

练习：泰坦尼克号乘客的命运

2023 年 9 月 18 日

作业：泰坦尼克号乘客的命运

1912 年 4 月 15 日，被广泛认为“永不沉没”的皇家邮轮泰坦尼克号在处女航中撞上冰山沉没。不幸的是，船上没有足够的救生艇，导致 2224 名乘客和船员中 1502 人死亡。

虽然在生存中有一些运气因素，但似乎有些群体的人比其他群体更有可能生存下来。

在这个挑战中，我们要求你建立一个预测模型来回答这个问题：“什么样的人更有可能存活？”

使用乘客数据 (如姓名、年龄、性别、社会经济阶层等)。

要求

在本次案例中，您将获得两个类似的数据集，其中包括姓名、年龄、性别、社会经济阶层等乘客信息。

数据集：1. [train.csv](#) 1. [test.csv](#) 1. [gender_submission.csv](#)

一个数据集是 `train.csv`，另一个数据集是 `test.csv`。`train.csv` 将包含机上乘客 (准确地说是 891 人) 的详细信息，更重要的是，它将揭示他们是否幸存，也被称为“真相”。

而 `test.csv` 数据集包含类似的信息，但没有披露每名乘客的“真实情况”。预测这些结果是你的工作。

使用您在 `train.csv` 数据中找到的模型，预测其他 418 名乘客是否幸存 (在 `test.csv` 中找到)。

属性	定义
<code>pclass</code>	舱位 (“1st” 为一等舱, “2nd” 为二等舱, “3rd” 为三等舱)
<code>survived</code>	是否生存 (1 为生存, 0 为未生存)
<code>name</code>	乘客姓名
<code>sex</code>	性别
<code>age</code>	年龄

属性	定义
sibsp	在船上的配偶和兄弟姐妹数量
parch	在船上的父母和子女数量
ticket	船票号码
fare	票价
cabin	房间号
embarked	登船地点

[]: