

何晨

(+33) 6-65-34-59-13 · mr.chen.he@outlook.com · 数据科学家 · hechen95.github.io

专业技能

- 职场软技能: 团队合作能力, 抗压能力, 分析能力, 报告撰写, 主动沟通以及表达能力。
- 语言水平: 英语 (专业工作水平, B2), 法语 (专业工作水平, B2), 普通话 (母语水平)。
- 报告撰写: Latex, PowerPoint, Word, Excel, MS Visio, Python(matplotlib), R(ggplot2)。
- 编程语言: Python (Numpy, SciPy, Pandas, Scikit-learn, PyTorch, TensorFlow, 等), R (Dplyr), C++ (Cplex, Gurobi), Java, Git, Unix shell 脚本, Spark, Hadoop。
- 数据检索: MySQL, MS SQL, PostgreSQL, AWS, Azure, NoSQL, Spark SQL, HiveQL。
- 操作系统: Linux (Debian, Ubuntu, CentOS), Windows。
- 数据驱动建模: 定量分析, 概率论, 统计学 (A/B 测试、假设检验等), 预测算法 (分类、回归、聚类、时间序列预测等), 关联规则挖掘, 异常检测, 深度学习, 运筹学建模, 数据驱动优化。

职业经历

圣艾蒂安国立高等矿业学校 | 圣埃蒂安, 法国 | 全奖博士生 2019.10-2022.12

- 我们通过将数据点投射到拓扑空间并识别高风险的误编码子空间来优化医院 ICD 编码的误编码率。我们的成果在 IEEE CASE 2021 国际会议上进行了展示。
- 将再入院的不确定性 (或风险) 纳入考量, 我们优化了法国一家地区性医院的再入院率。我们的结果发表在 IEEE CASE 2020 国际会议的论文集中。
- 担任研究生助教 (TA) 超过 120 小时, 履行所有助教职责, 包括讲课, 答疑和结业答辩评分工作。
- 所需技能: Python, R, 统计学, 贝叶斯推理, 机器学习算法, 关联规则挖掘, 数学规划, Git。

休伯特·居里实验室 | 圣埃蒂安, 法国 | 机器学习实习生 2019.03-2019.09

- 实现了一个用于图像轮廓检测的生成对抗网络 (GAN) 框架, 并优化了其性能 (准确性、精确度和召回率等)。
- 针对 FPGA 上的图像分类任务, 对 CNN 参数的精度变化对分类性能 (即推理速度和精确度) 的影响进行了定量研究。
- 所需技能: Python (Tensorflow), Shell 脚本, Git。

XXII Group 公司 | 巴黎, 法国 | 机器学习实习生 2018.07-2018.09

- 在 Nvidia Jetson TX2 嵌入式开发板上优化了一个基于 Yolo v3 的实时行人检测项目, 包括超参数调优、损失函数优化和 SGD 算法优化。
- 在 Nvidia Jetson TX2 上使用 NVIDIA TensorRT 将神经网络推理速度从 12FPS 提高到 22FPS。
- 定制了多个 docker 镜像, 用于简单快速地将行人检测项目部署到 Nvidia Jetson TX2 上。
- 所需技能: Python (Yolo 框架), Docker, Shell 脚本, Git。

项目经验

新闻过滤项目 | 圣埃蒂安电信学院 | 硕士研究生 2018.10

- 分析并自动选取与公司收入有关的新闻报道以向公司提供决策支持。在此项目中, 我贡献了 NLP 和机器学习方面的专业知识, 包括词嵌入、分类算法、性能评估等知识。项目源代码可以在 Github 上获取, @he-chen-95/M2-Data-Mining-Project。
- 所需技能: Python, 词嵌入 (TF-IDF, N-gram), ML 分类算法 (即 Naive Bayes, 决策树, 等等), Git。

还款能力预测项目 | 圣埃蒂安电信学院 | 硕士研究生 2018.06

- 自动分析贷款人的财务状况并预测客户的还款能力。在此项目中, 我在风险建模, 特别是通过应用 Logit 模型预测客户的还款能力方面贡献了专业知识。项目源代码可以在 Github 上获取, @he-chen-95/FISE3-Big-Data-Project。
- 所需技能: Python, 数据降维, 逻辑回归, NoSQL, Hadoop, Git。

教育背景

圣艾蒂安国立高等矿业学校 | 工业工程 | 在读博士研究生 2019.10 - 2022.12

工业工程应用于智慧健康领域。研究课题。”挖掘、理解和优化医疗误编码行为以改善编码实践”。论文指导人: 谢晓兰教授 (上海交通大学兼职教授)。

圣埃蒂安高等电信学院 | 计算机科学与技术 | 理学硕士 2016.09 - 2019.10

主修数据挖掘和机器学习相关课程。在 2018 年 9 月至 2019 年 10 月期间, 同时修读并获得《数据和连接系统》的第二硕士学位 (理学硕士, 圣埃蒂安大学)。

西安电子科技大学 | 电子科学与技术 | 工学学士 2013.09 - 2016.10

获得 2014 学年国家励志奖学金, 参加中法 3+3 本硕联合培养计划赴法国交流学习。