

周小越

手机：18948359839 · 邮箱：u3604403@connect.hku.hk



教育背景

香港大学, 数据科学, 在读硕士研究生	2022.9 - 至今
核心课程：高级数据库, 金融欺诈分析, 数据挖掘	
中山大学, 统计学, 理学学士, 辅修金融学	2018.9 - 2022.6
GPA：3.8/4, 核心课程：概率论、数理统计、数据库原理与应用、数据结构与算法、应用回归分析	
加州大学伯克利分校, 国际生交换项目	2020.1 - 2020.5
GPA：3.7/4, 核心课程：Data Structures、Concepts of Statistics	

技术能力

IT 技能: SQL、python、tableau、Trino 英语能力: 雅思 6.5 (6)、大学六级 599

实习经历

影石 insta360 Insta360, 数据分析师	2024.1-至今
<ul style="list-style-type: none">付费钻石分析：针对新上线的付费钻石功能, 分析其商业化情况。从付费渠道分布、用户转化率、付费套餐设计等多个方面进行拆解, 最终产出数据分析报告一份, tableau 看板一份。用户积分分析：针对积分商城活动, 用 trino (分布式 sql 引擎) 取数, 验收相关数据, 期间成功排查出四位异常用户。设计并搭建积分看板。埋点设计与验收：设计 APP 登录页面交互优化相关埋点, 成功交付一份埋点设计文档, 同时对开发好的埋点进行数据验收。	
华为技术有限公司 Huawei Technologies Co., Ltd, 数据工程师	2023.7-2023.9
<ul style="list-style-type: none">区域组织变更优化：针对区域组织变更中出现的版本更新时间窗小等问题, 和模块负责人沟通, 分析有关 hql 代码, 设计优化方案。最终产出一份总体变更优化方案, 实现 13 个模块变更与版本解耦。主数据集成监控：针对主数据集成不规范问题, 分析各模块数据集成链路, 识别数据源不明确、二次透传等集成不合规情况。最终产出主数据集成监控看板, 包含数据来源, 各模块合规率等信息。	

项目经历

网页弹窗效果分析 (AB test)	2023.12
<ul style="list-style-type: none">主要内容：分析网页弹窗对于课程订阅和付费人数是否有显著影响。将用户随机分成有无弹窗两组, 以课程订阅率、付费率为检验指标, 进行显著性分析。以界面浏览量等为不变指标, 进行健康性检验。实验效果：界面浏览量通过健康性检验, 说明实验不受无关变量影响。最终结果显示, 课程订阅率下降了 2.06%, 变化显著, 课程付费率下降了 0.4%, 不显著。	
联邦学习算法研究	2023.7-2023.11
<ul style="list-style-type: none">主要内容：为解决数据孤岛问题, 让参与方在不共享数据的基础上联合建模, 设计了以神经网络作为基模型, 模型转换 (permutation)、模型连接 (mode connectivity) 作为模型融合方法的联邦学习算法。算法构建：各个数据集上独立训练的神经网络基模型每隔 10 个 epoch, 通过将其他模型映射到同一个 loss basin 后再连接的方法进行参数更新, 重复训练融合直至各模型参数统一, 所得为最终模型。实验效果：以 CIFAR-10 为数据集, 随机分成四份并用 Resnet 18 训练初始模型, FedAvg 方法为 baseline。目前模型 accuracy 为 0.85, 比 baseline 高 0.13。	
保险欺诈分析	2022.11
<ul style="list-style-type: none">主要内容：为帮助保险公司判断投保人是否进行了保险欺诈, 使用 python 对一万多条数据, 33 个变量进行数据分析, 建立保险欺诈模型, 并最终完成一份分析报告数据清洗：在数据预处理部分, 排除了 5 个低关联性指标, 补充数据缺失值, 并将分类变量转为数字或哑变量。用 SMOTE 方法对处理好的数据进行重抽样模型搭建：在模型搭建部分, 用 logistic 二分类、随机森林、SVM 等六个经典分类模型进行模型训练, 通过 recall, AUC 等指标对比发现随机森林为最优模型。	

奖学金

中山大学优秀学生奖学金三等奖	2019 年 12 月
----------------	-------------