

庞玮敏

15057202758 | pangweiminlori@163.com
数据分析项目: <https://github.com/weiminpang-lori?tab=repositories>



教育经历

- 英国伯明翰大学 2018.09 - 2020.02
计算机科学 硕士
- 相关课程: 智能数据分析, 数据库, 机器学习, 评估方法与统计, 数据结构, 人工智能, 软件工程, 人机交互
- 英国莱斯特大学 2015.09 - 2018.06
会计与金融 本科
- 相关课程: 统计学, 计量应用经济, 金融衍生品定价, 金融市场产品, 公司治理

工作经历

- 德勤咨询(上海)有限公司 2020.10 - 至今
技术咨询分析师
- 【数据中台治理】基于数据生命周期(数据采集, 数据储存, 数据处理, 数据服务四个维度)分析数据中台(Dataphin, Dataworks等)面临的数据泄露, 访问机制和API可信认证等问题, 优化数据安全策略
 - 【云计算分析】基于部署在云计算产品(Azure, AWS, 阿里云等)上的系统架构和软件功能模块, 通过网络架构, 身份鉴别, 访问控制和入侵防范等维度分析虚拟机, 数据库, 云服务, 防火墙等组件, 设计制定并落实信息安全解决方案
- 爱奇艺 2019.11 - 2020.03
数据分析-算法策略实习生
- 【日常数据监测】运用SQL和Python处理热点频道feed流的核心效果数据(DAU, PV, UV和留存率等指标)生成数据分析报告, 并对短期异常波动进行归因分析
 - 【用户画像分析】通过漏斗, 用户分群等方法分析用户标签数据, 圈定个性化人群以及划分用户对短视频的兴趣偏好; 挖掘用户体验、产品推荐策略的优化点
 - 【A/B实验评估】基于排序模型的线上化, 通过A/B Test实验, 分析核心指标差异的因果关系, 评估推荐策略的最优版本, 最终相关推荐的良好导流效果, 提升了12.8%的用户点击率

项目经历

- 基于Python, TF-IDF算法分析Netflix的内容分布特点并搭建推荐系统 2020.10 - 2021.01
kaggle项目
- 【项目背景】Netflix是美国的流媒体视频平台, 个性化推荐系统是其核心的业务模式; 本项目旨在结合2000年~2020年12010条Netflix数据和爬取(web scraper爬虫插件)的TMDB数据, 探索性分析(EDA)Netflix的内容分布特点和用户行为特征, 并构建推荐系统, 模拟推荐用户喜欢的相似电影
 - 【推荐引擎搭建】基于地区, 流派, 导演, 演员和评论等数据维度分析内容分布和特点; 通过TF-IDF文本特征算法匹配内容相似度模拟推荐系统; 优化内容分发的策略
 - 【项目结果】结果显示Netflix上电影(68.9%)和电视节目(31.1%)大部分来自美国且侧重于动作和悬疑类型; 导演和演员的影响与作品受欢迎程度呈正相关关系; 推荐系统的精准度与用户行为和内容的相似度有关
- 基于MySQL, Tableau分析巴西Olist电商平台问题并制定策略 2020.07 - 2020.09
kaggle项目
- 【项目背景】Olist是巴西的电商交易平台, 本项目旨在对Olist的9个数据集(100,000条数据)分析巴西用户和商家的交易情况并针对异常问题制定优化策略
 - 【数据清洗】通过Python和MySQL数据库, 基于用户, 商家, 产品, 销售四个维度拆分清洗数据集; 分析用户复购率、转化率, 商家评分和订单延期交付率等各项指标
 - 【模型搭建】搭建RFM模型分类量化处理用户价值、随机森林模型挖掘商家评分相关特征; 运用Tableau可视化分析异常指标, 发现问题并制定合理的电商策略
 - 【项目成果】根据2.75%用户复购率, 70.87%用户好评率, Top10产品占64.57%销售额的明显头部效应等分析结果; 建议优化平台功能, 以提高用户的留存率; 追踪下单流程, 计算并合理提高转化率; 合理设定免息门槛和分期利息, 增强用户购买力
- 基于Python探索性分析PUBG游戏吃鸡的相关性因素 2020.04 - 2020.06
kaggle项目
- 【项目背景】《绝地求生》(PUBG)是一款战术竞技型大逃杀生存游戏, 本项目旨在通过探索性分析提高玩家游戏中吃鸡(获胜)概率的相关性策略
 - 【数据相关性挖掘】通过numpy,pandas预处理100万条海岛地图和沙漠地图的匿名玩家数据, 探索性分析落地安全区, 击杀武器, 代步工具和毒圈范围等相关性数据指标; 搭建相关系数矩阵和多维度特征结构分析多种数据维度的相关性
 - 【项目成果】最终优化的游戏吃鸡策略: 新手玩家降落在低危区域(海岸沿线以及桥头位置), 收集AKM、Kar98k等游击远战武器资源, 结合交通工具获取更多的游戏资源; 高手玩家降落在高频决赛圈区域(P港, 军事基地, 学校等), 采取近战模式掠夺游戏资源, 同时提高净击杀率至58%
- 基于D3.js可视化数据分析英国NHS医疗系统“床位堵塞”问题 2019.05 - 2019.10
毕业设计
- 【项目背景】英国NHS医疗系统的“床位堵塞”问题导致患者等待时间过长而错过最佳治疗时间, 本项目旨在构建数据可视化界面帮助用户(患者和医院决策者)作出有效决策来解决医院“床位堵塞”问题
 - 【仪表盘搭建】收集和清洗6000组医院患者的数据并划分成不同的数据维度; 构建仿真决策模型分析“床位堵塞”的潜在因素和用户需求; 通过D3.js和HTML等技术搭建多样化的数据可视化仪表盘, 并基于可用性测试三次迭代仪表盘
 - 【项目成果】最终实现20名用户(患者或医院决策者)与可视化界面进行交互, 帮助了60%以上患者快速选择合适的医院, 50%以上医院决策者及时发现不同科室的手术费用和患者满意度问题

核心技能

- 分析工具: 精通SQL(MySQL), Python(Numpy, Pandas, Matplotlib, Seaborn, Plotly), Jupyter Notebook, Tableau
- 模型算法: 熟悉回归(逻辑, 回归), 分类(决策树, 随机森林), K-means聚类分析等
- 其他工具: 精通Microsoft Office, Visio, Process-On, Xmind
- 英语能力: 听说读写全面, 可作为工作语言
- 兴趣爱好: 策划旅行攻略, 摄影剪辑Vlog, 健身(瑜伽, Boxing), 真人CS大赛