

王潇扬

✉ xywang191209@gmail.com · 📞 18513066510 · 🌐 WXYZ1209 · 🐙 wxys1209.github.io



🎓 教育背景

威斯康星大学麦迪逊分校

威斯康星, 美国

2022.09 – 2024.05

理学硕士 统计学与数据科学

华东师范大学

上海, 中国

2019.09 – 2023.06

理学硕士 统计学

🔧 工作经验

中国广视索福瑞媒介研究有限责任公司

北京, 中国

2024.07 - 2025.08

研究主管 体育部

1. 日常收视报告

- 负责足球五大联赛、欧足联、F1、ONE 冠军赛及 FIS 等赛事的收视报告制作, 设计自动化数据提取、清洗与分析程序, 有效缩短报告生成周期
- 设计并开发自动化程序和定制爬虫, 实现赛事日程与节目单的实时采集与更新, 大幅降低人工干预, 提高数据更新频率与质量

2. 体育数据库建设与数据平台优化

- 协同团队搭建并维护体育数据库, 开发交互式数据清洗平台, 将原始数据高效转化为结构化数据, 为后续深度数据分析和报告制作奠定基础

3. 综合语言模型

- 构建并优化服务于业务需求的语言模型, 实现文本分类、命名实体识别及信息检索的功能
- 通过微调大语言模型和应用机器学习方法, 完成了关键技术流程的整合, 构建出完整的系统架构

4. 交互式数据分析报告

- 开发并制作交互式数据分析报告, 设计标准化数据格式并整合多赛季数据
- 优化平台交互逻辑, 高效满足多样化数据分析与展示需求

5. DataEase 数据仪表盘

- 与团队共同设计并开发 DataEase 数据仪表盘, 整合电视和移动端数据, 支持多维度分析, 帮助客户以数据为驱动进行业务决策

Wisconsin Institute for Discovery

威斯康星, 美国

2023.07 – 2024.05

研究助理 Claudia Solís-Lemus 实验室

1. BHV 树空间上的统计检验研究

- 熟练掌握 Bowers-Hatcher-Vogtmann (BHV) 树空间理论, 以及该空间中统计量以及中心极限定律的应用
- 设计假设检验, 制定合适的检验统计量, 通过理论推导和模拟研究来验证统计量的渐进性质, 确保分析结果的准确性和可靠性

2. 病毒蛋白质序列相互作用的动态可视化研究

- 对上传的病毒蛋白质序列数据进行清洗和处理, 确保数据质量, 并将其格式化为用于进一步分析的结构化数据集
- 开发动态可视化工具来展现病毒蛋白质序列间的相互关系, 提供直观的数据表征, 辅助生物学家进行实验设计和决策过程

3. 威斯康星州马铃薯种薯认证计划线上数据分析平台

- 利用 Shiny 构建高效的在线数据分析平台, 集成数据验证、清洗、可视化、统计检验和预测功能, 提供全面的数据处理和分析
- 为威斯康星州马铃薯种薯认证计划开展多元化的数据分析, 深入研究马铃薯疾病流行趋势, 为疾病管理和防控提供了数据支持

上海数因信科智能科技有限公司

上海, 中国

2022.03 – 2022.08

助理数据分析师 生物信息部门

1. 成功执行差异表达基因分析和批次效应校正, 以评估不同病人状态对研究结果的影响

2. 运用逻辑回归等统计方法深入分析生物医学数据, 为疾病研究提供精确的数据洞察

3. 深入研究降采样技术的相关文献, 并成功复现了研究, 以评估数据量减少对差异表达基因分析的影响

华西证券股份有限公司

北京, 中国

2021.07 – 2022.03

助理分析师

1. 积极参与科创板与 A 股主板项目的立项报告制作, 负责分析和梳理关键项目信息, 确保报告的完整性和准确性

2. 协助分析师完成两份深入的企业研究报告

🔧 技能与特长

- 编程语言: R、Python、C++、Shell/Bash
- 技术工具: TensorFlow、PyTorch、Transformers、Pandas、Numpy、Scikit-learn、Shiny、Git、LLMs、 \LaTeX 、MySQL、Microsoft Office
- 其他: 足球、羽毛球、滑雪、量化投资研究、运用知识解决实际问题