

许智彤

13358070886 | zhitongx@163.com | 北京



个人总结

- 统计学专业大四在读，熟悉使用Word、PowerPoint、Excel、SPSS等软件，对文件管理非常熟悉
- 熟练使用R语言、python和SQL，机器学习方面的算法应用熟练，参加过kaggle竞赛排名达到前5%，跟学院老师做过知识图谱方面的科研项目，有Neo4j图形数据库开发实战经验，数据爬取、清理、可视化等数据分析及建模方面的经验丰富
- 熟悉HTML、CSS、JS、Django框架、d3可视化，能独立完成功能页面的设计与代码编写
- 主修多元统计分析、时间序列分析、抽样技术、算法设计与分析、机器学习、计量经济学、网络群体与市场等课程
- 拥有良好的沟通和协调能力，善于应变，能够快速适应新环境

项目经历

基于WOS统计学期刊数据的统计学科关键词知识图谱可视化

2019年10月 - 至今

项目组核心成员

- 基于原始数据与得到的adjacency matrix以及国内外关键词连通子图数据，针对学校地域信息，研究发文主题的侧重情况，实现可视化
- 将项目组的研究成果部署到网站，同时开发了其衍生产品，统计学知识自我测评系统，能够预测年轻学者与国际一流学者在进入面对面合作之前专业倾向意愿、知识储备的重组性和对协作机会的把握能力，从而有助于对研究者的未来的专业合作潜力进行预判和规划，应用推荐算法，提供相应的选课建议

产品责任延伸制度背景下电子废弃物资源化的消费者驱动模型的构建

2019年09月 - 2020年06月

项目核心成员

- 研究电子废弃物资源化相关理论，与从业者进行深度访谈，确定研究方法并制定研究计划
- 设计问卷并通过预调研确认问卷合理性，在北京、厦门、广东贵屿三地进行实地调研，收集数据对市场管理部门进行采访
- 使用联合分析法，构建消费者回收模式选择模型，探究消费者回收偏好，并对各种回归模式的市场占有量进行预测，将最终结果汇总成研究报告

基于评价数据的亚马逊商城部分产品竞品分析

2020年03月 - 2020年03月

小组长

- 从评论基本信息、评论与产品关联度、评论被点赞情况和参评用户的行为等多个维度，筛选出更有价值的评论
- 作出好评率、销量增长速度等一系列变化曲线，识别并讨论可能的基于时间的度量模式和模式，其反映出一个产品的声誉情况
- 根据评论基本信息，结合评论情感倾向和情绪强度等文本分析结果，进行逻辑回归，得到初步结论
- 使用LDA主题分析模型得到各类产品评论集中反馈的方面，构建新的变量深入刻画产品，使用决策树，得到进一步的结论
- 使用格兰杰因果检验分析特定的打分情况是否会引起更多的评论，构建评价矩阵，分析特定描述与打分水平的关联性

纽约出租车车程用时预测

2019年07月 - 2020年08月

项目核心成员

- 补充外部数据和预处理后进行特征工程，得到决策树、KNN、随机森林、神经网络等模型的结果，其中随机森林的效果最好，排名为前41.6%，作为baseline
- 根据XGBoost的重要性结果进行变量的筛选，并调参，该模型排名为前13.9%，后使用LightGBM模型并调参，模型排名提高到前4.9%，最终采用stacking，第一层模型为LightGBM和XGBoost，第二层模型为线性回归，调参后最终模型排名为前4.2%

教育经历

中国人民大学

2017年09月 - 2021年06月

统计学 本科

其他

- 语言：英语（CET-6）