

# ***Predict Readmission for Heart Failure Using Machine Learning***

## **Description:**

Heart failure is a chronic disease that affects millions of people in the US. Its treatment costs \$30.7 billion annually, expected to increase to \$50 million by 2030. Heart failure is also becoming the cause of high mortality and patient readmission rate. According to different sources, the heart failure readmission rate is about 23%. Thus, the primary objective is to reduce the readmission rate, which will help reduce the annual expenditure and hospital penalties.

In this competition, you will identify the essential features from available datasets to automate a process to predict heart failure readmission patients within 90 days. You must develop a supervised machine learning-based classification approach to predict patients' readmission days. Do not limit yourself to the Machine Learning approach; use deep learning, model optimization, and model reliability. Ideally, you must implement a top-notch approach to develop an optimized model to overcome overfitting and bias-variance trade-offs, ensuring balanced evaluation measures. Multiple models can be used to increase the accuracy.

Your work will contribute towards the hospital readmission reduction program initiated by CMS (Centre of Medicare & Medicaid Services). The implementation of the model will predict the patients' readmission (PredictedReadmissionWithin\_90Days). You will perform a crucial variable analysis on univariate, bivariate, and multivariate data. You can apply necessary data analysis, profiling, and visualization techniques to find correlations, covariance, causality (if required), and other insights from different variables. The data analysis and visualization steps will have the highest weightage in the final scoring.

## **Tasks:**

1. Data Preparation
2. Data Analysis (Univariate, Bivariate, Multivariate)
3. Oversampling Techniques
4. Features Engineering
5. Model Training
6. Model Testing

## **Data:**

The dataset is well-structured and about 70% is ready to train a model, hence convenient and time-saving. You need to go through pre-processing and visualization steps necessary for the machine learning process. The dataset description is mentioned in the feature\_description.xlsx file.

**Target Variable:** ReadmissionWithin\_90Days

**Training Set:** 8481 Examples

**Testing Set:** 2121 Examples

**Evaluation:**

Evaluation Criteria:

- Data Analysis,
- F1 score/ ROC curve,
- Training-validation accuracy/loss,
- TP if it exists in both a ground truth and a prediction,
- FN if it exists in a ground truth but not a prediction, and,
- FP if it exists in a prediction but not ground truth.

We are expecting balanced results with all readmission days in buckets.