

# 肖炬晔

手机号码: 13436826896 | 邮箱: jx2479@columbia.edu

## 教育背景

哥伦比亚大学 (Columbia University)

2021年9月 – 2023年2月 | 美国纽约

应用分析硕士 (M.S. in Applied Analytics)

课程: 数据管理与分析策略、应用分析框架与方法、数据叙事、组织环境中的应用分析、云计算、分析与变革领导、基于分析解决实际问题 (S&P Global资助的Capstone项目)

密歇根大学 (University of Michigan)

2018年8月 – 2020年12月 | 美国安娜堡

计算机科学学士 (B.S. in Computer Science) & 数据科学学士 (B.S. in Data Science)

课程: 编程与数据结构、数据结构与算法、计算机体系结构导论、计算机科学基础、面向企业家的移动应用开发、数据库管理系统、Web系统、机器学习导论、以人为中心的软件设计与开发、计算机视觉、数据挖掘、应用回归分析

## 工作经历

数据分析师 | C.O.D.E-Automotive

2023年9月 – 至今 | Camarillo, California

- 管理库存及销量数据, 优化库存策略, 提升供需匹配效率。
- 分析电商平台 (eBay、沃尔玛、亚马逊) 广告投放效果, 调整策略使广告ROI提升20%。
- 制定动态定价模型, 平衡利润与市场竞争力。

数据库管理专员 | China United Transport

2023年3月 – 至今 | Walnut, California

- 创建并维护综合货运数据库, 将数据检索速度提高30%。
- 优化数据库架构及流程, 提升了运输行程数据的完整性与实时性。

前端开发实习生 | 滴滴出行 (DiDi)

2019年6月 – 2019年8月 | 中国北京

- 开发网站性能监控工具, 运用React和Node.js构建高效模块。
- 分析性能数据, 提出优化方案, 使页面加载时间缩短15%。
- 参与滴滴的国际化进程, 了解全球网站部署实践和跨文化合作

## 项目经历

Climate Change Monitoring and Prediction using Satellite Data | S&P Global资助

- 收集来自NASA的MODIS陆地表面温度和NEX-GDDP近地表温度数据。构建包括热力图在内的数据可视化, 以阐明温度模式。
- 进行了全面的时间序列分析和分解, 揭示了MODIS和NEX模型之间的系统性变化。
- 开发并训练了一个ARIMA模型, 以MODIS数据为基础, 预测未来陆地表面温度。

Banking Customer Churn Prediction and Analysis

- 构建机器学习模型预测银行客户流失, 使用随机森林实现了86.35%准确率和较高的AUC得分。
- 通过One-Hot编码和标准化完成数据预处理, 识别了年龄、账户活跃度和地理位置等关键影响因素。
- 使用交叉验证评估模型 (随机森林、逻辑回归、KNN), 并通过网格搜索 (GridSearchCV) 优化模型性能。
- 帮助银行提前识别高流失风险客户, 制定针对性留存策略, 有效提高客户留存率和收益。

San Francisco Crime Analysis

- 运用Spark SQL技术, 确定旧金山最危险的三个地区, 并分析各类犯罪的时间分布, 明确高发时段和主要犯罪类型。
- 通过时间序列分解, 发现犯罪率在每年1月和10月达到高峰, 且2018年后呈下降趋势。
- 构建并运用网格搜索 (Grid Search) 确定ARIMA模型的最佳参数, 预测未来24个月旧金山的月均犯罪数量, 为警力资源规划提供支持。
- 提出优化巡逻安排、强化证据收集手段和推进社区安全项目的建议, 以降低犯罪率并提升解决率。

Amazon User Review Analysis and Topic Modeling

- 对亚马逊用户评论数据进行分析, 利用K-Means聚类和Latent Dirichlet Allocation (LDA) 模型挖掘潜在语义结构及用户洞察。
- 对评论数据进行预处理, 包括分词、去除停用词和词干提取; 通过TF-IDF特征提取识别代表用户情感的关键词。
- 确定用户关注的五大主题, 包括产品美观性、舒适性、质量、性价比及礼物潜力, 为营销和产品开发提供方向。
- 提供优化建议, 包括定价策略、设计改进及精准广告投放, 提升用户满意度与互动体验。

## 职业技能

- 编程语言: Python (Pandas, NumPy, Sklearn), Java, C++, SQL, Swift
- 数据工具: Tableau, Apache Spark, MongoDB, Metabase, Postgre SQL,
- 机器学习: 回归分析, 决策树, 聚类分析, 时间序列模型, Convolutional Neural Network, Recurrent Neural Network
- 其他: HTML, CSS, React, Figma, Node.JS,

## 个人评价

- 具备扎实的数据分析和统计建模能力, 精通Python、SQL等数据处理工具, 以及机器学习算法的应用。擅长从复杂数据中提取关键洞察, 为业务决策提供支持。拥有良好的逻辑思维与问题解决能力, 能够高效处理多任务需求并关注细节。熟悉电商、金融等行业的数据分析场景, 能够快速适应新环境并推动团队目标实现。注重团队协作与跨部门沟通, 具备数据驱动决策的强烈责任感和敏锐商业洞察力。