

张凯 (个人网页: <https://kaizhang0624.github.io/>)

教育经历

2019.09-2020.09 康奈尔大学 生物统计与数据科学 硕士

- 成绩排名: Overall GPA: 3.99/4.0 (学业成就奖)
- 核心课程: 机器学习、生存分析、因果推断、深度学习、层次模型与纵向数据分析

2015.09-2019.06 香港中文大学(深圳) 统计学 本科一等荣誉学位

- 成绩排名: Major GPA: 3.67/4.0 (专业排名 3/71) ;
- 奖项与荣誉: 2018-2019 学年本科生科研津贴; 2017-2018 学年理工学院院长嘉许名单 (学院 15%学生获此殊荣) ; 2017-2018 学年逸夫书院学业奖学金 (书院 10%学生获此殊荣)
- 核心课程: 回归分析、数据挖掘、数据管理 (SQL) 、统计推断、时间序列、随机过程、最优化、金融数据建模等



实习与研究经历

威尔康奈尔医学研究院-研究生研究助理

2020.01-至今

- 与两名教授和一名 PI 合作, 负责两个科研项目的统计与生物信息学分析。
- 运用非监督学习方法 (如 K-means, SNN 图聚类等) 分析老鼠视网膜单细胞 RNA 序列, 识别并特征化细胞子类别。
- 运用统计模型 (累积比数模型) 对纵向数据进行中介效应分析, 撰写计划摘要与学术论文 (in preparation) 。

深圳市大数据研究院-实习数据分析师

2018.09-2019.06

- 设计项目流程并进行特征工程, 从原始 WiFi 连接记录中提取描述大学生在校活动的行为特征。
- 基于行为数据, 运用机器学习方法 (XGBoost) 开发大学生 GPA 预警模型, 成果被港中深智慧校园平台接收。
- 基于 Python NetworkX 库开发算法, 利用 WiFi 连接数据追踪、特征化与可视化个人轨迹网络, 为研究院后续多个子研究项目提供框架基础与数据支持。

长江证券风险管理部-数据分析实习生

2018.07-2018.09

- 参与上市公司信用风险评分卡开发项目, 负责对财务数据进行特征工程, 并基于逻辑回归搭建评分卡模型。
- 日常向部门上级提交并汇报数据与行业报告。

项目经历

基于老鼠视网膜单细胞 RNA 序列的聚类分析与细胞类型识别 (指导老师: Kathy Zhou 教授)

- 该项目利用机器学习方法 (聚类) 分析老鼠 RPE 细胞的 RNA 序列数据, 并利用统计学方法发现与疾病有关的基因。
- 预处理原始单细胞 RNA 序列数据, 包括液滴筛选、质量控制、特征选择、降维与可视化。
- 运用多种聚类算法对 RPE 细胞分类, 包括 K-means、DBSCAN 和 SNN 图聚类等, 并利用配对 t 检验识别基因标记。

基于数据自适应方法的男性职业与薪酬关系的因果推断 (指导老师: Ivan Diaz 教授, [报告链接](#))

- 该项目基于机器学习方法 (Stacked Regressor) 与反事实分析探究男性职业与薪酬之间的因果关系。
- 负责预测模型 Stacked Regressor 的搭建, 元回归器包括逻辑回归、弹性网络模型、随机森林和 XGBoost 等, 并运用多种反事实统计方法 (包括 AIPW 和 TMLE 等) 估计因果效应并证实其显著性。

基于行为数据的 GPA 预警模型搭建

- 该项目通过 WiFi 连接收集多种描述大学生在校活动的行为特征, 并利用机器学习方法预测学生 GPA。
- 设计流程并开发算法, 从原始 WiFi 连接数据中提取 1710 名本科生的行为特征, 包括出勤率、各建筑月停留时间等。
- 基于行为特征和档案特征, 运用 XGBoost 算法开发 GPA 模型 (准确率 96.3%, 召回率 81.8%) 。

技能兴趣

- 技能与语言能力: R (tidyverse, ggplot2, CARET)、Python (pandas, scikit-learn, keras)、SQL (MySQL, PostgreSQL)