

李逸凡

18811019056 | liyifan.steven616@gmail.com

教育经历

宾夕法尼亚大学	2021 年 9 月 - 2023 年 5 月
工学硕士：数据科学，工程与应用科学学院	美国 - 费城
• A+ 课程：Algorithms & Computation, Machine Learning, Deep Learning, Artificial Intelligence, Computational Linguistics	
清华大学	2017 年 9 月 - 2021 年 6 月
管理学学士：信息管理与信息系统，经济与管理学院	中国 - 北京
辅修学位：数据科学与技术，软件学院	
• 平均分：3.84/4.0	
• 获 2019 年国家奖学金、2018 & 2020 年学业优秀奖学金等多项荣誉	
明尼苏达大学	2019 年 8 月 - 2019 年 12 月
交换生：卡尔森管理学院	美国 - 明尼阿波里斯
• 平均分：3.898/4.0	
• 选修课程：Big Data Engineering and Architecture (研究生课程，在课程大数据分析项目中获得班级第一), Interactive Data Visualization for Business Analytics, Analysis and Modeling of Business Systems	

工作经历

亚马逊	2024 年 3 月 - 至今
软件开发工程师	美国 - 西雅图
• 从零设计并搭建团队端到端数据流水线与机器学习平台，实现每天约 70GB 数据处理能力，用于扫描亚马逊内部数据库与 API 并进行基于元数据的分类；支撑下游针对高数据安全严重性与个人数据保护策略的执行	
• 架构并实现一套 AWS 原生流水线：使用 Step Functions 进行调度, AWS Glue (Scala) 做 ETL, SageMaker 覆盖模型全生命周期 (训练、评估、部署、推理), Amazon Bedrock 用于大模型推理；通过 SQS 与 Lambda 实现可扩展的 Bedrock 作业调度，管理并发、重试与运行可靠性	
亚马逊	2023 年 7 月 - 2024 年 3 月
软件开发工程师	美国 - 西雅图
• 使用 Scala-Spark 维护团队的 AWS Glue 数据流水线，保障系统平稳运行与数据一致性；设计并实现流水线中的新“员工消歧 (Employee Disambiguation)”组件，使转化率指标准确度提升 7%	
• 开发并维护网站门户：后端采用 AWS Lambda 与 DynamoDB, 前端使用 React；设计并实现一个“边界框 (bounding box) 绘制工具”的后端与前端组件，使用户能够定义关注区域并获取定制化指标	
微软亚洲研究院 (MSRA)	2020 年 8 月 - 2021 年 3 月
软件分析组，实习研究员	中国 - 北京
• 以第三作者身份在 SIGKDD 2021 发表论文 <i>HALO: Hierarchy-aware Fault Localization for Cloud Systems</i>	
• 与团队成员共同设计并实现“HALO”——用于云服务的大规模故障诊断算法；复现两项相关工作，并使用 Python 对准确率与运行时效率进行对比实验	
• 开发 Flask API 服务与 Kusto 插件，便于其他内部团队使用我们的新算法；根据 Office 365 与 Azure 团队成员反馈，该算法显著缩短了工程师排查异常的时间	
• 实习结束时获得明日之星卓越奖	
字节跳动	2020 年 1 月 - 2020 年 8 月
大数据研发 / 数据科学实习生	中国 - 北京
• 在公司广告数据仓库的建设与治理中应用 Hive SQL 与 Spark	
• 维护两张与 TikTok 广告产品“Dou+”相关的数据集市 (data mart) 级别表的表结构与数据流水线；实现新增需求，例如增加与优惠券、购物车、直播推广相关的新指标	
• 优化公司线上实验平台的 A/B 测试异常检测机制；调研并落地一种基于时间序列数据挖掘的新算法，使告警准确率提升 30%	
大陆汽车系统（上海）有限公司	2019 年 7 月 - 2019 年 8 月
实习生，工业工程部	中国 - 上海
• 从 Oracle 数据库中获取数据，使用 MicroStrategy 商业智能工具和 KNIME 数据分析平台对数据进行挖掘。分析并预测一次合格率、物流发货量等重要生产指标，并绘制交互式的可视化图表提供给相关决策者	

- 深度参与公司机器人流程自动化 (RPA) 战略项目。快速学习并独立使用 UiPath RPA 软件开发了公司试点的三个自动化流程项目。主持了三场内部研讨会，并完成了一份 110 余页的教程文档，推动 RPA 项目在整个公司范围内的实施与应用

武岳峰资本

实习生，投资部

2018 年 7 月 - 2018 年 8 月

中国 - 上海

- 独立完成 38 页的自动驾驶汽车传感器所在行业的研究报告文档，并将重点概括为 22 页 PPT（另附 8 页补充资料）进行汇报，为团队调研一系列同一行业的投资标的 - 车载毫米波雷达芯片提供支持
- 参与了对 2 个项目 CEO 的访谈调研，独立完成前期的问题准备，记录并整理访谈内容，综合前期研究结论给出初步的分析意见
- 完成各类报告的中英文翻译，根据调查材料汇总制作两份投资项目研究摘要 PPT，并完成会议记录

研究经历

研究助理

导师：宾夕法尼亚大学沃顿商学院 Prof. Ryan Dew

2021 年 9 月 - 2023 年 5 月

美国 - 宾夕法尼亚

- 为数据驱动的营销研究构建偏好测量 (preference measurement) Web 应用；对遗留的 Flask 与 React 代码进行大量调试、重构与优化，显著提升代码健壮性与可读性
- 协助维护多个项目的数据 ETL 流水线，使用 Python、MongoDB 与 AWS S3

Carlson Leader 时间使用情况分析

研究助理，导师：明尼苏达大学 Prof. Le Zhou

2019 年 9 月 - 2019 年 12 月

美国 - 明尼阿波里斯

- 使用 Python 提取 500 余份 Excel 文件中的数据，合并为一份 CSV 并进行了数据清洗。帮助完成了描述性统计、相关性和回归等初步数据分析
- 使用 Vue.js、Flask 和 MongoDB 搭建了用于替代 Excel 进行实验数据收集的网站，提供用户提交数据以及反馈实验结果的功能。通过用户提交时的验证机制减少了无效数据，同时使界面美观程度显著提升

虚拟经济中价格指数的构建及货币传导机制研究

2018 年 10 月 - 2019 年 6 月

研究导师：罗文澜教授，刘庆教授

中国 - 北京

- 与腾讯互联网创新技术联合实验室合作立项，使用 Python 对 FIFA Online 3 游戏中玩家间交易，以及玩家与游戏公司交易的逐笔数据进行挖掘和处理，建立计量模型
- 分析游戏币严重通货膨胀，并造成游戏玩家严重流失以至难以继续运营的原因，为 FIFA Online 4 的交易平台、货币机制、游戏币奖励等机制设计提出建议

其他项目经历

安永 NextWave 数据科学挑战赛

2019 年 4 月 - 2019 年 5 月

队长，地区冠军

中国 - 上海

- 带领团队在全球 2700 多名选手中获得中国地区决赛冠军
- 使用机器学习技术完成了居民活动路径终点的预测任务。为队伍贡献了 80% 的特征工程思路，使用 pandas 和 scikit-learn 独立完成了工作流框架的搭建，并与队友共同完成模型选择和参数调优
- 组织了一份清晰的展示报告，与队友一同阐述了整个探究过程的思路、对结果进行的分析，并提出实际应用的可能，给评委留下深刻印象并最终帮助我们获得冠军

CNKI-2-BibTeX

2019 年 2 月 - 至今

开源项目作者

- 构建开源库 CNKI-2-BibTeX，用于将中文学术数据库 CNKI 的文献信息转换为可用于 BibTeX 的格式
- 该项目帮助了许多中国学者开展科研工作，并获得 150+ GitHub Stars 与 10.5 万次下载

CYDP 中国青年发展项目

2019 年 1 月 - 2019 年 2 月

队长

美国 - 纽约

- 参加了由哥伦比亚大学教授和企业管理者讲演的课程，包括行为经济学、金融风险管理、价值投资、设计思维、市场营销、国际组织等
- 拜访了纽约的部分企业，包括摩根士丹利和 IBM 等
- 参与项目创业计划比赛，在队伍中担任队长，统筹任务规划，并使用 LATEX 完成报告的整合与排版，独立运用 Hexo 搭建宣传网站。最终带领其余 8 名队员，在 7 支队伍中取得第二名

校跆拳道代表队

2018 年 5 月 - 2019 年 6 月

队长

中国 - 北京

- 负责代表队日常训练与招新活动，统筹代表队在校庆等活动中表演节目的策划与排练
- 个人最佳成绩：首都高等学校跆拳道精英赛，个人 54-kg 级男子竞技第五名；团队最佳成绩：团体总分亚军

- 负责第八届“今经乐道”经济热点分析大赛（校级最高级别赛事）的筹备与策划，主赛场吸引到全国高校 200 多支队伍参赛，新生赛赛场规模比去年扩大了一倍
- 撰写参赛手册中最重要的“大赛主题与选题指导”部分，编写 Python 脚本高效完成近 200 份初赛报告的分组、打包、统计评委发回的打分表 Excel 文件等高度重复性任务

其他信息

- 技能：Python, Scala, Kotlin, Spark, SQL, Hive, L^AT_EX, C, Java, TypeScript, Node.js, AWS, 数据科学，机器学习