

季嘉伟

电话:+86-18530931097 邮箱: jasonjjw@umich.edu



教育经历:

硕士 | 密歇根大学安娜堡分校 09/2024 – 07/2026

专业: 健康数据科学
• 核心课程: 大数据计算与分析 | 机器学习 | 广义线性模型应用 | 线性回归 | 统计推断 | 生存分析 | 信息可视化
本科 | 西交利物浦大学 (西安交通大学 & 英国利物浦大学合作办学) 09/2020 – 06/2024

专业: 数据科学与大数据技术 (毕业获一等荣誉学位, 专业前 15%)
• 核心课程: 人工智能与数据分析 | 算法设计与分析 | 数据库开发与设计 | 神经网络 | 数据可视化 | 概率统计 | 高性能计算 | 模式识别 | C++/R 编程 | 线性代数 | 数值方法 | 并行计算
• 语言能力: 雅思 7, GRE320

工作经历:

数据分析岗 (实习) | 大专家.com (上海明品医学数据科技有限公司), 上海, 中国 05/2025 - 08/2025

- 负责处理从医院机构和数商提供的医学真实世界数据, 参与数据库管理、数据清洗与结构化处理, 推进报告生成自动化与数据质量控制流程建设:
 - 使用 MySQL 操作 Doris 数据库, 导入导出真实世界数据, 实现多模块间数据快速流转。撰写筛选脚本, 基于量表字段关系提取并清洗下游分析所需要核心数据;
 - 根据公司业务, 对指定药品的真实世界数据做深入分析。开发报告脚本, 用 JSON 模板清洗数据、做标准化处理、画图表, 整理药品分布、适应症等信息, 实现结构化报告自动输出。目前已交付企业分析报告 100 余份;
 - 审查真实世界数据的记录, 标出缺失、冲突或格式错误的值。设计数据检查流程, 并用 Python 脚本自动完成审核, 使审核效率提高大约 20%。

数据系统与验证研究助理 | 江苏省农科院, 南京, 中国 05/2023 - 08/2023

- 基于大语言模型 (LLM) 和 Prompt 工程, 从农药说明书中抽取知识图谱三元组, 使用 Python 脚本完成清洗与标准化, 并将结果导入 Neo4j 图数据库:
 - 参与农药领域知识图谱的构建工作, 负责设计三元组结构与实体关系方案, 明确农药名称、防治对象、适用作物等核心要素, 支持后续信息抽取与结构化整理;
 - 基于大语言模型 (LLM) 与 Prompt 工程方法, 构建提取策略与提示模板, 从农药说明书中自动识别并抽取三元组信息;
 - 使用 Python 脚本对抽取结果进行清洗与标准化, 修复缺失字段、统一格式, 建立字段标注与人工审查流程;
 - 将结构化数据导入图数据库 (Neo4j), 支持多层级实体与关系的展示与查询。

数据工程师 (实习) | 新华三集团大数据中心 (H3C), 郑州, 中国 01/2023 - 04/2023

- 使用 Hive SQL 与 PySpark 搭建用户行为数据的处理流程与指标口径, 输出覆盖率、转化率等核心指标, 支持后续建模与实验分析, 提高数据驱动策略优化的效率:
 - 协助优化 Hive SQL 与 PySpark 数据处理脚本, 修复字段异常、映射错误和时间格式不一致问题, 保障数据流程稳定运行;
 - 抽样 10,000 条用户行为日志, 利用 Pandas 执行缺失值填补、字段标准化与变量分箱, 并用 Seaborn 绘制清洗前后的分布对比图, 提升数据可用性;
 - 编写 SQL 脚本关联多张用户行为相关表, 构建样本标签, 配合建模团队筛选关键特征变量; 随后通过建立逻辑回归和随机森林模型, 对数据进行二分类实验并复盘效果;
 - 完成分析报告与 Tableau 可视化看板, 总结数据缺失、变量分布变化与样本平衡情况, 帮助团队统一指标口径并指导策略迭代。

研究 & 学术项目

百度飞桨 | 导师: Prof. Hongbin Liu 06/2022 - 07/2022

- 设计并训练出一款模型, 用于识别交通标志并提取关键信息:
 - 实施图像预处理流程, 包括尺寸重塑、归一化与数据增强, 有效提升模型鲁棒性与识别效果;
 - 构建、训练并比较了使用 PaddleClas 框架的卷积神经网络 (CNN) 模型的性能; 引入了 dropout 层和 ReLU 激活函数来提高模型的泛化性能。在调整 (参数如学习率和批量大小) 后, 模型在测试集上达到了 95.74% 以上的准确率。

技能

编程语言 / 软件: Python, R, C++, MATLAB, MySQL, Oracle 等
库 / 工具: NumPy, Pandas, PowerBI, Sklearn, PyTorch, Prompt Engineering, Matplotlib, Seaborn 等