**请用python完成以下内容：数据集/Volumes/yasen/CHARLS/2015/PSM/matched\_data.csv当中的变量Mental\_retardation的变量值1 Yes变成为数字1，2 No变成数字2，最后保存的原数据集当中。**

基线资料数据路径为/Volumes/yasen/cursor/matched\_data.csv，其中

自变量有：

年龄：Age为连续性变量

以下变量为分类变量：Sex Marital\_status ，Residence ，Smoking， Drinking Physical\_disabilities ，Mental\_retardation ，Vision\_problem ，Hearing\_problem Speech\_impediment ，Physical\_disabilities ，Mental\_retardation ，Vision\_problem Hearing\_problem ，Speech\_impediment ，difficulty\_with\_getting\_up ，difficulty\_with\_stooping ，difficulties\_with\_doing\_household\_chores ，difficulties\_with\_taking\_medications,new\_da007\_1\_到new\_da007\_14\_是14种慢性病，变量值1为有有此慢性病，2为无此慢性病。"new\_da007\_1\_": "Hypertension",

"new\_da007\_2\_": "Dyslipidemia",

"new\_da007\_3\_": "Diabetes",

"new\_da007\_4\_": "Cancer",

"new\_da007\_5\_": "Chronic Lung Disease",

"new\_da007\_6\_": "Liver Disease",

"new\_da007\_7\_": "Heart Disease",

"new\_da007\_8\_": "Stroke",

"new\_da007\_9\_": "Kidney Disease",

"new\_da007\_10\_": "Digestive Disease",

"new\_da007\_11\_": "Psychiatric Problems",

"new\_da007\_12\_": "Memory-related Disease",

"new\_da007\_13\_": "Arthritis",

"new\_da007\_14\_": "Asthma"

结局变量为fall\_2018为随访三年后是否发生跌倒，和fall\_2020为随访5年是否发生跌倒，0为未发生跌倒，1为发生了跌倒。

。我想利用**Random Forest** 模型训练发生跌倒的模型，通过训练使用的指标预测分别在3年和5年发生跌倒的风险，图片要美观，清晰，质量高，tiff格式。最后把模型部署到网站（用GitHub），医生和患者可以通过选择和输入预测指标预测，对每一个患者进行个精准预测3年和5年发生跌倒的风险。然后给每一个老年患者提高精准的预防跌倒的方案。