Cardiovascular disease (CVD) is a general term for heart and vascular diseases. Common cardiovascular diseases include hypertension (elevated blood pressure), coronary heart disease (heart attack), cerebrovascular disease (stroke), peripheral vascular disease, heart failure, rheumatic heart disease, congenital heart disease, and cardiomyopathy. According to the World Health Organization, approximately 17.5 million people died from cardiovascular disease in 2012, accounting for 31% of global deaths. Among these deaths, an estimated 7.4 million died from coronary heart disease and 6.7 million died from stroke; Due to the increasing number of patients with cardiovascular diseases, the diagnosis and treatment of cardiovascular diseases have become a major issue in the medical industry.

This Problem is based on Kaggle's publicly available cardiovascular patient diagnostic data.

(See <https://www.kaggle.com/datasets/pirogovskiy/cardio-train/data> for details)

Question:

1. Firstly, preprocess and conduct exploratory analysis on the data (See attached ‘files cardio\_train.csv’ with ‘data\_dictionary.xlsx’).
2. From the patient's physiological indicators, medical detection indicators, and subjective information provided by the patient, use the classification method (or variables methods) in machine learning to predict whether the patient has cardiovascular disease.
3. Finally, compare the prediction performance under different classifiers. And draw a conclusion.