

刘志远

15970627035 | fyhvyhj@gmail.com
研一在读 | 求职：数据分析实习



教育经历

南昌航空大学	2017年09月 - 2021年06月
计算机科学与技术 本科	南昌
海南大学 (211)	2021年09月 - 2024年06月
电子信息 硕士	海口

奖项及成就

- 英语四六级
- 国家励志奖学金，校一等奖学金
- 第十届蓝桥杯全国软件和信息技术专业江西赛区 C/C++ 程序设计大学 B 组二等奖
- 2019 年“中国高校计算机大赛-团体程序设计天梯赛”江西省“华山论剑”组中荣获团队三等奖
- csdn 技术博主，原创文章发布近 80 篇，最高实时排名 4000 左右
- 人工智能安全领域论文，在投

项目经历

天猫用户重复购买率预测

- 为了量化复购率问题，从三个维度进行拆解。从数据集中提取用户，商家以及用户与商家关联的画像与行为信息，以此作为特征工程的宏观目标。
- 在模型优化迭代方面，除去一套机器学习“组合拳”的基本操作。根据正负样本极度不平衡的现状，采用了三个方案：1. SMOTE 上采样 2. Ensemble 集成模型 3. 神经网络模型。
- 该任务在天池平台开放，在 7,041 个团队中排名第 132，获得 0.6871 的 AUC 值。

上海出租车数据集空间分析与可视化

- 目标是尽可能提取出有价值的信息，以便商业优化与政府决策。
- 以时间为变量，生成每个时刻上海出租车位置的分布图，并将图片组合成 gif 动态图片。由于数据量巨大，在地图上显示基本覆盖全区域，无法判断聚焦中心。采用栅格化系统，生成热力动态图。
- 统计出租车的起点与终点，在地图上生成热门行程标注图。相关部门可利用此信息，优化交通管理与密集人群安全管理。

豆瓣影评分析及可视化

- 该项目分三个阶段：1. 爬取影评数据 2. 数据清洗及分析 3. 生成有价值的可视化图
- 难点：豆瓣的反爬机制。解决：1. request 请求时，伪造客户端浏览器信息 2. 每次爬取操作后，都暂歇几秒钟。

其他

- 论文美化小工具。
- 更多项目细节，详见项目集 <https://liuyishou.netlify.app>

专业技能

- 编程语言及数据库：Python，Mysql
- 算法模型：逻辑回归，KNN，SVM，决策树，随机森林，GBDT，XGBoost
- 统计方法：假设检验，AB 实验，费米估算
- 数分工具：Numpy，Pandas，Sklearn，Scipy，Pytorch
- 可视化工具：Matplotlib，Seaborn