

简历



1998.02.16

9 浙江省丽水市

女

汉族

+86 13355916822

2569494688@qq.com

教育背景

浙江万里学院

数据科学与大数据技术-本科

相关课程： 数据结构与算法、面向对象程序设计(java) 、python、 计算机操作系统、mysql 数据库、
java Web,HTML,CSS,JS,机器学习

经历

北京哒线科技有限公司 （python 爬虫实习） 2022.03 到 2022.11

工作描述：

- 1： 对数据的清洗整理， 关联数据结构获取数据信息
- 2： 批量爬取学校网， 教育网， 医疗类网站各大网站信息， 提供数据， 解决反爬策

略上海芯化和云科技有限公司 （python 工程师） 2022.12--2024.10

工作描述：

- 1： 大模型数据供应： 为公司提供化学品数据、供应商数据， 获取的渠道有（化学书、化工网、1688、百度爱采购 化工云、化源网、包括一些海外港口数据等）
- 2： 自动化清洗数据： 定义从爬虫到清洗规则的自动化， 并且针对产品的合并联系方式以及去除无效手机等细节 达到业务的数据需求
- 3： 产品分析： 针对化学品产品的价格找到影响的因子关系， 通过皮尔逊/复相关等计算关系和回测等分析该因子 对产品的价格影响走势
- 4： 价格预测： 通过 PLS 算法预测价格的走向和趋势

索引教育 （NLP工程师） 2024.11---至今

- 1:针对英国，美国，澳大利亚，香港等大学批量爬虫，更新数据
- 2:AI择校项目的维护，智能回答，更新数据，中英文的翻译，RGA本地知识库生成
- 3:小红书数据爬虫（标题，正文，点赞，收藏和评论）
- 4:飞书文件的爬虫分析，生成每日报告
- 5: DIFY：
 - 1：邮件分类（学院申请类，基本信息类，住宿类，缴费...）；
 - 2：试卷出题
 - 3:试卷检测

技能

- 1: 掌握 python 语言、mysql 语言、selenium、scrapy 分布式框架、正则表达式、jieba、LSTM 神经网络、Transformer、word2vec
- 2: 熟悉: NER、TF-IDF、huggingface langchain、dify、微调, auto、autogens

项目

- 1: 腾讯招聘信息平台（输入希望招聘岗位可出现相对应信息）

随时同步腾讯招聘平台，在平台中输入向往的职业可得到相关信息并且保存到数据库中，显示网页上并且计算此行业在所有行业之中招聘的情况绘制图像

- 2: 大模型数据供应+自动化清洗

通过1688，百度爱采购，化学书，化工网，化工云渠道每日从数据库中读取产品爬取数据，并且根据业务要求清洗成需要的数据合并一家公司多个产品和联系方式等

- 3: 价格预测模型

使用PLS（最小二乘法）通过化学品的上游或者下游预测产品未来1天，7天，15天的价格走势

- 4: 命名体识别

提取和标注大量数据文本的关键信息，使用NER，先对数据进行标注，在通过huggingface中选择模型微调，训练好保存加载预测分类

- 5: 中文词预测

使用LSTM训练模型，预测文本一些不完整的语句加以补充，用于预测下一个词

- 6: 垃圾邮件分

类 7: 文本分类

:

采用TEXTCNN模型训练多分类，对不痛渠道类型的数据进行训练，对杂乱无章的数据文本进行分类处理

CSDN博客: https://blog.csdn.net/qq_52491868?spm=1011.2415.3001.5343

自我评价

责任心强，对工作耐心细致，具有强烈的责任感，遇到问题会迎难而上，不断学习不断进步，有较好的沟通能力，理解力，有积极的心态面对问题，有良好的协作力,善于发现问题并且解决问题