**BSAN 6070 CA05 – Logistic Regression**

**MUHAMMAD AOUN HAMID**

**Necessary Packages**

import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.model\_selection import train\_test\_split

from sklearn import linear\_model

from sklearn.linear\_model import LogisticRegression

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier

from sklearn.ensemble import AdaBoostClassifier

from sklearn.ensemble import GradientBoostingClassifier

from sklearn.preprocessing import LabelEncoder

from sklearn import metrics

from sklearn.metrics import classification\_report, confusion\_matrix

from sklearn.metrics import roc\_curve

from sklearn.metrics import auc

from sklearn.metrics import plot\_roc\_curve

from sklearn.metrics import roc\_auc\_score

from sklearn.metrics import accuracy\_score

from sklearn.metrics import precision\_score

from sklearn.metrics import f1\_score

from sklearn.metrics import recall\_score

**Description**

Cardiovascular Disease (CVD) kills more people than cancer globally. A dataset of real heart patients collected from a 15 year heart study cohort is made available for this assignment. The dataset has 16 patient features. Note that none of the features include any Blood Test information.

**Required Dataset**

For this project the dataset can be downloaded on github at '[https://github.com/ArinB/CA05-B-Logistic-Regression/raw/master/cvd\_data.csv](https://github.com/ArinB/CA05-B-Logistic-Regression/raw/master/cvd_data.csv%20) '. However, within the code you can run the appropriate code to load the dataset.

**Additional Comments**

For any additional comments or improvements please contact the author