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.

心血管疾病（CVD）是心脏和血管疾病的总称。常见的心血管疾病包括高血压（血压升高）、冠心病（心脏病发作）、脑血管疾病（中风）、外周血管疾病、心力衰竭、风湿性心脏病、先天性心脏病和心肌病。根据世界卫生组织的数据，2012年约有1750万人死于心血管疾病，占全球死亡人数的31%。在这些死亡病例中，大约有740万人死于冠心病，670万人死于中风；由于心血管疾病患者数量不断增加，心血管疾病的诊断和治疗已成为医疗行业的重大问题。这个问题基于Kaggle上公开的心血管患者诊断数据。

问题：

1. 首先，对数据进行预处理并进行探索性分析（参见附带的'files cardio\_train.csv'和'data\_dictionary.xlsx'）。

2. 利用患者的生理指标、医学检测指标和患者提供的主观信息，使用机器学习中的分类方法（或变量方法）来预测患者是否患有心血管疾病。

3. 最后，比较在不同分类器下的预测性能。并得出结论。