Outliers : Only one outlier in the cholestrol feature
Duplicates : Only one duplicate sample found

Risk of Heart Attack Prediction

	PARAMETERS		PERFORMANCE					
MODELS	Test size	Random state	Outlier removal	Duplicate values removal	Scaling features (Standard Scaler)	Accuracy Score (Training)	Accuracy Score (Testing)	Precision(P) Recall(R) F1 Score(F1)
Logistic Regression	0.2	14	No effect. As the feature had insignificant correlation with the target variable.	Positively affected the performance by just a little.	Did not improve the performance of the model.	83.8	93.4	P - 94 R - 93 F1 - 93
Decision Tree Classifier	0.2	4	No effect. Tree based methods are insensitive to outliers.	Positive effect. Duplicate values put weight on the nodes and can make them biased.	No effect. Tree based methods are not sensitive to the variance in the data.	100	85.2	P - 85 R - 84 F1 - 83
Random Forest Classifier	0.2	8		No effect in the performance was observed.		100	91.8	P - 92 R - 92 F1 - 92
Support Vector Machine Classifier	0.2	13	Postive effect. Since SVM is sensitive to outliers.	Did not improve the performance of the model. SVM is not generally sensitive to duplicates.	Positive effect. Feature scaling is important in SVM as it uses data to find margins around hyperplanes and gets biased for variance in high values.	90	95	P - 95 R - 95 F1 - 95


Heart Dataset
120
100
80
60
40
20
0
83.8
93.4
100
85.2
100
91.8
90
95
77
70
Logistic Regression
Decision Tree
Random Forest
SVM
KNN
Training
83.8
100
100
90
77
Testing
93.4
85.2
91.8
95
70
Training
Testing