Heart Disease Predictions via Random Forest and Logistic Regression with SMOTE

A data science project that required the use of SMOTE for the construction of a random forest model and a logistic regression model.

About

This project centers on a business problem of rising health insurance claim costs focusing on a major contributor, heart disease. Two classifier models leverage survey data from the CDC's Behavioral Risk Factor Surveillance System to predict whether an individual has heart disease. Please note, the classes in the dataset are unbalanced; therefore, SMOTE was used to balance the classes.

Link

Dataset: https://www.kaggle.com/datasets/kamilpytlak/personal-key-indicators-of-heart-disease

Setup

This Python coding project was constructed on Jupyter Notebook. Please import the following libraries:

- Matplotlib
- Numpy
- Pandas
- Sklearn
- Imblearn
- Seaborn