LABEL ENCODING AND Z_SCORE

ALGORITHM	PRE-PROCESSED DATA ACCURACY	PRE-PROCESSED DATA + HYPER PARAMETER TUNING
LOGISTIC REGRESSION	Accuracy of train data = 85.48707753479125 % Accuracy of test data = 87.03703703703704 %	Accuracy of train data = 85.88469184890656 % Accuracy of test data = 86.11111111111111 %
SUPPORT VECTOR MACHINE	Accuracy of train data = 84.89065606361828 % Accuracy of test data = 86.11111111111111 %	Accuracy of train data = 92.44532803180915 % Accuracy of test data = 85.64814814814815 %
DECISION TREE	Accuracy of train data = 100.0 % Accuracy of test data = 87.5 %	Accuracy of train data = 90.25844930417495 % Accuracy of test data = 86.57407407407408 %
NAÏVE BAYES	Accuracy of train data = 85.2882703777336 % Accuracy of test data = 87.03703703703704 %	accuracy_score on test dataset : 0.8935185185185185
RANDOM FOREST	Accuracy of train data = 100.0 % Accuracy of test data = 89.81481481481481 %	Accuracy of train data = 100.0 % Accuracy of test data = 90.74074074074075 %
K NEAREST NEIGHBOUR	Accuracy of train data = 89.26441351888667 % Accuracy of test data = 84.72222222222221 %	Accuracy of train data = 89.06560636182903 % Accuracy of test data = 86.57407407407408 %

ONE-HOT ENCODING AND Z_SCORE

ALGORITHM	PRE-PROCESSED DATA ACCURACY	PRE-PROCESSED DATA + HYPER PARAMETER TUNING
LOGISTIC REGRESSION	Accuracy of train data = 85.88469184890656 % Accuracy of test data = 85.64814814814815 %	Accuracy of train data = 86.08349900596421 % Accuracy of test data = 87.5 %
SUPPORT VECTOR MACHINE	Accuracy of train data = 84.89065606361828 % Accuracy of test data = 86.11111111111111 %	Accuracy of train data = 92.44532803180915 % Accuracy of test data = 85.648148148148 15 %
DECISION TREE	Accuracy of train data = 100.0 % Accuracy of test data = 86.11111111111111 %	Accuracy of train data = 90.25844930417495 % Accuracy of test data = 86.57407407407408 %
NAÏVE BAYES	Accuracy of train data = 57.65407554671968 % Accuracy of test data = 69.9074074074074 %	accuracy_score on test dataset : 0.8379629629629629
RANDOM FOREST	Accuracy of train data = 100.0 % Accuracy of test data = 88.42592592592592 %	Accuracy of train data = 100.0 % Accuracy of test data = 87.5 %
K NEAREST NEIGHBOUR	Accuracy of train data = 89.26441351888667 % Accuracy of test data = 84.72222222222221 %	Accuracy of train data = 89.06560636182903 % Accuracy of test data = 86.57407407407408 %

LABEL ENCODING AND IQR METHOD

ALGORITHM	PRE-PROCESSED DATA ACCURACY	PRE-PROCESSED DATA + HYPER PARAMETER TUNING
LOGISTIC REGRESSION	Accuracy of train data = 83.9851024208566 % Accuracy of test data = 86.14718614718615 %	Accuracy of train data = 85.28864059590316 % Accuracy of test data = 87.44588744588745 %
SUPPORT VECTOR MACHINE	Accuracy of train data = 86.03351955307262 % Accuracy of test data = 81.38528138528139 %	Accuracy of train data = 87.33705772811918 % Accuracy of test data = 82.25108225108225 %
DECISION TREE	Accuracy of train data = 100.0 % Accuracy of test data = 83.54978354978356 %	Accuracy of train data = 88.8268156424581 % Accuracy of test data = 86.14718614718615 %
NAÏVE BAYES	Accuracy of train data = 84.17132216014897 % Accuracy of test data = 87.01298701298701 %	accuracy_score on test dataset : 0.8571428571428571
RANDOM FOREST	Accuracy of train data = 100.0 % Accuracy of test data = 90.47619047619048 %	Accuracy of train data = 100.0 % Accuracy of test data = 90.04329004329004 %
K NEAREST NEIGHBOUR	Accuracy of train data = 88.64059590316573 % Accuracy of test data = 85.28138528138528 %	Accuracy of train data = 92.17877094972067 % Accuracy of test data = 86.14718614718615 %

ONE-HOT ENCODING AND IQR METHOD

ALGORITHM	PRE-PROCESSED DATA ACCURACY	PRE-PROCESSED DATA + HYPER PARAMETER TUNING
LOGISTIC REGRESSION	Accuracy of train data = 83.79888268156425 % Accuracy of test data = 85.28138528138528 %	Accuracy of train data = 85.84729981378027 % Accuracy of test data = 86.14718614718615 %
SUPPORT VECTOR MACHINE	Accuracy of train data = 85.6610800744879 % Accuracy of test data = 81.38528138528139 %	Accuracy of train data = 87.33705772811918 % Accuracy of test data = 82.25108225108225 %
DECISION TREE	Accuracy of train data = 100.0 % Accuracy of test data = 83.98268398268398 %	Accuracy of train data = 88.8268156424581 % Accuracy of test data = 86.14718614718615 %
NAÏVE BAYES	Accuracy of train data = 57.16945996275605 % Accuracy of test data = 62.77056277056276 %	accuracy_score on test dataset : 0.8181818181818182
RANDOM FOREST	Accuracy of train data = 100.0 % Accuracy of test data = 88.74458874458875 %	Accuracy of train data = 100.0 % Accuracy of test data = 87.01298701298701 %
K NEAREST NEIGHBOUR	Accuracy of train data = 88.64059590316573 % Accuracy of test data = 85.28138528138528 %	Accuracy of train data = 92.17877094972067 % Accuracy of test data = 86.14718614718615 %