

硕士 | 26 岁 | 男 | 在职 看机会

电子邮箱: e190049@e.ntu.edu.sg

目标/兴趣岗位: 数据分析 | 数据挖掘 | 数据科学

偏好地区: 南京 苏州 无锡 上海 杭州

林先生

教育经历

新加坡南洋理工大学

硕士 全日制 / MSc. Communications Engineering

GPA: 4.56/5.00

2019.07 - 2020.06

东南大学

本科 全日制 / 测控技术与仪器

加权分: 86.39/100

2015.09 - 2019.06

技能经验

具有工程思维，工科学历背景和 IT 工作经验。INTJ，倾向于研究、分析和创造性工作。熟悉常见的数据挖掘模型及算法。

使用 Python/SQL 及相应科学计算包对数据进行预处理和 EDA，第三方机器学习库进行机器学习建模和搭建深度学习模型。

使用 Echarts / d3.js 进行数据可视化。

使用常见的思维框架对问题进行分析。(e.g. MECE, AARRR, SWOT, BCG matrix, 3C, PEST)

其他工具经历: FineBI, Tableau, SPSS, matlab, Linux

工作经历 & GAP

浙江省农业融资担保有限公司

2022.11 - 至今

职位: 信息技术高级经理 (数据方向)

数据质量管理: 包括数据采集、清洗、整合、存储、处理和分析等环节的质量控制，设计和管理。

主要工作: 面向多数据来源 (政务数据、厂商接口数据、离线数据、业务数据)，设计数据采集加工存储方案，要求时效性和准确性的同时降低获取数据 (付费接口或受限数据) 的成本。

数据分类和标准化管理: 对企业数据进行分类和标准化，建立数据字典和元数据管理体系。

主要工作：数仓分层规划，制定表命名规则，并为重要数字资产表编号归档；对表的元数据、数据字典进行归档；对不同数据来源统一编码归档；设计数据资产目录。参与制定核心业务数据标准。

数据架构管理和优化：设计和优化企业的数据架构，包括数据模型、数据流程和数据集成等。

主要工作：根据上级部门对数据管控的要求，配合整改落地云端数据的方案设计，同时尽可能满足公司对数据的需求。

数据治理组织和流程建设：建立完善的数据治理组织和流程，明确职责分工和 workflows。

主要工作：在数据清洗方面，提出明确核心字段和临时表检查机制，重新规范并不断完善清洗流程，提高文档编写效率，提高开发及验收效率，提升清洗质量，降低沟通成本。

模型方面：参与并主要负责各类风险模型的（与业务人员探讨）需求计划制定、开发进度跟踪，模型验证。参与对于模型建设相关的数据挖掘分析。模型所需数据加工存储方案设计，如数据指标变动监控。对外部厂商数据进行建模分析、测试评估，为筛选采购外部数据提供决策意见。

数据应用/分析

- 1 涉农补贴项目名称挖掘扩充。
- 2 分析各地区客户经营状况，辅助业务寻找合适区域开发新产品。
- 3 预研并开发某识别接口，减轻开发人员工作量，缩短功能上线时间，保证识别率。
- 4 对模型输出分数进行等级区间分析，向风控部门输出推荐分级。

其他

- 1 原有模型验证方式主要是根据主体身份信息手动查询前端页面或者数据库获取信息然后人工汇总计算，效率低且易出错。借助 python 编写脚本将验证过程标准化步骤化，极大提升验证效率，降低人因错误率。
- 2 外部厂商数据产品（数据加密、决策引擎等）试用、测试工作，与厂商交流提出自己的建议，并对产品结合我司使用场景、适配可行性等情况向上级汇报。
- 3 日常性事务：取数、跑数、数据统计、分析；编写数据建设/应用需求，发布并跟踪进度、沟通协调开发人员；系统异常问题排查；接口管控模块运营管理，少量接口封装开发。
- 4 代码版本控制管理工作。

个人自主学习、动手实践项目和求职阶段

2022.04 – 2022.11

离职之前已经有调整职业方向的想法，开始主动学习相关知识和内容。

离职后继续学习并积极参与一些动手实践项目来巩固所学和加深理解。

兴业数字金融服务(上海)股份有限公司

2020.08 - 2022.04

职位：Java 后端工程师

原有系统接口维护和 bug 修复，开发新功能，参与新项目开发。

- 自动接口测试套件开发，快速验证系统基本功能运行有效性，节省大量繁琐手工验证工作量，极大提升验证效率。
- 推动旧版本系统用户进行数据改造并迁移至新版本系统。
- 使用新组件完成旧的可视化功能改造升级，用户体验提升。

- 独立负责过 ES 服务端搭建并开发、维护相关应用模块、上线自动部署脚本编写。
 - 数据产品-数据资产模块部分功能研发。
- Springboot / Mybatis / SpringCloud / Elasticsearch / GitLab / Mysql

项目经历

陌生人社交平台指标体系搭建 个人实践项目，无关联公司 2022.07 - 2022.08

- 【背景】 陌生人社交行业存在竞争，用户可能同时下载多款陌交 APP，流失情况难以控制，用户黏性不高消费不稳定，容易导致平台营收状况不佳。
 - 【任务】 本项目从 0 到 1 搭建一套指标体系雏形，用以客观量化业务的发展情况。
 - 【行动】 按照人货场模型将产品分成人货场三大块，人即用户，场即平台，货代表了平台的业务，分别梳理了各块的可能指标并使用思维脑图呈现。挑选若干指标进行数据埋点事件设计，并随机生成了用户行为数据。基于这些数据进行指标计算，并定制了指标看板。
 - 【意义】
 - 1 指标体系的梳理助于以更加全面的视角去观察和理解业务的发展情况。
 - 2 基于各项指标的历史波动结合实际情况，可评估出合理的业务运行区间建立预警。
 - 3 指标看板相比于直接阅读报表上的数字会更加的直观，趋势更加明显，时效性更强。
- 最终输出一套[指标体系（思维导图）](#) 和 [指标看板（web 前端）](#)

商店商品销量预测建模 个人实践项目，无关联公司 2022.07 - 2022.08

- 【背景 & 任务】 基于提供的日粒度历史销售数据和商品商店等信息，希望预测未来一个月(月粒度)的特定商店商品销量情况。
- 【行动】 从数据预处理、特征工程、模型(LightGBM)建立训练评估到最终预测结果产生全流程(包括后期优化)，均独立完成。
- 【结果】

最终预测结果在测试集公开榜上 RMSE 得分为 1.23xxx。该项目主要是为了锻炼个人的数据科学技能和熟悉机器学习建模流程。

电商销售数据挖掘 个人实践项目，无关联公司 2022.06

- 【背景】 某品牌客户在驱虫剂市场已经研发了多款上市产品，这些产品在电商平台已经产生了比较可观的收益，但也存在其他实力强劲竞争的对手。希望通过数据挖掘分析寻求市场增长点，进一步扩大盈利。
- 【分析思路】 从大到小，由总到分——由整个大市场占比开始，宏观分析未来发展的可行性，再针对潜力板块进行细拆。对于具备时间属性的特征，观察随时间的趋势。对于静态数据选择维度进行对比分析，必要时采用波士顿矩阵辅助产品分组。
- 【结果】 由数据得出若干分析结论辅助业务决策。

[项目链接](#)

- 【背景】人机交互是人工智能的重要应用领域，其中语音交互目前来说比较符合人们生活习惯，但是相关产品往往很难满足用户的各类别复杂要求，其根源在于自然语言本身较高的复杂性使得用户意图无法被较好的理解。自然语言理解（NLU）任务旨在让计算机具备理解用户语言的能力，从而进行下一步决策或完成交互动作，是具备语音交互能力的产品需要处理的重要任务。
- 【任务】本项目对经过语音识别技术处理过后的用户指令文本信息，使用模型去识别意图和抽取关键槽位信息。
- 【行动】对 baseline 模型微改，并进行语料增强。
- 【结果与收获】子任务表现获得明显提升，该项目获得了飞桨平台加精。该项目使本人对从数据处理到模型训练推理的流程有了更进一步的认知。

基于激光雷达的着落平面选择系统研究

本科毕业设计

2019.01 - 2019.06

- 【背景】借助激光雷达对周围物理环境进行感知具有速度快，精准度高且不受光线明暗的天然优势，使用软件算法对获取到的点云数据进行加工处理才能得到有用的信息，其中点云目标识别是一个广泛的研究问题。
- 【任务】本项目期望开发一个软件，对某 16 线激光雷达获取到的点云数据文件进行解析处理并计算识别出有效的目标平面。
- 【解决方案】C++/MFC 编程实现软件框架，OpenGL 绘制点云播放动画。平面识别起初使用了主流开源库 PCL，但是框架过重，在自己的笔记本上运行不畅。于是根据采集方式和数据文件结构的特点，自己定制了一种识别方法，在识别准确度上达到了视觉可接受范围，速度上每帧计算约 300ms，相比直接调用 PCL 库 API 播放流畅性有明显提升。
- 【优化尝试】后在满足功能需求后，从代码优化、尝试引入 SIMD 技术和 CUDA 编程等方面进行了相应优化探索，并总结效果差异和原因分析。

RoboCup 竞赛 救援 2D 仿真组

校队成员

2017.07 - 2018.04

- 作为团队成员主要负责维护预计算模块的分区算法，并与其他模块的队员联调测试。任期内将分区模块的算法 Kmeans 升级为 Kmeans++。
- 在 2018 年于绍兴举办的 RoboCup 机器人世界杯中国赛中，其他队伍耗时几分钟的预计算环节，我们组只用了半分钟左右，最终也获得该项目的团队冠军。

自学 & 在线学习

数据分析技能课程

2017.08 - 2018.03

- 使用 Mysql 存储数据及分析。
- 探索分析数据中的单变量、多变量相互关系，数据分布，异常，缺失值等。
- 使用机器学习模型进行预测。
- 使用开源库开发展板，可视化数据问题。

[完课证书](#)

数据分析 (互联网/运营/数据科学)

2021.11 – 2022.9

- 对 Banner 广告图片调整使用 ABtest，设计一类二类检验指标和相应假设，设计事件和埋点，基于样本数据使用 python 进行统计计算（最小样本量、统计量、P 值），并根据检验结果得出测试结论。
- 短视频平台日活指标异动拆解分析，并给出针对性建议。
- 使用 Numpy/Pandas/sklearn 包编写代码对销售数据时间序列进行构建乘法模型预测：提取季节项、拟合趋势，并使用 MSE 进行效果评估。 [\[plot\]](#)
- 使用 FineBI 对样例数据进行看板设计。 [\[看板\]](#)

获奖情况

12/2018	2017-2018 学年东南大学三好学生
06/2018	华为校级奖学金
04/2018	RoboCup 2018, 中国公开赛, 救援仿真组, 团队冠军
07/2017	RoboCup 2017, 日本名古屋, 救援仿真组, 团队第四名

求职标签

数据分析、挖掘建模、A/B Test、指标体系建设、数据可视化、框架分析、工程思维、coding/debug

兴趣爱好

电影、排球、乒乓球、羽毛球、徒步、烧菜、音乐

————— 感 谢 查 阅 —————