Exam#3

Big Data

Download [processed.hungarian.data](https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/processed.hungarian.data) data set from the following link <https://archive.ics.uci.edu/ml/machine-learning-databases/heart-disease/>

Write a scala code on spark to create a machine learning system that can predict someones heart condition (e.g. diagnosis of heart disease (angiographic disease status) -- Value 0: < 50% diameter narrowing, -- Value 1: > 50% diameter narrowing) based on two parameters – age and cholesterol level (e.g. serum cholestoral in mg/dl).

Q1) Implement using random forest machine learning model.

Q2) Implement using K-means clustering model.