**课后作业**

**参加过的项目：**

Kaggle---Titanic Meachine learning Competition

Kaggle数据挖掘竞赛中的初学者任务，利用机器学习算法对人员幸存概率的预测

**使用范围：**

二分类的预测问题

**时间：**

约一周

**成本：**

时间成本，以及相关知识的学习成本

**目标：**

使自己预测的结果在Kaggle LeaderBoard 上面的排名尽可能地高

同时使自己的数据处理能力，算法实践能力得到较大的提升

**参加项目的原因：**

是一次对机器学习算法的实践

开启自己的Kaggle实践的道路

在实际项目中可以使自己的数据处理能力得到极大的提升

**如何定义项目的成功：**

能否得出合理的预测的结果，以及在LeaderBoard 上面的预测率排名，由于该项目是在Kaggle上多年的经典项目，很多人应用的模型已经将预测结果刷的零误差，因此看来，取得前10%的排名已经算是不错，同时该项目作为Kaggle上的入门项目；以及在项目过程中是否能够提升自己的数据处理和分析能力，是更加重要的事情

**如何接触到这个项目：**

在机器学习的课程中了解到Kaggle是用于算法实践的极佳平台

**项目亮点：**

由于存在数据的缺失项，故使用了随机森林算法对缺失数据进行拟合

使用Logistics线性，OLS线性回归，随机森林回归模型三种模型进行综合预测

对最终得到的模型进行了交叉验证，是结果更为准确

**关于项目经理和发起人：**

由于本项目是个人练手项目，故无项目经理和发起人的设置

**可以改进的地方：**

正确率有很大的提升空间

可以去尝试更多的模型

特征工程做的不够多

**Kaggle比赛平台地址：**

[Kaggle: Your Home for Data Science](https://www.kaggle.com/)