# 谁会提出保险索赔？

## ——请参考数学建模论文，完成大作业报告

### 序言

在保险行业，准确预测索赔的可能性对于风险评估和保单定价至关重要。然而，保险索赔数据集经常出现类别不平衡的问题，其中非索赔实例的数量远远超过实际索赔实例的数量。本次大作业，你将化身Python “保险侦探”，利用在课堂上掌握的数据分析与建模方法，让模型在真实的车险数据中精准捕捉风险信号，帮助保险公司准确识别高风险投保人，优化保费定价、降低理赔成本，并为广大车主带来更公平合理的保障。

准备好了吗？让我们一起体会数据驱动决策在现实中的重要价值，用代码为保险行业带来更智慧的未来！

**截至时间 2025 年 6 月 29 日 23:59 前**

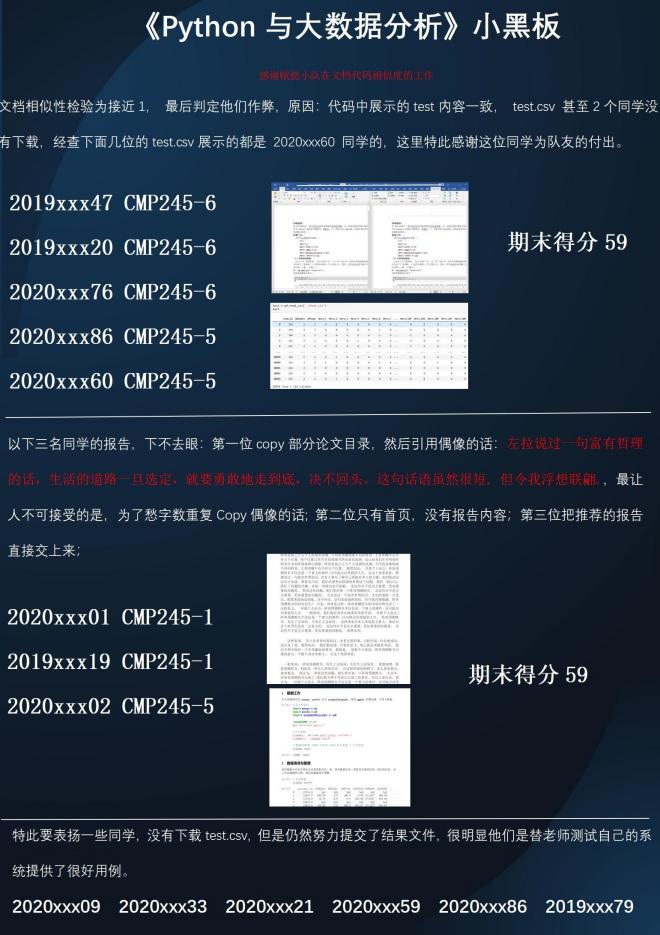
注意：（**务必注意！！！！！！！！！！！！！！！！不要错过截止时间**）

**1 需要提交：**

1. **结果文件，csv文件， 提交时间段登录见公告**
2. **首页，doc或pdf文件，截至时间登录见公告**
3. **正文、doc或pdf文件，截至时间登录见公告**
4. **代码、ipynb或py或zip压缩文件，截至时间登录见公告**

**2 请勿尝试考验老师对抄袭的厌恶程度和检测能力。如果被发现有抄袭嫌疑，无论是单人还是多人，最终成绩都将是59.99分。**

**3 请勿提交他人的结果文件或使用他人的测试集。我们会对提交结果进行甄别，一经发现将会严肃处理。**



### 数据介绍

本次大作业题目的train数据位于（6-数据集）文件夹中，test数据会在指定时间在期末考试.exe中获取，请在这段时间内提交结果，其他时间建议在训练集上构建模型：

训练集train.csv：位于（6-数据集）文件夹中，共41015行×40列，其中第一行是属性名称，剩余的行是每份保单的相关信息，最后一列为该保单生效期间客户是否提出了索赔申请（1-提出，0-未提出）。

测试集test.csv：[通过期末考试.exe 中获取](https://lab.uibe.edu.cn/show?filename=2024-2025-2-CMP245-Python%E4%B8%8E%E5%A4%A7%E6%95%B0%E6%8D%AE%E5%88%86%E6%9E%90-%E8%AF%BE%E7%A8%8B%E5%A4%A7%E4%BD%9C%E4%B8%9A%E9%80%9A%E7%9F%A5.md)，共17579行×39列，其中第一行是属性名称，剩余的行是每份保单的相关信息。

请预测测试集中每份保单中客户会提出保险索赔的概率，越接近1越有可能进行索赔）。

数据属性及含义如下：

|  |  |  |
| --- | --- | --- |
| 字段名称 | 数据类型 | 字段描述 |
| subscription\_length | float | 保单生效时长（年） |
| vehicle\_age | float | 车辆使用时长（年） |
| customer\_age | int | 投保人年龄（岁） |
| region\_code | string | 投保人所在地区代码 |
| region\_density | int | 区域人口密度（人/平方公里） |
| segment | string | 车辆细分市场级别（如 A、B1、C2 等） |
| model | string | 车型编号 |
| fuel\_type | string | 燃料类型（汽油/柴油/CNG等） |
| max\_torque | string | 最大扭矩及对应转速（格式“XXXNm@XXXXrpm”） |
| max\_power | string | 最大功率及对应转速（格式“XXXbhp@XXXXrpm”） |
| engine\_type | string | 发动机类型（如 V型、直列等） |
| airbags | int | 安全气囊数量 |
| is\_esc, is\_adjustable\_steering, is\_tpms, is\_parking\_sensors, is\_parking\_camera, is\_front\_fog\_lights, is\_rear\_window\_wiper, is\_rear\_window\_washer, is\_rear\_window\_defogger, is\_brake\_assist, is\_power\_door\_locks, is\_central\_locking, is\_power\_steering, is\_driver\_seat\_height\_adjustable, is\_day\_night\_rear\_view\_mirror, is\_ecw, is\_speed\_alert | string | 是否配备相应车辆功能（Yes/No） |
| rear\_brakes\_type | string | 后制动器类型（如鼓式/碟式） |
| displacement | float | 发动机排量（升） |
| cylinder | int | 气缸数量 |
| transmission\_type | string | 变速箱类型（手动/自动等） |
| steering\_type | string | 转向方式（如液压助力/电动助力） |
| turning\_radius | float | 转弯半径（米） |
| length | float | 车身长度（毫米） |
| width | float | 车身宽度（毫米） |
| gross\_weight | int | 车辆总质量（千克） |
| ncap\_rating | int | NCAP 安全碰撞测试评级（星级） |
| claim\_status | int | 索赔状态（1=已索赔，0=未索赔），模型预测目标 |

### 结果文件格式要求

你需要提交的结果文件为一个.csv格式文件，总计17578行×1列（不包含第一行的属性名称）。每一行为索赔的概率值。

注意：

**1 再次警告，请用自己的test.csv文件生成结果，否则会爆炸；**

**2 结果须为17578行，请不要添加列标题或增减行数；**

**3 实在不行，请运行下面代码。**

参考保存csv代码如下：

#注意 predictY是pandas DataFrame 类型

import pandas as pd

import numpy as np

predictY = pd.DataFrame(np.random.uniform(0,1,**17578**).reshape(**17578**,1)) # 产生随机数

predictY.to\_csv('Results\_1.csv', encoding = 'utf-8', index=False , header=False)

### 评价标准——AUC

详细介绍请见B开头网站：

https://baike.baidu.com/item/AUC/19282953

使用示例

http://scikit-learn.org/stable/modules/generated/sklearn.metrics.auc.html

import sklearn

Score\_AUC = sklearn.metrics.auc(y\_true, y\_predict)

例如：

>>> import numpy as np

>>> from sklearn import metrics

>>> y = np.array([1, 1, 2, 2])

>>> pred = np.array([0.1, 0.4, 0.35, 0.8])

>>> fpr, tpr, thresholds = metrics.roc\_curve(y, pred, pos\_label=2)

>>> metrics.auc(fpr, tpr)

### 参考代码及报告

见压缩包 5-参考案例-强力推荐 文件夹