

戴凌雪

15200866485 | Lingxue_dai@163.com | 杭州
离职，随时到岗 | 杭州/上海/深圳/北京/广州/成都 | Java研发工程师



教育经历

中山大学	985 双一流	2020年09月 - 2022年06月
计算机技术(推荐免试) 硕士	计算机学院	广州
中南大学	985 双一流	2016年09月 - 2020年06月
计算机科学与技术 本科	计算机学院	长沙

专业技能

- 熟练使用Java、熟悉springboot、mybatis等框架
- 熟悉MySQL、PostgreSQL及Linux常用命令
- 熟悉redis、rocketMQ等中间件

工作经历

上海恒生聚源数据服务有限公司	2019年11月 - 2020年01月
技术实习生 数据业务研发部	上海
阿里巴巴	2021年06月 - 2021年08月
java研发实习生 cto数据技术及产品部-QuickDecision	杭州
阿里巴巴	2022年07月 - 2023年07月
java研发工程师 瓴羊-Dataphin	杭州

1. 数据集成

- **长板建设**：支持hadoop多引擎，涵盖主流版本，如cdh5、tdh、argodb等，并且支持对orc/parquet/text/holedesk主流数据格式的读取写入，同时提供一键建表和整库迁移能力，提高产品竞争力。
- **能力建设**：新增支持多种数据源，例如starrocks、kudu、redis以及argodb，进一步完善离线集成管道组件功能。
- **体验优化**：针对客户使用体验以及业务场景复杂性，对数据集成相关产品进行体验优化及功能扩展，比如调度依赖自动解析，文件系统类字段列表自动识别、引擎分区不存在时选择策略等等，覆盖更多使用场景，提升用户体验。

2. 资产治理

- **资产目录**：汇聚dataphin各种类型的数据资产，提供高效搜索、查看详情的能力。支持资产的分层管理能力建设，提高资产查询管理的效率并提高数据质量以简化资产后链路的消费。
- **数据标准**：对元数据进行标准化定义，以覆盖全域资产、全生命周期落标稽核。增强与研发链路的关联、支持标准的批量下载与上传，增强事中监控并方便客户对数据进行迁移与批量标准化。

项目经历

分子连接物生成（自然语言处理&强化学习）-国家超级计算广州中心	2020年08月 - 2022年10月
描述：搭建基于transformer的强化学习模型，生成具有预期化学性质的分子连接物。	
方法及模型：将SMILES连接物生成问题类比为句子补全问题，借助机器翻译模型并利用化学信息学开源工具包RDkit及预训练的活性打分器设计多种针对分子目标化学性质的Rewards，通过策略梯度法优化序列生成过程，实现了生成分子向目标性质值的偏移。	
论文：DRlinker:Deep Reinforcement learning for Optimization in Fragment Linking Design	
专利：一种基于深度学习连接片段的化合物优化方法	
量化投资市场预测（时间序列问题 & 机器学习）-Kaggle & Jane Street	2020年12月 - 2021年03月
描述：通过历史市场数据，构建交易模型识别盈利机会并快速判断是否执行交易	
方法及模型：对百万级数据构建了适用于时间序列的交叉验证方式，借助EDA特征分析结果筛选特征，通过映射将回归简化为分类，使用特征中和方法降低过拟合，构建MLP\CNN\XGBoost模型并进行模型融合进而对历史市场数据进行建模，获得较为	

理想的效果，排名114/4245 (Top 3%)

荣誉奖项

中南大学三等奖学金、优秀毕业生	2017/2018/2019、2020
微信小程序应用开发赛华中赛区二等奖	2019
美国大学生数学建模大赛 (MCM) M奖 (一等奖)	2018