations with 391 variables. The raw data was split into training and testing data using 0.9 partition, which resulted in a training data size of 48439 and a testing data size of 5383. Logistic regression was used as the for data analysis with the training dataset.

# Data analysis summary

The goodness-of-fit measured in adj R-squared is 0.251512275651. Based on the best model selected, the prediction accuracy on the testing data is 0.851012446591. The residual plot is attached in the following section

# The performance measure table

|  |  |  |
| --- | --- | --- |
| Performance measure index | Performance measure name | Value |
| 1 | goodness-of-fit | 0.251512275651 |
| 2 | training accuracy | 0.854703854332 |
| 3 | testing accuracy | 0.851012446591 |
| 4 | training type I error | 0.41517166718218373 |
| 5 | training type II error | 0.04700481595178416 |

# The residual plot

