Classification Performance on Heart Disease Dataset

# 1. Explanation of Selected CI Techniques

## a) SMOTE (Synthetic Minority Over-sampling Technique)

SMOTE is a widely used technique to handle class imbalance. It works by generating synthetic samples for the minority class by interpolating between existing minority class instances. This helps to create a more balanced dataset for training classifiers.

## b) Random Undersampling

Random undersampling involves randomly removing samples from the majority class to balance the dataset. This technique can lead to loss of information, as many samples from the majority class are discarded.

## c) SMOTEENN (SMOTE + Edited Nearest Neighbors)

SMOTEENN is a hybrid technique that combines SMOTE and Edited Nearest Neighbors (ENN). After applying SMOTE to generate synthetic samples, ENN is used to clean the dataset by removing noisy examples, thereby improving the quality of the resampled dataset.

# 2. Impact of CI Solutions on Classification Performance

The impact of CI techniques on classification performance can be observed through various metrics such as ROC-AUC, Precision, Recall, and F1 Score. Below, we compare the performance of different classifiers before and after applying CI techniques.



# 3. Impact of Different Algorithms

Different algorithms may respond differently to CI techniques. Some algorithms, like Decision Trees, may benefit significantly from techniques like SMOTE, while others, like Logistic Regression, may see more modest improvements.

The performance metrics (ROC-AUC, Precision, Recall, F1 Score) for each algorithm before and after applying CI techniques are compared to assess the impact of the choice of algorithm on the effectiveness of the CI techniques.