

《机器学习导论》· 期末大作业 · 船运到达时间预测

参赛说明

- 1、受疫情影响，本学期没有期末考试，本次比赛为期末评分，各位需要谨慎、认真对待
- 2、**1-3**人组队，建议选出一名队长统一安排、组织
- 3、需要提供队伍命名，按照要求，队名里**不能出现学号、学校、院系**等有关太过明显的标识，命名由**字母与数字组成，不限大小写**

队伍名需要包含队员姓拼音，ETA，不限次序，长度不多余30个字符，一般不少于15字符：

如：ETAliuyuan2020winwinwin

注：队名及组队详情尽快确定并QQ提交给助教（刘驭壬、袁雷），一经提交，不能更改

提交内容

- 1、队伍积分榜实时排名
- 2、完整代码与报告，其中提交给助教的代码需要注释清楚、逻辑清楚；报告需要清晰完整，包括但不限于设计思路、设计流程图、测试结果、队员分工等详细内容，会提供报告模板

评分截止时间

2020年06月19日中午12：00

评分说明

待定

赛题内容

在企业全球化业务体系中，海运物流作为其最重要的一项支撑。其中，船运公司会和数据供应公司进行合作，对运输用的船通过GPS进行定位以监控船的位置；在运输管理的过程中，货物到达目的港的时间是非常重要的一项数据，那么需要通过船运的历史数据构建模型，对目的港到达时间进行预测，预测时间简称为ETA（estimated time of arrival），目的港到达时间预测为ARRIVAL_ETA。

本次大赛提供历史运单GPS数据、历史运单事件数据、港口坐标数据，预测货物运单的到达时间，对应“历史运单事件”数据中EVENT_CODE字段值为ARRIVAL AT PORT时EVENT_CONVOLUTION_DATE的时间值。

赛事说明详见（<https://competition.huaweicloud.com/information/1000037843/circumstance?track=107>）

官方对数据格式、比赛要求等都有详细说明，链接底端有数据集下载说明，可以下载查看

最终名次判定依据‘评分截止时间’初赛实时排名

赛题引导

经过一个学期的学习，想必同学们对机器学习的基础知识、一些基本的模型已经了如指掌。通过参加比赛我们希望同学们能够体验一个数据科学家工作的完整流程是什么样的。需要注意的是，对于比赛而言，**数据处理 (包括EDA Exploratory data analysis)、特征工程** 与模型选择、模型训练一样重要。而且在大多数情况下，针对某些特征明显的问题，所有人使用的模型可能都是类似的，最终决定成绩的很可能就是提取的特征好坏。这次的比赛试题和kaggle中举办的很多比赛类型相似，所以建议同学阅读一些Kaggle比赛入门的资料。在这里我们列举了一些推荐阅读材料，希望对同学们的参赛有所帮助。

1. Kaggle比赛入门介绍: [1](#) [2](#)
2. Kaggle上的各类问题的[Notebook教程](#) (可以选择经典问题如[Titanic](#), [Predict Boston housing price](#)了解一个项目的完整流程)
3. [数据科学家初学者教程](#)(有选择地阅读)
4. 书本外的常用模型: 如[XGBoost](#), [ARIMA](#)等,
5. [时间序列问题](#)
6. [Ensembling & Stacking models](#)
7. [Kaggle竞赛\(很全\)的获胜解答](#) (选择性阅读)

注意事项

报名截止时间为**2020年5月31日 17:00**

报名截止时间为**2020年5月31日 17:00**

报名截止时间为**2020年5月31日 17:00**

助教联系

刘驭壬、袁雷

最后，预祝同学们在比赛中取得好的成绩！