

附件 1

人工智能选拔赛题目

一、题目背景

1912 年 4 月，泰坦尼克号邮轮在首航中与冰山相撞后沉没，造成乘客和船员共计 1502 人死亡。本题要求参赛者利用提供的数据，分析数据特征，建立预测模型，预测乘客的幸存情况。

二、数据说明

在本次比赛中，参赛者将能够访问两个关键的数据集，这些数据集详细记录了乘客的各类信息，包括姓名、年龄、性别、社会经济阶层等。具体数据集说明如下：

`train.csv`：该数据集包含泰坦尼克号上乘客的详细信息，以及他们是否幸存的记录。这主要用于训练预测模型。

`test.csv`：该数据集包含另一组乘客的信息，但缺少“Survived”列，即这些乘客的幸存状态未知。参赛者需要使用训练的模型来预测这些乘客的幸存状态。

数据集字段说明：

PassengerId	乘客的唯一标识符
Survived	生存状态（0 = 没有生存，1 = 生存）
Pclass	船票等级
Name	乘客姓名
Sex	性别
Age	年龄
SibSp	同行的兄弟姐妹或配偶数量
Parch	同行的父母或孩子数量
Fare	乘客支付的船票费用
Ticket	船票编号
Cabin	船舱编号
Embarked	登船码头

四、评分标准

数据处理	数据清洗和编码的准确性和完整性。	10分
	数据预处理过程的清晰度和逻辑性。	10分
乘客特征分析	乘客特征分析的准确性。	15分
	乘客特征与生存率分析的相关性。	15分
模型训练与评估	模型选择的准确性。	10分
	模型构建的完整性。	10分
	性能评估的准确的与完整性。	10分
结果展示与报告撰写	结果展示的准确性、清晰度和直观性。	10分
	报告撰写的条理性和逻辑性，完整性与深度。	10分