

杜智浩 (Johnson Du)

Tel: +1 (510) 833-4417 | E-mail: zhihao617@berkeley.edu | 目标岗位: 机器学习方向

教育背景

加州大学伯克利分校

GPA: 3.7/4.0

2019.08 – 2023.05

- 统计学 – 现代统计预测与机器学习; 线性模型的理论和应用; 数据计算概念; 基础数据科学; 概率论; 基础统计学; 线性代数
- 计算机科学 – 机器框架与结构 (A+); 数据结构和编程法; 离散数学和概率论; 信息设备和系统; 计算机程序架构

上海包玉刚实验学校

最终成绩: 40(+2)/42(+3)

2007.09 – 2019.06

- IB 国际课程双语文凭 – 高等物理, 高等数学

项目经历

基于空气指标对北京 PM2.5/PM10 进行实时预测

数据分析项目组长

2021.11 – 2021.12

- 该项目旨在利用北京环境监测中心 2013 年至 2017 年的数据 (30,000 项, 12 个特征) 对北京的 PM2.5/PM10 指数进行实时预测
- 首先, 通过观察单变量分布、双变量关系和分类变量的交互影响来执行深入探索数据分析 (EDA), 再使用向后逐步 AIC 信息准则进行变量选择, 对变量选择后的模型进行了前提条件和异常值诊断
- 不断优化模型, 最终使其均方根误差达到 0.6, 并为其构建预测区间; 最终, 将模型及所得数据可视化, 主导并撰写小组最终研究报告

HOYO 实验室

项目组长

2021.03 – 2021.08

- 该项目是以兴趣驱动、数据为中心, 围绕 RPG 游戏 Genshin Impact 的自发性项目:
- 项目一“伤害计算器”: 使用带有随机回归量和特征映射 (feature mapping) 的线性回归 (16 个特征) 预测游戏的伤害计算机制; 基于从游戏内模拟收集的 500+ 行数据, 用原始的特征映射函数初始化模型, 并通过交叉验证最终确定特征映射函数, 达到了接近于 0 的误差; 使用了 R 中的 Shinyapp 包制作了网络应用程序以展示最终模型^[1], 该模型制作的教程在游戏社区平台获得了 400,000+ 浏览量及 50,000+ 点赞量
- 项目二“天空岛数据库”: 该项目旨在根据战斗挑战中角色使用率建立角色强度排行榜; 带领 3 人小组, 使用 Python requests 爬虫及虚拟标题克服了应用程序保护屏障, 在与游戏相关的移动应用程序中爬取了 1000+ 用户个人资料; 解码 JSON 文件以提取相关数据, 在 MySQL 数据库中每三周存储并更新一次数据, 最终在网站上展示了最终的角色强度排名^[2]

上海数聚软件系统股份有限公司

实习生

2021.05 – 2021.08

- 采用 Kettle 和 MySQL 执行数据提取-转换-加载 (ETL) 技术, 与小组三人将来自两个过时数据源的 300+ 张表格合并到最新数据源上
- 在 SQL 中筛选并审定多个主键以及必要的元数据列来构建目标表格的架构, 独立完成数据清洗, 并通过数据分组、unpivot、及修改数据类型来重新调整不兼容列, 对最终合并后的表格进行诊断检测
- 该 ETL 项目已帮助客户复星医药升级到最新的、性能更高的数据库系统

技能

熟练: R (机器学习类: caret, MASS, rpart; 数据类: ggplot2, dplyr), C, MySQL, MS Office, RISC-V

良好: Python, HTML, CSS, Java, SQL Server

证书及奖项

包陪庆奖: 荣获包玉刚实验学校 2019 级前三名学生

兴趣爱好

钢琴 – 于高中毕业典礼为全校师生家长演奏; 合唱 – 参予了 ISCMS XII 与 ISCMS XIII (国际学校合唱协会);

小提琴 – 高中管弦乐队第二小提琴; 足球 – 高中足球校队后腰

[1]: https://zhihao617.shinyapps.io/genshin_basic_damage_calculator

[2]: <https://www.youngmoe.com>