“数据仓库与数据挖掘”第二次作业

# 医疗数据分析挖掘

医生在进行疾病诊断时，通常会依据患者的基本信息、生命体征、检测数据以及患者当前的生理状态等信息，判断患者是否存在某种疾病以及疾病的严重程度。然而在一些特殊情况下，例如灾害现场存在大批量伤员，而专业的医护人员短缺，就希望有一种自动化医疗诊断工具对伤员的伤情进行分析，辅助医护人员进行救治决策。

本次实验数据集中，包含了患者的基本信息以及部分检测指标，希望通过回归模型或分类模型，实现自动化伤情诊断，以辅助医生进行决策。作业包含了回归（氧分压指标预测）和分类（气道梗阻严重程度分类）两个任务，可**任选其一**完成。

# 问题一：氧分压指标回归预测

1. **任务描述**

根据患者的基本信息（年龄、性别）和基本生命体征（呼吸率、心率等），构建回归模型预测患者发生气道梗阻的严重程度。

1. **数据描述**

训练集：datasets/regression\_train.xlsx

测试集：datasets/regression\_test.xlsx

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段描述 | 正常值参考范围 |
| PVID | Int | 患者唯一标识 |  |
| Age | Int | 年龄 |  |
| gender | Int | 性别 | 0女性 1男性 |
| HR | Int | 心率 | 60-100 |
| RR | Int | 呼吸率 | 12-24 |
| SBP | Int | 收缩压 | 80-120 |
| Spo2 | Int | 氧饱和度 | 95-100 |
| DBP | Int | 舒张压 | 60-90 |
| MAP | Int | 平均动脉压 | 70-105 |
| PO2 | Int | 需要预测的氧分压值 |  |

# 问题二：气道梗阻严重程度分类预测

1. **任务描述**

根据患者的基本信息（年龄、性别），基本生命体征（呼吸率、心率等）、检测化验指标（乳酸、碱剩余等）以及患者状态（昏迷状态、躁动状态）等信息，构建分类模型预测患者发生气道梗阻的严重程度。

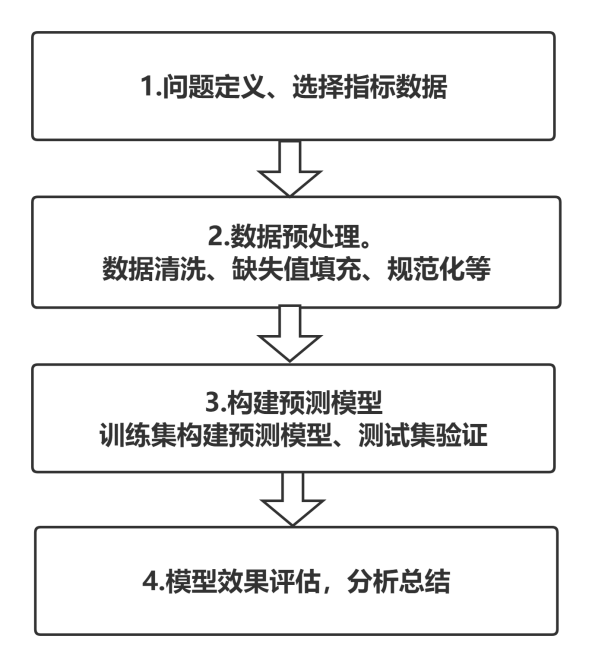
1. **数据描述**

训练集：datasets/classification\_train.xlsx

测试集：datasets/classification\_test.xlsx

|  |  |  |  |
| --- | --- | --- | --- |
| 字段名称 | 数据类型 | 字段描述 | 正常值参考范围 |
| PVID | Int | 患者唯一标识 |  |
| Age | Int | 年龄 |  |
| gender | Int | 性别 | 0女性 1男性 |
| HR | Int | 心率 | 60-100 |
| RR | Int | 呼吸率 | 12-24 |
| SBP | Int | 收缩压 | 80-120 |
| Spo2 | Int | 氧饱和度 | 95-100 |
| Po2 | Int | 氧分压测定 | 70-100 |
| Pco2 | Int | 二氧化碳分压测定 | 35-45 |
| BE | Float | 剩余碱测定 | -3-3 |
| LA | Float | 乳酸 | 0.7-2 |
| PH | Float | PH值 | 7.35-7.45 |
| BS | Float | 血糖 | 3.9-6.1 |
| Conscious | Float | 昏迷状态 | 1严重昏迷 0.5轻微昏迷 0未昏迷 |
| agitate | Int | 躁动状态 | 1躁动 0未躁动 |
| Label | Int | 需要预测的气道梗阻严重程度 |  |

# 作业步骤

****

# 作业要求

n **数据处理**：由于数据集中存在数据缺失和数据异常的情况，因此需要对数据进行一定的预处理，包括但不限于数据修复、数据清理等，可根据背景知识自行决定，例如负值或严重超出正常范围的值可认为无效，以及对数据进行规范化等。

n **模型算法**：使用**至少三种回归器（或分类器）**完成该回归（或分类）任务，可使用如 scikit-learn 或 matlab 等常用库中提供的回归器（或分类器）。

n **评价指标**：采用 K 折交叉验证评估模型算法；根据任务特点采用合理的评价指标，并说明选择依据 。

n **结果分析**：对比同一回归器（或分类器）的不同参数和不同回归器（或分类器）在该任务上的效果差异。

## n 实验报告：提交的设计报告文档应至少包含上述内容。报告文档页数最多不超过4页。文档中英文不限。

**提交的作业内容包括：**

1. 代码：实现作业要求所需的代码，编程语言不限
2. 文档：根据作业要求撰写文档，转换为pdf格式提交
3. 无需上传数据

## 请把以上内容打包成一个压缩文件上传。

作业须在 2022年 4 月 8 日 23:59:59（含）之前通过网络学堂提交。本次作业由个人独立完成，占课程总成绩的 25%。

文件名格式:学号\_姓名\_作业 2.zip，如 2021123456\_张小明\_作业2.zip