**Development and Validation of Cardiovascular Disease Risk Prediction Model**

**Background and Motivation**

According to the World Health Organization, Cardiovascular disease (CVD) is the leading cause of deaths globally. An estimated 17.9 million people died from heart disease in 2016, representing 31% of all global deaths. In the United States, about 655,000 people die from heart disease each year – that’s 1 in every 4 deaths. CVD includes [coronary artery diseases](https://en.wikipedia.org/wiki/Coronary_artery_disease) such as [angina](https://en.wikipedia.org/wiki/Angina_pectoris) and myocardial infarction, stroke, heart failure, hypertensive heart disease, rheumatic heart disease, etc.

The leading but controllable risk factors for CVD are:

* High blood pressure
* High cholesterol
* Cigarette smoking
* Diabetes
* Unhealthy diet and physical inactivity
* Overweight and obesity

Accurate risk assessment of an individuals’ propensity to develop CVD is crucial for the prevention of these conditions. Risk prediction models can be a component of CVD prevention and control efforts, because they can help to identify people at high risk of CVD who should benefit the most from preventive interventions. The goal of this study is to establish a CVD risk prediction models for adults and to collect comparative information on their relative prognostic performance. The present study will focus on drawing a scenario of CVD risk assessment methods and a stepwise approach to population risk stratification paying more emphasis to nonclinical factors.

**Dataset**

The data was collected from [NHANES](https://wwwn.cdc.gov/nchs/nhanes/Default.aspx) website from 2005-2016. The objective is to classify the individuals for their relative risk of CVD prediction based on data available on demographics, dietary information, laboratory results and survey questions.

**References:**

1. <https://www.who.int/news-room/fact-sheets/detail/cardiovascular-diseases-(cvds)>
2. https://www.cdc.gov/heartdisease/facts.htm