

罗政

邮箱: zluoqaz327@outlook.com

联系方式: +86 15366217281



教育背景

纽约大学 (New York University), 生物统计, 硕士 2023.9 - 2025.5

- 相关课程: 生存分析应用 (A), 数据库与数据科学实训 (A), 文本数据分析 (A), 统计推断 (A), 缺失数据分析 (B+), 机器学习 (B), 流行病学 (B)

利物浦大学 (University of Liverpool), 应用数学, 本科 2019.9 - 2023.7

- 相关课程: 概率论与数理统计, 数学分析, 常微分方程, 线性代数, JAVA

实习经历

汇丰银行 (环球工商数据分析部), Data Analytics 实习生 2024.12 - 2025.1

- 负责客户体验提升项目, 基于 10,000 名客户数据, 运用 Excel 完成数据清洗与星座模型建模。
- 独立使用 Power BI 设计并开发了三层级可视化分析报表, 包含宏观 KPI 监控、客户经理 (RM) 个人绩效评估和潜在流失客户预警三大模块; 通过报表实现了对客户流失率 (20.37%)、平均满意度 (45.45%) 等核心指标的动态追踪, 并下钻分析至每位客户经理的具体服务表现, 定位其改进方向。
- 构建了基于负面反馈率的流失风险客户识别模型, 为业务部门提供了精准的挽留名单; 利用 Python 脚本实现报表生成与分发自动化, 有效支持了数据驱动的精细化运营。

微软 (商业分析平台), 数据分析实习生 2024.7 - 2024.8

- 负责巴西电商用户行为分析项目, 利用 KMeans 对用户进行价值分层, 并对各分层用户独立开展同期群 (Cohort) 分析, 识别出不同客群的关键留存节点与运营机会。
- 基于留存分析洞察, 为“高客单价”、“中等价值”等不同客群量身定制了精细化运营策略, 如“VIP 关怀”和“积分激励”, 以提升用户全生命周期价值。
- 独立构建 LightGBM 回归模型, 以预测用户未来 6 个月的消费总额 (LTV)。特征工程涵盖 RFM、品类多样性、支付方式等多个维度。对模型进行了深度诊断, 发现其对高价值活跃用户的预测误差 ($MAE \approx 150$) 远高于整体误差 ($MAE \approx 3$), 并主动提出采用两阶段模型 (Two-stage Model) 等一系列优化方案以提升预测精度。

学术经历

毕业论文 (硕士), 预测高血压风险的机器学习方法: 基于 NHANES 数据的实证研究 2024.5 - 2025.4

- 研究框架与数据处理: 基于美国 NHANES (2017-2020) 数据集 ($n=6,175$), 独立设计并执行全流程分析。使用 R 语言 (tidyverse, dplyr) 进行复杂数据清洗、缺失值处理及分类变量重编码与特征工程。
- 模型开发与优化: 为精准预测高血压风险, 构建并比较了 SVM、XGBoost 和随机森林三类机器学习模型。针对严重的类别不平衡问题, 综合应用了 SMOTE 过采样技术和加权损失函数进行优化。

- 模型性能：采用 5 折交叉验证进行超参数调优与评估。最终，经优化的 SVM 模型表现显著优于其他模型，在测试集上取得了近乎完美的性能：AUC-ROC 为 0.997，灵敏度 0.96，特异性 0.98。XGBoost 与随机森林的 AUC 分别为 0.728 和 0.730。
- 模型解释：采用先进的 SHAP 分析对最优模型（SVM & XGBoost）进行可解释性研究，不仅定量确定了年龄、BMI、酒精摄入为最重要预测因子，还发现了睡眠障碍、低收入水平的中等影响。通过 SHAP 依赖图，揭示了关键变量（如收缩压、舒张压）与高血压风险之间的非线性关系。
- 成果与产出：本研究系统性地比较了不同机器学习算法在复杂生物统计数据中的性能，为健康风险预测提供了实证经验。研究成果已撰写成硕士论文并通过纽约大学全球公共卫生学院答辩，所有分析代码与数据处理流程已在 GitHub 开源。

线性回归，课题项目，预测 2025 至 2026 年全加州房价中位值的课程分析项目

2024.4 - 2024.5

- 应用线性回归模型进行房价预测，并探索关键经济指标对房价的影响。
- 管理了项目的端到端生命周期，包括使用 R Studio 进行数据收集与清洗、特征工程、模型开发与验证；运用方差膨胀因子（VIF）诊断并有效缓解多重共线性对模型稳定性的影响；综合利用 Q-Q 图、残差分析以及调整后 R²（Adjusted R-Squared）等统计指标进行严谨的模型选择与迭代优化。
- 1. 成功构建并验证了一个能够有效预测目标时期内加州房价中位值的线性回归模型，测试集上模型的 R² 达到 0.82。
2. 通过系统地识别和处理多重共线性问题（如移除相关变量），显著提升了模型的稳定性和结果的可解释性。

课外活动

科技与创新协会社团，媒体部成员

2019.12 - 2021.8

- 描述：主要进行活动策划和宣传
- 职责：组织各种活动例如密室逃脱；制作并在公众号出版活动相关海报

Apple iOS 社团，媒体部成员

2020.12 - 2021.8

- 描述：主要进行活动策划和宣传
- 职责：制作并在公众号出版活动相关海报

大学生“互联网+”创新创业大赛，技术顾问

2021.10 - 2022.2

- 描述：设计提升生活品质的物品
- 职责：撰写大赛报告，提供应用数学专业相关的意见

专业技能

计算机方面：R, Java, Python, SAS, Stata, Power BI

工具方面：Office, Photoshop, Premiere Pro