Classification of Arrhythmia using ECG Data.

Objective: To distinguish between the presence and absence of cardiac arrhythmia and to classify it in one of the 16 groups. For the time being, there exists a computer program that makes such a classification. However, there are differences between the cardio log's and the program's classification. Taking the cardio log's as a gold standard we aim to minimize this difference by means of machine learning tools.

Dataset:

This data is extracted from the UCI Repository available at:

https://archive.ics.uci.edu/ml/datasets/Arrhythmia

The following is a description of our dataset:

- **of attributes (Columns):** 280
- **of instances (Rows):** 452

Expected Algorithm:

- Logistic Regression
- Tree classifiers
- Support Vector Machines (SVM)
- Naive Bayes
- Random Forest
- KNN

Expected Outcome: Prediction and classification of Cardiac Arrhythmia with maximum accuracy.

Application: These machine learning techniques can be deployed in hospitals where a large dataset is available and can help the doctors in Diagnosis of Cardiac Arrhythmia and to cut down the number of casualties due to heart diseases in the future.

IMPLEMENTATION & RESULT ANALYSIS

When trained over original data, Kernelized SVM proved to be the best among other classifiers in terms of recall value, with an accuracy percent of 79.12%. Also, Logistic Regression showed better training accuracy.

After performing Principal Component Analysis, when we trained our models, we found no improvement in KNN results, also, Random Forest too did not yield better results. However, we obtained improvised results from Kernelized SVM model with an accuracy of **80.21%**.