

肖琦伟

+86 15899013261 | kiwixiaowork@163.com
应届毕业生

教育经历

加州大学尔湾分校 UCI

2019年09月 - 2023年09月

计算机科学 本科 Donald Bren School of Information and Computer Sciences

尔湾

GPA: 3.7/4.0

院长嘉许名单 (Dean's Honor List) 2019-2023

相关课程: 数据结构, 数据库, 机器学习, 信息检索(Information Retrieval), 算法, 计算机网络, 操作系统原理, 软件测试与分析, 自然语言处理, 计算机视觉

实习经历

加州大学尔湾分校附属医院 (UC Irvine Health)

2022年06月 - 2022年10月

数据分析实习生

奥兰治

- 使用Excel VBA和Formula为9个项目编写超过1000行代码, 设计工作分配流程, 自动预处理数据, 大幅提高了工作效率并规范了数据格式。
- 调整并优化了员工管理系统和患者流系统, 提高了数据的可读性, 使系统的运行检索效率提高了31%, 降低了后期数据分析的难度。
- 分析比较了大量的数据, 例如劳务派遣收据, 前台电话记录等, 辅助管理层做出对医院运行有利的决策。
- 与其他部门的同事交流合作, 共同维护跨部门的患者系统。

中国建设银行股份有限公司

2020年12月 - 2021年02月

财务顾问助理

库尔勒

- 参与分析潜在客户并根据客户资料提供适当的财务建议, 评估不同客户群体以协助财务顾问进行更深层次的财务咨询。
- 使用 Excel 整理客户数据, 并协助财务顾问将任务分配给银行出纳员。
- 与新客户对接, 协助进行投资风险测试, 并就符合其背景的金融产品提供初步建议。
- 协助大堂经理接待处理对公及对私业务的引导办理。

项目经历

开发类似ElasticSearch的新型搜索引擎

2021年10月 - 2022年01月

项目组长

尔湾

技术栈: Python, Porter2Stemmer, nltk, BeautifulSoup, re, numpy

这是一种新的搜索引擎, 能够在苛刻的操作限制下处理数以万计的文档或网页, 并确保查询响应时间小于 150 毫秒。

- 编写了超过 950 行代码, 通过标记解析 (tokenizing) 和词干提取 (stemming) 来使用爬虫爬取和处理来自网站的信息, 并对其进行索引以大幅度减少搜索时间。
- 设计了个人搜索引擎, 并了解了其整体结构和优化方法, 对新型搜索方法有了更细致的了解。
- 组织了一个四人项目小组进行项目开发, 统筹规划了项目的整体架构, 并制定了模块化目标和项目进度表, 使项目按时高质量的上线。

使用机器学习预测灾害中的幸存者

2022年03月 - 2022年06月

项目组长

尔湾

技术栈: Python, numpy, matplotlib.pyplot, sklearn

- 大量清洗并预处理数据, 以选择可能对分析乘客有用的特征, 使数据集更加的干净与高效。
- 训练和调整 K-Nearest Neighbors, Linear model, Neural Networks, 以及Random Forest, 预测哪种类型的乘客更有可能在灾难中幸存下来, 类型包括性别, 年龄和乘客仓位等级等等。
- 所有预测模型的准确率都在82%以上。

通用聊天机器人 (Chatbot)

2022年09月 - 2022年12月

项目组长

尔湾

技术栈: Python, Pytorch, nltk, numpy, csv, json, Transformer T5, RNN

- 收集整理"Cornell movie dialogs corpus"和"Reddit comments"两个数据集, 分析数据的内容和结构提取需要的部分, 最后使用nltk对原始数据进行预处理。
- 运用pytorch搭建RNN模型并使用处理过的数据集进行训练, 再使用相同数据集微调T5模型, 并进行可用性测试。
- 设计用户研究 (User studies) 评估准则, 随机抽样超过100人进行用户研究, 对比两种不同大小和技术路线的模型, 评估两种模型在不同使用场景的效果。最终微调后的T5模型的回答更流利易懂。

技能/证书及其他

- 技能:** Python, C++, C, MySQL, MIPS, VBA, Git, Linux, PyTorch
- 语言:** 普通话, 英语
- 兴趣爱好:** 公路车骑行, 摄影, 旅游, 健身, 吉他, 铁人三项