Heart Attack Prediction Machine Learning Project

**Problem Statement**

The aim of this project is to analyze and understand various reasons for heart attack and build a predictive model to proactively identify the same.

**Dataset**

Data dictionary can be found [here](https://daxg39y63pxwu.cloudfront.net/hands-on-labs/3471/heart-attack-prediction-machine-learning-project/datasets/Data+Dictionary.pdf).

Kindly download the data from [here](https://daxg39y63pxwu.cloudfront.net/hands-on-labs/3471/heart-attack-prediction-machine-learning-project/datasets/heart.csv).

**Projects for this Exercise**

* [Build a Credit Default Risk Prediction Model with LightGBM](https://www.projectpro.io/project-use-case/credit-risk-prediction-modelling-in-python-with-lightgbm)

**Tasks**

1. Check the count and percentage of missing values in each column. Drop columns which have more than 80% null values.
2. EDA:
   * Find top 3 highly correlated numerical features with the target variable. Plot Correlation matrix.
   * How do you check outliers in each column? Create a function that takes in each column and caps the value in the Inter Quartile Range (effectively removing outliers).
   * How is age and heart attacks related? What age group of people are more prone to heart attacks?
   * Does Gender have any impact on heart attacks?
3. Identify the categorical variables (variables with less than 4 unique numbers) and use target encoding method to encode them.
4. Select features by taking union of top 7 features by building Decision Tree and Logistic regression models.
5. Build Classification model using LightGBM to predict the heart attacks:
   * Build an LGBM Classifier on the data (LGBM - 1).
   * Tune the hyperparameters and select the best model from the experiment (LGBM - 2)
   * Compare and contrast the evaluation metrics of the 2 models - precision, recall, accuracy, roc-auc score.

**FAQs**

**Q1. How do you check outliers in each column?**

Outliers can be detected by capping values based on the Interquartile Range (IQR) using a custom function.

**Q2. How are age and heart attacks related? What age group is more prone to heart attacks?**

Analyzing the data can reveal the relationship between age and heart attacks, identifying the age group with higher vulnerability.

**Q3. Does gender have any impact on heart attacks?**

The data analysis can provide insights into the impact of gender on heart attacks, highlighting any notable patterns or correlations.

**Features**

●Age : Age of the patient

● Sex : Sex of the patient

● exang: exercise induced angina (1 = yes; 0 = no)

● ca: number of major vessels (0-3)

● cp : Chest Pain type chest pain type

○ Value 1: typical angina

○ Value 2: atypical angina

○ Value 3: non-anginal pain

○ Value 4: asymptomatic

● trtbps : resting blood pressure (in mm Hg)

● chol : cholestoral in mg/dl fetched via BMI sensor

● fbs : (fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)

● rest\_ecg : resting electrocardiographic results

○ Value 0: normal

○ Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)

○ Value 2: showing probable or definite left ventricular hypertrophy by Estes'

criteria

● thalach : maximum heart rate achieved

● target : 0= less chance of heart attack 1= more chance of heart attack