Project Milestone 1

Kiran Chowdary Komati

College of Science and Technology, Bellevue University

DSC680-T301: Applied Data Science (2231-1)

Dr. Catherine Williams

December 2, 2022

Proposal:

Topic: The topic I have selected as my first project is “Heart Attack prediction”.

Business Problem:

In the current day fast paced modern world, the no. of cases of heart attack are on a rise. There are several factors that contribute to it. Using machine learning models, we’ll analyze different attributes of persons to predict if a person is prone to heart attack or not.

Data sets:

I’m using the data set “Heart Attack Analysis & Prediction Dataset “ from Kaggle. It contains one csv file heart.csv. This has 16472 records and 14 different attributes of a patient which are listed below.

Age : Age of the patient

Sex : Sex of the patient

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

ca: number of major vessels

cp : Chest Pain type

trtbps : resting blood pressure (in mm Hg)

chol : cholesterol in mg/dl fetched via BMI sensor

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

rest_ecg : resting electrocardiographic results

thalach : maximum heart rate achieved

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

Methods:

The analysis methods for this topic may change as the project progress. I’ll first perform some EDA to understand the data. If the data has any NULLs, I’ll eliminate the records that has any NULLs as it may not be appropriate to replace them with a mean or any other value as they are specific to individuals. Also, if the classes, in this case the gender is not balanced, I’ll use SMOTE or any other methodologies to balance them. I’m planning to use Logistic regression initially and may add other models such as Random Forest and Decision Tree to compare the accuracy against the logistic regression model.

Ethical considerations:

As this project is related to the health of individuals and as it contains sensitive information, I made sure to not use any PII information such as individual names. Also, the data needs to be presented in accurate form with out any modifications or misrepresentations. Also, steps must be taken to avoid any bias towards any gender.

Challenges/Issues: Potential challenges and issues must be identified as a best practice when exploring data science projects. I need to validate if the data is incomplete and must take measures to handle it ethically. Also, the feature engineering may pose challenges in determining the best features. Despite these potential challenges, I'm looking forward to this interesting data science project, "Heart Attack prediction."

References:

https://www.kaggle.com/datasets/rashikrahmanpritom/heart-attack-analysis-prediction-dataset