

# 钟俊彦

广东省广州市 (城市)  
510080 (邮编)

17665458796 (电话)  
pzsx005@gmail.com (邮箱)  
github.com/W55699 (GitHub)



## 教育经历

- **华东交通大学** 机电与车辆工程学院 · 车辆工程 2015.09 – 2019.07 学士学位  
– 数学建模, 数值计算, 控制器原理, 程序设计 (C++), 自动控制原理, 列车网络控制
- **澳门科技大学** 资讯科技学院 · 应用数学与数据科学 2019.09 – 2021.08 硕士学位  
– 数值线性代数, 应用统计分析, 机器学习, 时间序列分析, 数据挖掘, 数据库系统, 数据科学程序设计

## 技能及研究方向

- **语言:** Python, C/C++, MATLAB,  $\text{\LaTeX}$ , Java
- **大数据类工具:** MySQL, Hive
- **英语:** 英语能力较强, 可用英语作为工作语言 (CET-6 527)
- **研究方向:** 图神经网络, 推荐系统安全, 对抗训练

## 工作经历

- **汇丰科技** *python* 开发工程师 2022.07-2023.03  
– 利用 python, hive, shell 脚本对金融数据进行清洗, 利用 jenkins, git 对代码进行 CI/CD 管理。  
– 利用 GCP 平台对于清洗后的数据进行欺诈可能性的预测。
- **品友互动** 策略算法工程师 2021.02-2022.06  
– 利用 Hive, 以及 Sqoop 等工具等数据进行同步以及清洗。  
– 结合业务理解构建机器学习广汽本田复购模型 (特征工程、模型训练), 并进行代码管理和优化。  
– 和思必驰合作参与构建语音语义文本分析模型, 主要参与文本的清洗以及后续业务报表构建工作。

## 实习经历

- **AIATSS(友邦资讯科技公司)** 测试组数据分析实习 2020.04-2020.6  
– 撰写 SQL 以及 python 脚本校验公司内部数据。  
– 利用 jira 实时监控工作流程进度, 并通过 Excel pivot table 绘制组内测试进度报告。  
– 对测试流程以及 ETL 开发流程有了更深入的了解。
- **TCL 工业研究院** 数据挖掘实习 2020.07-2020.09  
– 通过组内讨论, 参与制定推荐系统 CTR 的业务指标, 并基于以上指标进行统计分析。  
– 利用 spark 负责数据清洗以及异常数据的核验。  
– 参与组内的论文讨论, 并参与大规模特征数据的分类 (Random Fourier features SVM)、聚类 (minitach kmeans) 工作, 并参与特征筛选以及特征交叉工作。  
– 参与组内爬虫代码的日常维护, 丰富自身挖掘经验。

## 项目经历

- **美国大学生数学建模竞赛** 队长 2017.02  
– 通过层次分析法对城市居住环境影响因素进行统计分析, 分析影响因素及其权重。

- 通过模糊综合评价与灰度预测法预测各项指标的变化。
- 研究城市可持续发展模型，与 2 位队友共同讨论后，撰写可持续发展数学模型英文论文，描述各项指标变化。
- **泰迪杯 队长** 2018.03 – 2018.04
  - 对电视节目收视率进行量化分析，根据收视分析改进收费节目策略，策略波动显著降低
  - 研究了协同过滤算法 (Item-based, User-based)、SVD、FM 等算法在推荐系统中的应用
  - 利用 py-spark mllib 对算法进行实现，最终成绩位列 top 15%。
- **Douban web crawler data analysis 个人项目** 2019.09 – 2019.12
  - <https://github.com/W55699/doubanbook-web-crawler>
  - 利用正则表达式，并通过创建线程池，多线程爬取豆瓣书籍信息。
  - 将信息生成 csv 文件，并将信息存入 mysql 数据库。
  - 利用 pandas 读取 csv，并做数据可视化分析以及统计分析。
  - 通过 pca 将数据进行降维，提取关键信息，然后通过 k-means 算法进行聚类分析。
  - 根据 pca 降维后的信息，同时结合数据的标记，将数据分为训练集和测试集，并将数据进行二分类，比较各种分类方法如 SVM, LR, 决策树，随机森林算法的优劣。
- **IMDB sentiment analysis 个人项目** 2020.10 – 2020.12
  - 利用 stop-words 对数据集进行清洗，并通过 wordcloud 进行词云可视化。
  - 利用 python gensim word2vec 对文本进行向量化处理。
  - 训练并调整 bi-lstm 模型，使模型准确率在测试集中达到 85%。
  - 利用 docker, Tensorflow-serving, streamlit 对模型进行部署，实现可视化。
- **广汽本田增换购预测 品友互动** 2021.02 – 2022.06
  - 该项目主要利用车辆修理特征以及车主在官方微博、微信的特征对车主增换购进行推荐。
  - 数据预处理：使用 sqoop 将数据从数据库导入到 Hive。然后使用 hive SQL 使每个进程的数据（数据清洁并定义正样本和负样本）
  - 机器学习模型训练以及预测：使用 pandas 导入数据，做特征工程。然后我们使用 catboost、xgboost、逻辑回归模型来训练模型和预测。同时，我们根据得到的反馈调整每个模型的输出权重。
  - 数据输出：结合业务逻辑过滤，输出模型前 0.5% 预测结果到输出数据库，最后 ETL 团队将数据发送到各个 4s 店。ETL 团队会将这些数据传输到 4s 店以呼叫客户进行换购推广。
- **Credit card financial risk analysis 汇丰科技** 2022.07 – 2023.02
  - 这个项目主要是利用用户特征和消费特征对 HSBC 信用卡进行风险评级。
  - 预处理：利用 python, hive, shell 脚本对金融数据进行清洗，利用 jenkins, git 对代码进行 CI/CD 管理。
  - 数据建模：在 gcp 平台利用 python 对数据进行特征工程，利用 tensorflow 构建 tabnet 模型进行预测，并与 LR 模型进行比较。
  - 模型输出：将风险最高的用户 1% 进行输出，并利用 seaborn 对其特征进行可视化。

## 获奖情况

- **2017 美国大学生数学建模竞赛 二等奖** 2017.02
- **第六届“泰迪杯”数据挖掘 三等奖** 2018.04

## 论文以及专利

- **一种可单轨行走的转向架 实用新型专利** 2017.11
- **一种联网联动小型化多功能晾衣机 实用新型专利** 2017.11

- 一种基于激光检测独立供源的共享单车区域护车系统 实用新型专利 2017.11
- JUNYAN ZHONG,HUIBIN WANG,KIN TAK U. Matryoshka Attack: Research on an Attack Method of Recommender System Based on Adversarial Learning and Optimization Solution,ICWAPR, 2020.
- JUNYAN ZHONG,HUIBIN WANG, Robust graph embedding recommendation against data posioning attack CIKM 2022(under review)