

Assignment03 - Classification

General Instructions – Must Read

- **Number of Questions:** 01
- **Submission Guidelines:** You need to submit one python and one CSV file.
 1. **Single** csv file consist of your data extracted from Assignment02. File Name: <YourRollNum>.csv |
Example: 10155.csv
 2. **Single** python file that generate comparison tables and graphs. File Name: <YourRollNum>.py |
Example: 10155.csv
 - Your program must be run from **command line** only:
 - **Usages:** python <program.py> <InputDataFile>
 - **Example:** pytohn 10155.py 10155.csv
 - **Output:** You need to generates **09** multiple result files explain below.

Write a python program that generate the following result files.

1. Accuracy using Normalization

Output File Name: <output>-<YourRollNum>-<Normalization>.csv

Example: output-10155-Normalization.csv

Model	Accuracy without Normalization	Accuracy using Normalization			
		zscore	minmax	maxabs	robust
ada	0.42	0.81	0.09	0.38	0.91
dt	0.75	0.31	0.31	0.61	0.08
et	0.74	0.75	0.75	0.56	0.91
gbc	0.87	0.31	1.00	0.52	0.21
knn	0.50	0.69	0.54	0.30	0.78
lda	0.30	0.27	0.96	0.37	0.81
lightgbm	0.71	0.56	0.44	0.87	0.33
lr	0.05	0.10	0.00	0.02	0.56
nb	0.90	0.34	0.31	0.71	0.52
qda	0.13	0.57	0.12	0.19	0.64
rf	0.50	0.66	0.90	0.61	0.03
ridge	0.63	0.66	0.68	0.76	0.21
svm	0.33	0.72	0.89	0.00	0.34

2. Accuracy using **Feature Selection**

Output File Name: <output>-<YourRollNum>-<FeatureSelection>.csv

Example: output-10155-FeatureSelection.csv

Model	Accuracy without Feature Selection	Accuracy using Feature Selection			
		Classic = 0.2	Classic = 0.5	Boruta = 0.2	Boruta = 0.5
ada	0.42	0.81	0.09	0.38	0.91
dt	0.75	0.31	0.31	0.61	0.08
et	0.74	0.75	0.75	0.56	0.91
gbc	0.87	0.31	1.00	0.52	0.21
knn	0.50	0.69	0.54	0.30	0.78
lda	0.30	0.27	0.96	0.37	0.81
lightgbm	0.71	0.56	0.44	0.87	0.33
lr	0.05	0.10	0.00	0.02	0.56
nb	0.90	0.34	0.31	0.71	0.52
qda	0.13	0.57	0.12	0.19	0.64
rf	0.50	0.66	0.90	0.61	0.03
ridge	0.63	0.66	0.68	0.76	0.21
svm	0.33	0.72	0.89	0.00	0.34

3. Accuracy using **Outlier Removal**

Output File Name: <output>-<YourRollNum>-<OutlierRemoval>.csv

Example: output-10155-OutlierRemoval.csv

Model	Accuracy without Outlier Removal	Accuracy using Outlier Removal			
		Threshold=0.02	Threshold=0.04	Threshold=0.06	Threshold=0.08
ada	0.42	0.81	0.09	0.38	0.91
dt	0.75	0.31	0.31	0.61	0.08
et	0.74	0.75	0.75	0.56	0.91
gbc	0.87	0.31	1.00	0.52	0.21
knn	0.50	0.69	0.54	0.30	0.78
lda	0.30	0.27	0.96	0.37	0.81
lightgbm	0.71	0.56	0.44	0.87	0.33
lr	0.05	0.10	0.00	0.02	0.56
nb	0.90	0.34	0.31	0.71	0.52
qda	0.13	0.57	0.12	0.19	0.64
rf	0.50	0.66	0.90	0.61	0.03
ridge	0.63	0.66	0.68	0.76	0.21
svm	0.33	0.72	0.89	0.00	0.34

4. Accuracy using **PCA**

Output File Name: <output>-<YourRollNum>-<PCA>.csv

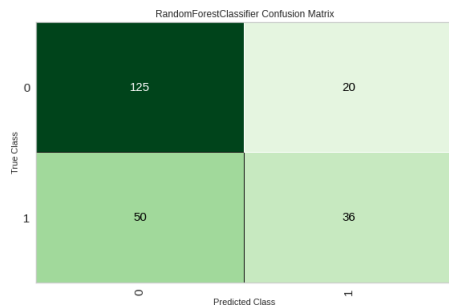
Example: output-10155-PCA.csv

Model	Accuracy without PCA	Accuracy using PCA		
		Method=linear	Method=kernel	Method=incremental
ada	0.42	0.81	0.09	0.38
dt	0.75	0.31	0.31	0.61
et	0.74	0.75	0.75	0.56
gbc	0.87	0.31	1.00	0.52
knn	0.50	0.69	0.54	0.30
lda	0.30	0.27	0.96	0.37
lightgbm	0.71	0.56	0.44	0.87
lr	0.05	0.10	0.00	0.02
nb	0.90	0.34	0.31	0.71
qda	0.13	0.57	0.12	0.19
rf	0.50	0.66	0.90	0.61
ridge	0.63	0.66	0.68	0.76
svm	0.33	0.72	0.89	0.00

5. Graph for **Confusion Matrix** by Best Model

Output File Name: <output>-<YourRollNum>-<ConfusionMatrix>.png

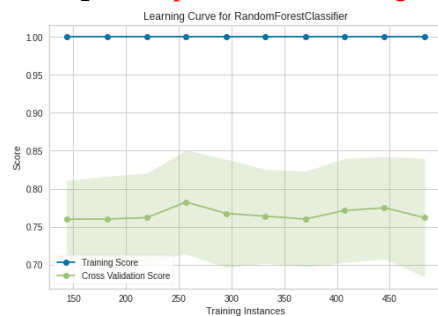
Example: output-10155-ConfusionMatrix.png



6. Graph for **Learning Curve** by Best Model

Output File Name: <output>-<YourRollNum>-<LearningCurve>.png

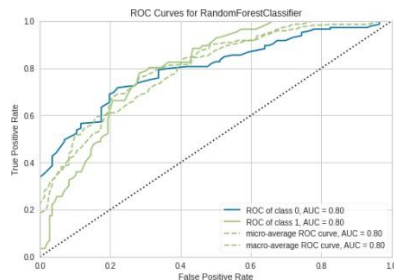
Example: output-10155-LearningCurve.png



7. Graph for **AUC Curve** by Best Model

Output File Name: <output>-<YourRollNum>-<AUC>.png

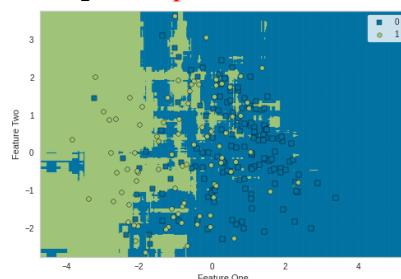
Example: output-10155-AUC.png



8. Graph for **Decision Boundary** by Best Model

Output File Name: <output>-<YourRollNum>-<DecisionBoundary>.png

Example: output-10155-DecisionBoundary.png



9. Graph for **Feature Importance** by Best Model

Output File Name: <output>-<YourRollNum>-<FeatureImportance>.png

Example: output-10155-FeatureImportance.png

