# 题目3 预测幸存者

该数据集为kaggle的竞赛之一（<https://www.kaggle.com/c/titanic>），数据包括train.csv和test.csv，最后可以生成gender\_submission.csv格式的文件上传kaggle查看得分。评价标准为准确率Accuracy。

要求1：观测数据离群点（异常点），选择是否去掉，可以通过训练结果决定是否应该去掉离群点

要求2：特征数量较少，可以逐个进行相关性分析。

要求3：对某些特征可以进行适当变换，使其更符合正态分布，转换类型可以考虑对数变换，指数变换，幂变换等。

要求4：缺失值处理

要求4：特征工程，特征提取，从Name中提取有用信息

要求5：利用决策树或随机森林查看不同特征的权重

要求6：采用多种模型调参，得到最佳模型

要求7：试对多模型结果进行综合得到预测结果，和单模型结果进行比较。

备注：可以对数据做其他认为应该进行的处理，尽量取得较好结果。