

---

# 期末考核大作业

---

---

# 交通事故中人员伤亡多分类预测

## □ 任务内容

- 根据交通事故中收集的信息对交通事故中的人员伤亡情况进行分类预测，预测分为四类：
  - 1 - Unharmed
  - 2 - Killed
  - 3 - Injured hospitalized
  - 4 - Slightly injured
- 分类的groundtruth在训练集和验证集中对应于“grav”列
- 文件说明：
  - train.csv 训练集
  - val.csv 验证集（包含groundtruth）
  - test.csv 测试集（不包含groundtruth）
  - Columns\_Description.txt 数据集各列属性说明文件
  - submission\_example.csv 分类预测结果提交文件样例
  - data-analysis-example.ipynb 数据分析样例程序
- 利用训练集和验证集进行学习并验证模型，然后在测试集上进行预测生成预测结果文件。预测结果文件为csv格式文件，包含两列“sample\_id”和“grav”，每一行中“grav”列为该行“sample\_id”对应的测试集中同一“sample\_id”样本的分类预测结果。具体格式参见“submission\_example.csv”

□ 提交内容包括以下内容，以zip或rar压缩打包，打包后的文件名以各自的姓名学号命名，格式为：“姓名\_学号.zip”或“姓名\_学号.rar”：

1. 作业报告以ipynb文件完成：包括程序、程序运行的实验结果和图表、以markdown格式写的说明和分析

- 用python进行编程，只可使用网上公开的库，程序可复现实验结果
- 不能只有最终模型，需包含思路、方法和算法等的说明以及中间实验结果
- 最终模型必需是只用训练集数据进行训练得到的模型
- 最终模型需在验证集进行验证，给出验证集上的Accuracy和每个类别的precision、recall和F1-score
- 最终模型需在测试集上进行预测，生成预测结果文件，以test\_result.csv命名，格式和submission\_example.csv的格式一致

2. 最终模型在测试集上进行预测的结果文件：test\_result.csv

3. 大作业汇报ppt（如需提供）

□ 将压缩文件以附件形式发送至cenfengcourses@163.com

□ 邮件主题以各自的姓名学号命名，格式为：[2022\_ML\_FP]\_姓名\_学号

---

## ■ Submission\_example.csv格式

sample\_id,grav

0,1

1,2

2,1

3,1

4,1

5,4

6,1

7,1

8,3

9,4

# 评分

---

## □ 期末大作业评分和作业提交截止时间：

- 准备得90分以上的同学，需在17周课上进行汇报，
  - 5分钟ppt对思路、方法和结果进行线上汇报，**严禁超时**
  - **ppt中不得出现程序代码**
  - 作业和ppt提交截止时间：17周周五(6月17日) 上午9:00
- 其余同学作业提交截止时间：18周周三(6月22日) 之前

## □ 期末大作业90分以下评分：

- ipynb报告：60%
- test\_result.csv：40%，以测试集上的Accuracy进行评分，最低要求：测试集上的Accuracy>0.5

# 注意事项

---

- ❑ 各位同学独立完成程序，并确保程序不被他人获取。任何抄袭将取消抄袭和被抄袭同学成绩。
- ❑ 可在班级微信讨论群中对技术问题进行交流，不可私下交流。任何私下交流（包括但不限于微信群中的私下交流）视为考核违规。
- ❑ 微信交流群只可进行技术理论分析交流，不可发表具体程序实现，发表程序代码的行为视为考核违规。