

唐珮雯

ptang0114@gmail.com | (+86)15996292630

专业技能

编程语言: Python (Pandas, scikit-learn, FastAPI), SQL, JavaScript (Next.js)

机器学习: 回归分析, 聚类, 分类, 时间序列分析, 多模态模型, 特征工程, 模型评估

MLOps: 云平台 (Google Cloud Platform, AWS), Docker & Docker Compose

全栈开发: RESTful API 设计, 前后端系统集成, 交互式数据看板开发 (Streamlit, Gradio)

教育背景

2017 - 2018 数据科学硕士, 哥伦比亚大学, 美国

2012 - 2016 心理学学士, 布林茅尔学院, 美国

工作经历

数据科学家 / 机器学习工程师, Zenni Optical

上海, 中国

(Zenni是北美第一的线上眼镜零售商)

2023.04 - 2024.09

【邮件营销 & CRM推荐系统】

- 目标: 开发数据驱动的个性化推荐系统, 替代低效的人工选品流程, 提高邮件营销效果
- 设计数据结构和ETL流程以整合Google Analytics埋点数据, 解构用户浏览点击链路, 分析用户点击促销邮件以及进行回购的频率
- 优化多个机器学习模型 (regression, random forests, XGBoost)提升SKU点击率和转化率, 构建A/B测试框架监控算法表现
- 使用BigQuery和Python拉取埋点数据计算关键指标, 实现“邮件点击率-推荐商品点击率-推荐商品转化率”的漏斗分析, 为市场团队提供实时反馈
- 创建自动化流程, 收集模型表现数据产出可视化图表和报告, 给业务提供数据驱动的营销决策
- 成果: 推荐商品覆盖率提升15倍(5%→70%); 转化率提升13.16%; 邮件营收增长31.17%; 人工选品时间减少85%

【库存预测和供应链优化】

- 目标: 开发精准的SKU级销量预测系统, 优化库存管理, 降低缺货率和库存成本
- 设计数据清洗流程, 处理SKU编码不一致、缺失值和异常值, 提高基础数据质量达95%以上
- 构建特征工程框架, 从原始销售数据中提取时间特征、季节性模式和价格弹性等40+维度特征
- 开发和对比多种预测模型 (Naive Forecast, Prophet, GBDT), 针对不同SKU特性(高/低销量、稳定/波动、季节性)选择最优模型
- 基于Google Cloud Run部署自动化模型服务, 设计监控和再训练流程, 确保模型持续适应市场变化, 维持长期预测准确性
- 开发Streamlit交互式数据看板, 实现预测结果可视化和异常分析, 支持业务团队直观分析实际销售与预测偏差, 主动调整库存决策
- 成果: 预测准确率提升41.6%(WMAPE从24.5%降至14.3%); 缺货率降低35%; 库存周转率提升28%

【生成式AI应用】

- 结合业务团队经验和指标, 设计了基于多模态模型(GPT-4V, Gemini, Claude)的商品描述自动生成流程
- 设计了质量评估框架和人工审核流程, 实现99%准确率和95%审核通过率, 将产品上线准备时间缩短65%

- 与产品团队紧密合作, 根据业务反馈持续优化提示词结构, 确保生成内容符合品牌风格和营销策略
- 开发基于GPT-4o的对话式推荐系统, 集成企业商品数据库, 通过自然语言交互实现个性化推荐, 提升加购转化率38%
- 设计结构化对话流程, 引导用户表达需求和偏好, 通过用户反馈不断调整推荐结果, 提高推荐产品匹配精度

数据科学家, 特赞 (上海) 信息科技有限公司

(特赞为创意方和设计师提供AI辅助工具和平台)

上海, 中国

2021.05 – 2022.06

【信息流内容分析与优化】

- 对信息流短视频全链路数据进行拆解, 使用Python提取帧级元数据标签, 识别关键内容特征, 并基于历史投放和行业效果数据进行分析
- 运用聚类算法(k-means, DBSCAN), 对机器打标效果进行后处理和优化, 提高人工二次审核的效率, 节约了50% 的人力成本
- 构建整合抖音、小红书等平台数据的ETL流程, 使用Python处理超过5000个账号的非结构化数据, 构建多维度标签体系 (内容风格、粉丝画像、历史带货表现)
- 开发“达人-商品”匹配推荐矩阵, 结合标签与效果指标 (互动量/播放量/搜索次数), 提升营销活动ROI 20%
- 创建标准化的数据建模流程(稳定性80%), 实现跨活动分析框架, 为不同品牌提供一致的数据洞察

【搜索体验优化】

- 挖掘和分析用户搜索历史数据, 定位异常值, 发现隐藏的问题和用户搜索习惯, 优化现有素材库和搜索字段的方法
- 利用自然语言处理工具, 对每条搜索路线进行拆解, 并运用聚类、语义相似度等算法, 推断用户搜索行为和动机进行
- 根据分析结果进行假设检验和A/B测试, 优化后提升搜索位54% (44 至23), 搜索链路缩短19%

商业分析师, 金融壹账通AI研究院

(金融壹账通隶属平安集团旗下, 提供各种AI产品的开发和服务)

纽约, 美国

2018.09 – 2021.04

- 爬取用户社交媒体数据, 运用机器学习模型搭建数据 workflow, 对关键数据缺失值进行填补预测, 丰富用户画像, 使模型准确率达到90%
- 使用SQL和Python从多源数据(社交媒体文本/ 互动数据/ 交易记录)中提取用户行为特征, 构建300+维度的用户画像模型
- 应用TF-IDF、情感分析和主题建模等NLP技术, 从社交媒体文本中挖掘用户偏好和行为模式
- 优化用户属性预测模型(LightGBM, RandomForest), 实现年龄/性别/关系网预测, 准确率92%
- 利用用户文本数据拟合模型, 以80%准确率预测用户性格特征, 利用五大性格模型来识别消费者习惯和动机, 为客户维系提供有效策略