



王晋一

19816899984 | jinyiwang2022@163.com | 23岁 | 男 | 网站: [王晋一个人网站](#)

## 教育经历

南加利福尼亚大学(U. S. News 28)

美国加利福尼亚州洛杉矶 2023年08月 - 2025年05月

空间数据科学 硕士(3.95/4, 年级第一)

杭州电子科技大学

浙江杭州 2019年09月 - 2023年06月

信息管理与信息系统 本科(3.96/5)

荣誉/奖项: 获案例分析竞赛二等奖, 获电子商务竞赛二等奖, 获互联网+ 竞赛三等奖, 获挑战杯三等奖  
2022-2023校三等奖学金

## 工作经历

Lenovo美国

2024年9月-2024年12月

数据分析师

- AI模型微调与优化:** 主导BART和Qwen模型的微调, 利用Confluence、WikiHow和CNN数据提升模型泛化能力。构建高效训练流程, 使用PyTorch和Hugging Face优化训练效果, 并通过ROUGE和BLEU等指标评估模型质量, 成功优化Qwen tokenizer, 提高模型性能。优化GPU资源利用率, 将单次训练时间缩短25小时, 并开发可复用的AI微调框架, 获团队认可并应用于多个模型。
- 数据分析与质量评估:** 对销售、库存等业务数据进行清洗、处理和分析, 优化数据流, 提高数据准确性。构建销售转化率、客户留存率和库存周转率等关键指标体系, 并进行持续优化。利用SQL和Tableau开发自动化报表, 实时监控销售趋势和市场推广效果, 为管理层提供数据驱动决策支持。
- 销量预测与库存优化:** 设计并优化库存需求预测算法, 基于历史销量数据和市场趋势预测未来需求, 结合Python/R开发回归与分类模型, 与推动模型在库存管理和销售预测中的应用, 成功降低了库存积压20%, 提升了库存周转率。

杭州世平信息科技有限公司

2024年7月-2024年8月

数据分析师(社区安全隐患检测与响应系统)

- 安全隐患处理效率指标监控:** 开发自动化报表系统, 实时监控包括隐患识别率、响应时间和处理完成率等指标, 并对社区安全隐患处理中的指标进行异动分析, 发现某些区域的隐患处理完成率显著下降, 进一步分析发现是由于巡查人员短缺所致。提出相应的巡查人员调度优化方案后, 该区域的隐患处理完成率提升了12%。
- 社区隐患识别与分类预测:** 利用BERT-Chinese预训练模型对大量社区巡查报告和安全隐患记录进行文本分类和预测。模型优化后, 隐患识别准确率从85%提升至98%。
- 模型参数优化与Web化调试:** 通过Flask框架构建Web端调试工具, 实现了BERT模型参数动态调整与训练监控, 降低了训练时间35%, 加快了模型的优化迭代, 适应不同社区的安全隐患数据分类需求。

## 项目经历

膳食数据分析与预测系统(浙江大学)(技术支持)

2022年5月 - 2023年5月

【项目背景】: 通过机器学习技术对预包装食品的营养成分进行深度分析, 并预测用户的饮食偏好与营养需求, 提供更个性化的饮食建议, 提升用户健康管理的体验。

- 数据清洗与整理:** 通过使用Web抓取工具(BeautifulSoup和Scrapy), 自动化收集超过50,000条预包装食品的营养成分数据, 包括蛋白质、脂肪、过期日期等。对数据进行了清洗处理, 包括处理缺失值、去除异常数据以及归一化处理。
- 数据探索性分析EDA:** 使用Python的Pandas、Matplotlib和Seaborn库对数据进行了探索性分析, 绘制特征分布直方图、箱线图和热力图, 分析各营养成分的分布情况以及它们之间的相关性。重点分析了关键营养素(如蛋白质、脂肪、碳水化合物)对用户饮食选择的影响。
- 特征工程与模型选择:** 基于EDA筛选了与用户饮食偏好和健康管理最相关的特征, 并将其用于模型训练。使用回归分析和时间序列模型ARIMA、Prophet进行预测, 通过GridSearchCV调优超参数, 最终提升模型的预测精度。
- 模型评估与落地:** 使用均方误差RMSE、平均绝对误差MAE等指标对模型进行评估, 并通过交叉验证保证模型的泛化能力。模型预测的准确率提高了15%, 显著优化了用户的饮食管理体验。

智能化经济地理数据分析与预测系统(南加州大学)

2024年2月 - 2024年5月

【项目背景】: 通过机器学习技术分析2008-2023年间多个经济数据源(如住房价格、GDP、薪资水平), 从中探索经济因素与住房价格的关联性, 并对未来的房价趋势进行预测。

- 数据清洗与预处理:** 整合Zillow、Build On、BEA等数据源, 处理超过15万条记录, 涵盖住房价格、GDP和薪资水平。利用分布式计算和多线程技术加快数据清洗流程, 处理缺失值、异常值并进行归一化处理, 确保数据质量。
- 影响因素分析:** 通过Python Pandas、Matplotlib库进行数据探索性分析, 深入研究了各特征的分布及特征之间的相关性, 尤其是薪资水平、GDP与住房价格的关系。分析发现GDP与房价的线性相关性较弱, 而薪资水平与房价正相关性显著。同时基于回归模型得到各特征的重要性, 为理解住房价格的影响因素提供了数据支持。
- 时间序列预测:** 基于ARIMA模型对2008-2023年的经济数据进行了时间序列预测, 捕捉长期趋势和周期波动。对主要城市的未来房价进行了预测, 分析结果显示经济增长较快城市的房价将继续上升, 而经济放缓城市的房价将保持稳定或下降。

## 专业技能

- 分析技能:** Python, SQL, Excel, Tableau, Elasticsearch, Java, HTML, CSS5, JavaScript, Go, 小程序文档
- 分析方法:** 熟悉常用数据分析方法和模型: 漏斗分析、异动分析、相关分析、归因分析、RFM、AARRR、AB Test、假设检验等
- 语言能力:** 英语(托福103), 中文