

Project Proposal:

Harigovind Padmarajan

Data Description: I would like to work on a dataset containing health-related information of individuals. The dataset contains 21 independent variables including, Diabetes_012, HighBP, HighChol, CholCheck, BMI, Smoker, Stroke, HeartDiseaseorAttack, PhysActivity, Fruits, Veggies, HvyAlcoholConsump, AnyHealthcare, NoDocbcCost, GenHlth, MentHlth, PhysHlth, DiffWalk, Sex, Age, Education, and Income. The target variable that I am interested in analyzing is HeartDiseaseorAttack, which is a binary variable indicating whether an individual has heart disease or has had a heart attack.

Data Source: [DiabetesPrediction](#)

The data was collected from the National Health Interview Survey (NHIS), which is an annual survey conducted by the Centers for Disease Control and Prevention (CDC) to assess the health status of the US population. The dataset contains information from 23,393 individuals who participated in the survey from 2014 to 2017. The survey was conducted through in-person interviews, and participants provided information on their health status, lifestyle, and demographic characteristics.

Project Objective: The objective of my project is to develop a predictive model that can accurately identify individuals at risk of developing heart disease or having a heart attack. I will explore the relationship between the target variable and the independent variables using various statistical techniques. I will also perform data preprocessing, feature selection, and model evaluation to ensure the accuracy and generalizability of the model. My goal is to provide insights that can help healthcare professionals identify individuals at risk of heart disease and develop effective prevention strategies.