UCCB3224 Data Mining Techniques

Mock Practical Test

Question:

Assume that you have been assigned a data mining task to build a model to predict whether a patient has low chance of heart attack or high chance of heart attack. This task involves applying data mining steps to evolve a model using a training set and eventually to test the model using a test set. The data has been provided to you in a csv file named "heart.csv". You will need to read the data using pandas package. You need to use Scikit Learn package to apply various data mining process. For visualization you may use Matplotlib and/or Seaborn

About this dataset

- Age : Age of the patient
- Sex : Sex of the patient (male/female)
- exang: exercise induced angina (yes; no)
- ca: number of major vessels (0-3)
- cp : Chest Pain type chest pain type
    - typical angina
    - atypical angina
    - non-anginal pain
    - asymptomatic
- trtbps : resting blood pressure (in mm Hg)
- chol : cholestoral in mg/dl fetched via BMI sensor
- fbs : (fasting blood sugar > 120 mg/dl) (true; false)
- rest_ecg : resting electrocardiographic results
    - Value 0: normal
    - Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV)
    - Value 2: showing probable or definite left ventricular hypertrophy by Estes' criteria
- thalach : maximum heart rate achieved
- oldpeak: Previous peak
- slp: the slope of the peak exercise ST segment
  -- Value 1: upsloping
  -- Value 2: flat
  -- Value 3: downsloping
- thall: Thalium Stress Test result ~ (0,3)
- target : low: low chance of heart attack; high = high chance of heart attack