|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Algorithm/Model** | **Unscaled Training Score** | **Unscaled Test Score** | **Scaled Training Score** | **Scaled Test Score** |
| Logistics Regression Model | 0.6497536945812807 | 0.8236920459378988 | 0.7078817733990148 | 0.8898341131433433 |
| Random Forest | 1.0 | 0.6461080391322841 | 1.0 | 0.646958740961293 |
|  |  |  |  |  |

This clearly shows that scaled data has better score compare to unscaled data for Logistics Regression Model model.

I also have similar prediction for logistic regression, this behavior as unscaled high value features have more influence on the model compare to low value features/attributes.

Scaling is done to Normalize data so that priority is not given to a particular feature. Role of Scaling is mostly important in algorithms that are distance based and require Euclidean Distance.

Random Forest is a tree-based model and hence **does not require** feature scaling.

Logistic Regression Model performed better than Random Forest model for Testing dataset for the selected dataset.