

# 周梓洋

电话: +1 604 603 6794

国内: 157 5781 7693

邮箱: zhouziyang0630@163.com

生日: 2005 年 6 月



## 教育背景

2023.09-2027.6

英属哥伦比亚大学(University of British Columbia)

统计学 (理学学士)

- 主修课程: 软件工程原理、数据分析(R, Python)、数据结构与算法(C, C++)、机器学习(Python)、网页开发 (JavaScript)、NLP (Python)、应用程序开发(Java)、随机过程分析
- 学业表现: 3.75/4.33 (CS GPA)

## 实践经历

2025 年 7 月-8 月

国泰海通证券 IPO 项目组实习

IPO 实习生

- 参与某企业北交所 IPO 项目, 协助完成行业研究、财务数据整理与上市申报材料分析。
- 使用 Excel / Python 对企业经营数据、财务指标及行业可比公司进行分析, 提高数据处理效率与准确性。
- 协助项目组进行招股书相关数据核验与信息披露整理, 支持上市材料准备流程。
- 参与市场与行业数据调研, 对企业收入结构、成长性及风险因素进行分析与总结。

2025 年 1 月-5 月

SpeedUp 教育机构

兼职数学教师

- 提供一对一辅导与学术支持, 帮助学生成绩平均超出班级 10%以上。
- 因教学条理清晰、擅长简化抽象数学概念而获得学生好评。
- 向本科生讲授高级微积分课程, 重点讲解多元积分与复杂问题解决技巧。

## 项目经历

2026 年 1 月 - 2026 年 4 月

NLP 文本分析与主题建模项目

独立工作者

- 基于 Python 与 Scikit-learn 完成 NLP 文本分析与主题建模, 对大规模非结构化文本进行语义分析与向量化处理。
- 使用 GloVe、TF-IDF 与 LDA 完成词向量建模、主题提取与关键词识别, 实现文本语义关系分析与 Topic Modeling。
- 实现 "king : queen" 等词语类比推理任务, 并结合 Matplotlib 完成主题分布与词向量关系可视化。

2026 年 1 月 - 2026 年 3 月

牛油果价格时间序列预测项目

独立工作者

- 基于 2015-2018 年多地区牛油果销售数据构建时间序列预测模型, 完成时序数据清洗、特征工程与价格趋势预测。
- 构建 Lag Features 与 month/week 等时间特征, 并通过时间序列切分方式避免数据泄露问题。
- 使用 RandomForestRegressor 完成建模与预测, 测试集  $R^2$  达到 0.787, 显著优于基线预测模型。

2026 年 1 月 - 2026 年 2 月

Airbnb 房价预测项目

独立工作者

- 基于纽约 Airbnb 公开房源数据 (4 万+记录) 构建房价预测模型, 对房型、地理位置、评论数量等特征完成数据清洗、EDA 与特征工程。
- 使用 Scikit-learn Pipeline 实现缺失值处理、标准化与 One-Hot Encoding, 并对异常值与长尾价格分布进行优化。
- 对比 线性回归、Gradient Boosting、Random Forest 等模型并结合 GridSearchCV 完成调优, 最终将 RMSE 从约 248 降低至 216。
- 基于 Random Forest 特征重要性分析识别核心定价因素, 提高模型可解释性与业务分析能力。

2024 年 1 月-2024 年 5 月

睡眠追踪 APP

独立工作者

- 技术栈:Java。
- 使用 Java 开发应用程序, 追踪与分析用户的睡眠模式。
- 与用户高效互动的睡眠数据分析 app, 生成个性化可视化优化报告, 包括睡眠规律、异常与改进空间等。

2024 年 9 月-现在

期货交易策略-中国玻璃期货

独立工作者

- 技术栈: Python, PineScript(TradingView), pandas, matplotlib。
- 实现: 用 pinescript 形成自动化的交易模型。
- 正在做什么: 加入 API 来用实盘进行检验, 多周期测试 防止过拟合。
- 在 13 年历史回测中实现 1685.76%的累计收益, 远超市场基准。

## 技术技能

- 编程语言:** Python, JavaScript, Java, C++, C, R, Vibe Coding
- 数据分析与可视化:** Pandas, NumPy, Matplotlib, ggplot, EDA (探索性数据分析)、数据清洗、统计分析、数据可视化
- 机器学习:** Scikit-learn, tidymodels、回归模型、分类模型、KNN、Random Forest、Gradient Boosting、XGBoost、特征工程、交叉验证、超参数调优、时间序列预测
- 无监督学习与 NLP:** K-Means、DBSCAN、Topic Modeling (LDA)、TF-IDF、Word Embedding、文本向量化、语义分析、文本聚类
- 开发工具:** Git, GitHub Actions、Streamlit、TradingView (Pine Script)
- 计算机基础:** 数据结构与算法、面向对象编程 (OOP)、多线程、数据库管理、软件开发基础