

芦耀辉

13615524679|huberylu@foxmail.com



教育经历:

中央民族大学 金融硕士 2019.08-2022.07

- GPA 3.6|4.0 金融科技方向

太原理工大学 地质学学士 2014.09-2018.07

- GPA 3.4|4.0 辅修经济学第二学位
- 中国金融博物馆-蚂蚁金服金融知识挑战赛优胜奖
- SAC 发布证券研究报告业务资格

个人技能:

- 数据库:** 熟练使用 MySQL, hive sql。了解查询优化、事务并发, 索引, Hadoop, spark 等。
- 编程能力:** 熟悉 java, python。熟悉面向对象语法, 单例、工厂等设计模式。良好的注释习惯与代码品味, 熟悉各类数据结构: 顺序表、链表、二叉树、哈希表; 熟悉各类查找与排序算法: 折半查找, 快排、冒泡、归并排序等。
- DS 工具:** 精通 pandas、numpy、matplotlib、scikit-learn 等, 大量实际工作中代码实践, 并以个人和组队的形式参加过天池、kaggle 多个数据挖掘项目竞赛。在可视化方面, 熟练使用 seaborn、pyecharts, 工具: tableau、think-cell 等。
- 计量与建模:** 熟悉因果分析、注重模型解释性。良好的计量功底, 熟练使用 Eviews 进行相关分析、线性回归、假设检验、区间估计、异方差处理等, 做过的项目有: 中国石油消耗量影响因素与预测, 房价影响因素与回归分析等项目。
- 机器学习:** 熟悉分类、回归、聚类机器学习算法, 可推导公式并熟悉适用范围。包括: LR、决策树、SVM、ANN, k-means, PCA 等。个人熟悉集成学习领域研究, 从 bagging/boosting、到 xgboost、lightGBM。熟悉模型推导、超参数含义, 并在公司实习、竞赛中多次使用。
- 深度学习:** 熟悉 CNN 与 RNN, 熟悉 tensorflow.2.0。
- 爬虫与前端:** 熟悉 HTML\CSS\JS, 以及盒子模型、滑动显示等。熟悉爬虫框架 scrapy, 熟悉正则与 xpath 爬取各类文字、图片、视频资源, 熟悉 selenium 从 AJAX 中爬取内容。熟悉常见反反爬方式。

实习经历:

2021.01 - 至今

快手科技-数据分析部-内容与增长分析中心

数据挖掘/数据分析

- 支持快手二次元垂类、与新业务 ACFun 的 DA/DS 工作。团队负责个人主导的阶段项目有以下内容:

ACFUN 快手信息流广告投放项目

- 为了提高信息流广告投放的精准性, 从而提升新增设备的时长、留存指标, 需要从快手用户中找出对二次元题材高度感兴趣且有下载 A 站意愿的用户。将正样本定为从快手信息流广告跳转而下载 A 站, 并重度消费将要投放的 IP 素材的用户, 取出其在快手 APP 上有效播放过的二次元内容标签以及 APPList, 负样本定为对二次元垂类感兴趣, 但对投放素材反感(曝光次数>5 有效播放=0 次)的用户, 取出对应标签与 APPList。得到全部为 0-1 型变量, 选取不易过拟合的 lr, +FM 提升性能, 模型 AUC 为 0.84。base 为快手磁力引擎, A/B 结果: treat 首日消费时长+12%, 次留+3%。并进行 PSM 对比, 差异通过 MDE 检验。

新增 23 岁以下用户 Aha-Moment 探索

- 北极星指标由时长向留存转换, 协助拆解指标并计算关键行为阈值, 挑选 1 月新增 12-23 岁设备的留存与各类消费数据 (共计 150 万条), 分为登陆用户和为登陆用户两组数据集, 以是否留存作为被解释变量, 进行相关性检验、xgboost 特征筛选 (AUC=0.73,F1=0.97), 得出影响留存的主要用户行为因素, 并在此基础上编写程序计算 aha 极值, 反推阈值点, 以及大于/小于阈值的用户 7、14、30 留存差异。输出结果给产品侧, 协助进行用户激励体系与登陆引导设计。

UGC 稿件综合分模型

- 从业务需求出发, 为缩小人工审核优质稿件与 23 岁以下年轻用户偏爱内容的调性差异, 以该人群的稿件消费结果: ctr、ftr、wtr 等 8 个指标为特征, 以人均播放时间和完播率为被解释变量, 进行相关性分析和 RF 特征筛选后, 进行线性回归, 分别得到两个方程的线性回归系数, 并将回归系数进行加权平均作为综合分的各指标权重, 以此分来指导审核和运营人员调整筛选依据, 使得曝光给目标受众人群的优质稿件整体 ctr 上升 4%, 差异通过 MDE 检验。

2020.01 - 2020.06

滴滴出行-两轮车事业部-用户中心

数据分析

- 支持各类 SQL 查询需求。配合数据研发部门底层表更新, 编写并优化 SQL 模板。支持同团队各个项目的数据分析需求。

私占车召回模型

- 依据用户是否私占车辆数据, 提取车辆末次成单时间、末次使用距离、月卡剩余时间、超出运营区次数等特征, 使用逻辑回归, 决策树, xgboost 方法建立机器学习模型, 得出样本数据集规律, 模型 F1 值达到 0.89。