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. Elastic net was used as the for data analysis and its tuning parameters were selected using 5-fold cv with the training dataset.

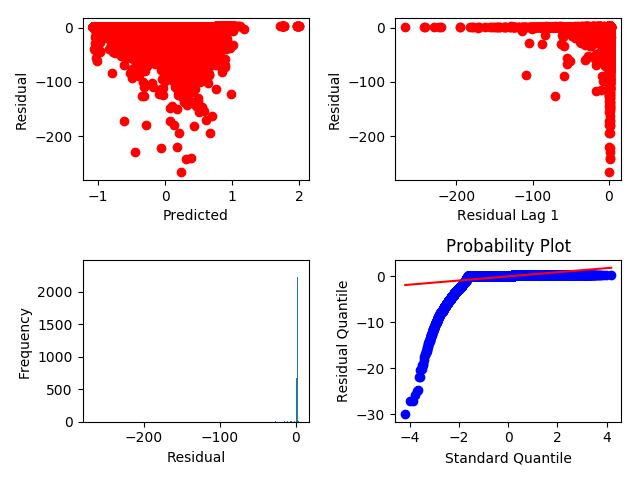
# Data analysis summary

The goodness-of-fit measured in adj R-squared is -0.00458790957118. Based on the best model selected, the prediction error on the testing data is 21.29322076211212. The residual plot is attached in the following section

# The performance measure table

|  |  |  |
| --- | --- | --- |
| Performance measure index | Performance measure name | Value |
| 1 | goodness-of-fit | -0.00458790957118 |
| 2 | training rmse | 8.91526085765032 |
| 3 | testing rmse | 21.29322076211212 |

# The residual plot



# The tuning plot

