INFO-659 Final group project proposal.

Abstract:

Our project focuses on creating a predictive model for identifying heart disease indicators and risk factors, utilizing Kaggle's extensive datasets. Heart disease is a major public health concern, being a leading cause of death in the United States for various racial groups. With approximately 47% of Americans possessing at least one of three significant risk factors—high blood pressure, high cholesterol, and smoking—early detection and risk factor prevention are vital. We will harness Kaggle's healthcare datasets, primarily derived from the CDC's Behavioral Risk Factor Surveillance System (BRFSS), to work with heart disease indicators and associated risk factors.

Our approach involves data cleaning, preprocessing, and predictor creation to ensure data quality and relevance. By applying machine learning algorithms like logistic regression and random forests, we will analyze from basic descriptive and build predictive models to identify individuals at risk of heart disease based on clinical and demographic characteristics.

By leveraging Kaggle's datasets and advanced analytics, our project will comprehensively address the data analytics process, offering insights and potential solutions for heart disease prevention and intervention.

Group Members:

1. Kassa Workineh
2. Primohammed Mahir
3. Hobbs Emily

https://www.cdc.gov/heartdisease/risk_factors.htm#:~:text=Several%20health%20conditions%2C%20your%20lifestyle,%2C%20high%20cholesterol%2C%20and%20smoking.