

教育经历

Master of Science in Data Science (Offer)

School of Data Science and School of Management and Economics [🔗](#)

09/2022

计算机科学与技术(数据科学方向) 学士学位

兰州大学, 信息科学与工程学院, 数据科学系 [🔗](#)

09/2018 - 06/2022

GPA: 85.44/100

主修课程:

- | | | |
|----------------|-----------------|-----------------|
| ▫ 数据科学导论 (95) | ▫ 机器学习 (100) | ▫ 信息检索系统 (91) |
| ▫ 云计算与大数据 (93) | ▫ 数据挖掘应用 (96) | ▫ 信息可视化 (91) |
| ▫ 探索性数据分析 (99) | ▫ 数据科学编程II (91) | ▫ 面向对象程序设计 (97) |

实习经历

大数据开发实习生 - 数据开发部

袋鼠云(杭州玳数科技有限公司) [🔗](#)

08/2020 - 09/2020

职责:

- 通过阿里云 Dataphin 智能大数据平台根据客户需求建立报表, 源数据来自客户提供的 SQL Server 中的亿级销售数据
- 将 Dataphin 生成的报表按需定时全量/增量同步到 Quick BI 数据可视化分析工具中
- 利用 Quick BI 对报表进行可视化, 方便客户对全国门店的销售情况进行分析

项目与竞赛

IEEE Radar Conference Radar Challenge - EMSIG Hackathon "Radar-based Human Activity Recognition Challenge 2021" (09/2021 - 10/2021) [🔗](#)

- 对给出的六种不同人体动作的雷达信号进行处理, 训练模型, 对其进行分类
- 将雷达信号转换为时频谱图, 进行数据增强, 使用 Cait(Class-Attention in Image Transformers) 模型结合5折交叉验证进行训练与验证
- 分类准确率在 14 个参赛队伍中排名第一, 同时在此数据集的研究中达到世界第一 (98.5%)
- 完成论文 *A ViT Approach for Short-range behaviour recognition using Radar Signals*, 负责所有代码工作和实验结果分析

国家级大学生创新创业训练计划 - 多重分形在金融市场的应用 (03/2021 - 09/2021) [🔗](#)

- 从金融数据中利用多重分形消除趋势波动分析法 (MF-DFA) 提取分钟级时间序列的多重分形特征, 并对该多变量时间序列进行预测
- 负责实验, 使用多种机器学习及深度学习模型 (e.g. Prophet, DeepAR, Temporal fusion transformers) 对该多变量时间序列进行回归预测, 对比在该数据集上的表现
- 参与撰写论文《基于多重分形谱的股价指数特征提取及预测》的实验结果分析部分 (投递中)

Kaggle Competition-CommonLit Readability Prize (05/2021 - 08/2021) [🔗](#)

- 使用 NLP 来识别一段文字的适当阅读水平, 为 3-12 年級的课堂使用的阅读段落的复杂性进行评级
- 在给定数据集上对预训练的主干模型 (RoBERTa, Electra, DeBERTa) 进行微调, 使用不同的 classification heads (Attention/CNN/Dense) 得到预测结果, 最后使用 ensemble learning 来结合所有模型的结果
- 在 Private Leaderboard 上取得了 38/3682 的排名 (前 2%, 银牌)

项目与竞赛

云计算与大数据课程设计 - 基于 Spark 与 LightGBM 的脉冲星甄别模型 (05/2021 - 06/2021) [🔗](#)

- 使用 Databricks 的 DBFS 分布式文件系统进行数据管理
- 使用 Spark SQL, matplotlib, Seaborn, Plotly 进行探索性数据分析并进行特征选择
- 使用 MMLSpark(Microsoft Machine Learning for Apache Spark) 构建分布式 LightGBM 模型, 使用网格搜索寻找最佳超参数, 在验证集上获得 F1-Score = 98%

技能

Python 机器学习 & 深度学习	Numpy, Pandas; Scikit-learn, XGBoost, LightGBM; PyTorch, PyTorch Lightning; TensorFlow, Keras	数据科学工具 & 编程环境	Git, SQL, Linux, Docker, Weka, Excel, Conda
可视化及探索性数据分析	Python: Matplotlib, Plotly, Seaborn; R: Tidyverse	英语	TOEFL iBT: 98(MyBest Scores: 100); CET-6: 535

PUBLICATIONS

Conference

A ViT Approach for Short-range behaviour recognition using Radar Signals [🔗](#)

03/2022

2022 IEEE Radar Conference (Accepted for Publication)

Second Author

Conference

Empirical Study on the Development of Big Data Boosting the Economic Development of Tourism Industry -- based on Spatial Panel Data [🔗](#)

03/2022

2022 7th International Conference on Big Data Analytics (Accepted for Publication)

Second Author

荣誉奖项

校级二等奖学金 (10/2021)

兰州大学

证书

Feature Engineering [🔗](#)

Kaggle

SQL for Data Science [🔗](#)

Coursera(University of California, Davis)

Data Science Ethics [🔗](#)

Coursera(University of Michigan)

Digital Media and Marketing Strategies [🔗](#)

Coursera (University of Illinois at Urbana-Champaign)

Using Machine Learning in Trading and Finance [🔗](#)

Coursera (New York Institute of Finance and Google Cloud)

Machine Learning Explainability [🔗](#)

Kaggle

Hadoop Platform and Application Framework [🔗](#)

Coursera (University of California San Diego)

Data Science Math Skills [🔗](#)

Coursera (Duke University)

Corporate Financial Decision-Making for Value Creation [🔗](#)

Coursera (The University of Melbourne)

Using Python to Access Web Data [🔗](#)

Coursera (University of Michigan)