

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: 3.98/4.0
- 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: Overall GPA: 3.40/4.0 (专业 7/71) ; Major GPA: 3.66/4.0 (专业 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)。
- 定期 (两周) 向导师与客户提交研究报告并汇报项目进展。

深圳市大数据研究院-实习数据分析师

2018.09-2019.06

- 设计流程并开发算法, 将原始 WiFi 连接数据转换成描述大学生在校活动的行为数据。
- 基于行为数据开发预测大学生 GPA 的机器学习模型, 包括随机森林和 XGBoost 等算法, 最终模型达到 96.3%准确率与 81.8%敏感度。
- 基于 Python NetworkX 库开发算法, 利用 WiFi 连接数据追踪并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-信用风险实习生

2018.07-2018.08

- 参与信用风险评分卡开发项目, 负责对财务数据进行数据预处理, 并基于逻辑回归搭建评分卡模型。
- 日常向项目负责人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)

- 该项目基于机器学习方法 (Super Learner) 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 Super Learner 的搭建, 并运用两种统计方法 AIPW 和 TMLE 估计因果效应并证实其显著性。

预测 COVID-19 新冠病毒住院患者呼吸机需求的模型搭建

- 基于生命特征检测数据, 搭建神经网络模型预测纽约长老会医院新冠病毒住院病人五天内对呼吸机的需求。

技能兴趣

- 技能与语言能力: R (tidyverse, ggplot2, shiny, CARET)、Python (pandas, scikit-learn, keras)、SQL (MySQL, PostgreSQL)
- 英语 (GRE 324 V: 154/Q: 174)