|  |  |  |  |
| --- | --- | --- | --- |
| readme |  |  |  |
|  |  |  |  |
| **To follow along with the report, the following RMD files are used:** | | | |
|  |  |  | Part\_1\_data\_info |
|  |  |  | \* Part\_2\_3\_clean\_convert\_scale |
|  |  |  |  |
|  | **Step 1: Pre-processing the data** |  |  |
|  | Basic file: |  | 1\_BASIC\_UNB |
|  | Then we have these added to it: |  | 1\_BASIC\_UNB\_SC |
|  |  |  | 1\_BASIC\_UNB\_dummy |
|  |  |  | 1\_BASIC\_UNB\_95CI |
|  |  |  | 1\_BASIC\_UNB\_95CI\_SC |
|  |  |  |  |
|  | **Step 2: Rebalance the Data** |  |  |
|  | Basic file: |  | 2\_UNB\_95CI\_SC |
|  | Then we have these added to it: |  | 2\_Bal\_DOWN\_95CI\_SC |
|  |  |  | 2\_Bal\_UP\_95CI\_SC |
|  |  |  | 2\_Bal\_ROSE\_95CI\_SC |
|  |  |  | 2\_Bal\_SMOTE\_95CI\_SC |
|  |  |  |  |
|  | **Step 3: Feature Selection:** |  |  |
|  | *file to test the models first:* |  | 3\_ALL\_Features |
|  | Basic file: |  | 2\_Bal\_SMOTE\_95CI\_SC |
|  | Then we have these added to it: |  | 3\_F1\_RFE\_AFTER\_SMOTE\_95CI\_SC |
|  |  |  | 3\_F2\_BORUTA\_AFTER\_SMOTE\_95CI\_SC |
|  |  |  | 3\_F3\_LVQ\_AFTER\_SMOTE\_95CI\_SC |
|  |  |  | 3\_F4\_ANOVA\_AFTER\_SMOTE\_95CI\_SC |
|  |  |  | 3\_F5\_AIC\_AFTER\_SMOTE\_95CI\_SC |
|  |  |  |  |
|  | **Step5 - Model Comparison** |  | 2\_Bal\_SMOTE\_95CI\_SC |
|  |  |  |  |
|  | \* Part\_2\_3\_clean\_convert\_scale file: components copied to other files |  |  |