萧晋熙 (Charles)

邮箱 me@chunheisiu.com | 微信 chunheisiu | 手机 +852 6391 5470 / +86 170 9892 5470

教育经历

数据科学硕士 美国加州旧金山
美国旧金山大学 2019 年 7 月 – 2020 年 6 月

- GPA：3.8/4.0
- 入学奖学金，2019 秋季至 2020 春季

数据科学学士，副修设计 美国加州旧金山
美国旧金山大学 2014 年 8 月 – 2018 年 5 月

- GPA：3.64/4.0（优秀毕业生）
- 副校长奖学金，2014 – 2018
- 院长嘉许名单，2014 秋季至 2018 春季

工作经历

国际商业机器 (IBM) 中国香港
初級数据科学家 2020 年 11 月 – 现在

- 设计及开发超過 10 个用于 Apache Airflow 工作流程的 Python 数据处理脚本，其中包括從 XML 和 JSON 转换到 CSV 格式的转换脚本，系统化云端数据库的数据摄取及导入流程
- 开发及整合超過 20 个 Informatica ETL 工作流程及映射，应用在云端信息仓库中处理多层面的商业数据，方便其他部门的人员查询 OLAP 数据作业务分析及优化

香港应用科技研究院 (ASTRI) 中国香港
软件开发工程师 2018 年 8 月 – 2019 年 6 月

- 设计及开发智慧广告管理平台项目中用户管理的多个后端 REST API
- 把 DevOps 理念如 Jenkins 及 GitLab CI/CD 等应用于团队的开发环境，减少用于构建及部署 Docker 镜像的时间
- 以 Istio 收集用于开发环境的 Kubernetes 服务网格中的各种测量数据并以 Kibana 仪表盘呈现
- 参与用于开发及客户演示的系统及服务器维护并向项目经理提供采购建议

实习经历

Valimail 美国加州旧金山
数据科学实习生 2019 年 10 月 – 2020 年 6 月

- 通过机器学习链路，识别分类超过 10 万个由 Valimail Defend 系统收集的未识别域名
- 链路通过评定域名的各种参数， 评估该网站是否属于如房地产等低风险类别，最终采用精确率达 95%、基于自然语言处理（NLP）的梯度提升模型（GBM），为公司减少对人工标签的依赖
- 设计及开发链路中多个基于 Docker 及 Flask 的 RPC API
- 开发多个爬虫以收集网站信息作机器学习模型训练数据

高威电信 中国香港
软件开发暑期实习生 2016 年 6 月 – 2016 年 8 月

- 测试公司内部多个对接思科统一通信应用（Cisco UC）的软件
- 开发及展示一个基于思科 Jabber 网页开发工具包的演示应用

项目经历

- **HireReady：** 担任此项目的后端开发工程师。此项目是一个通过机器学习算法分析求职网站的职位信息，为数据科学家求职者提供量身定做的模拟面试网站。此项目在产品分析课堂中被评为优秀项目之一并在旧金山风险投资家路演中获得正面评价。
- **NBA 本周球星奖及薪酬分析：** 担任此项目的数据分析员。此项目使用了多项关于 NBA 球星的数据，以逻辑回归预测该球星在本季度获得本周球星奖的机会，以及通过线性回归预测他本季度的薪酬
- **USF MEDA 实习:** 担任此项目的数据分析员及网站开发。此项目与旧金山教会区经济发展委员会（MEDA SF）合作，基于由委员会提供的各项会员数据，分析上学距离、家庭经济状况以及学区派位的关系，并以可视化形式展示

技术及工具

- 编程语言：Java, Python, R, JavaScript, SQL, HTML, CSS
- 操作系统：Windows, macOS, Ubuntu, CentOS
- 工具及平台：Jupyter Notebook, Apache Spark, Docker, Kubernetes, Jenkins, Elasticsearch, Kibana, Redis, StreamSets Data Collector, Apache Kafka, Apache Airflow, Informatica, RStudio