Song jie

教育背景

2007/9——2011/7　　上海大学　　全日制本科

英语水平 CET-6

工作技能

1) 熟悉 Python 以及相关的核心知识，熟悉 Python 后端开发框架，具备良好的面向对象的编程思想和编码能力，有阅读源码的良好习惯，精通数据处理相关的 pandas 等第三方包和工具，熟悉常用的数据结构与算法。
2) 熟悉 SQL 语言，掌握存储过程的编写和开发，postgresql 和 MySQL 数据模型，熟悉常用的调优方法。
3) 熟悉 Hadoop 相关的大数据生态体系和分布式知识，大数据主要计算框架 hive，MapReduce 过程；Spark 核心处理机制 DAG，RDD 过程，深入研究和阅读相关源码，理解相关的原理和发展历史。
4) 熟悉数据仓库的建模理论，分层方法（ODS、DWD、DWM、DWS、ADS）和数仓模型：星型模型，雪花模型，星座模型，业务知识丰富，能根据公司业务和产品进行梳理分析和设计原型结构，熟悉 datalake 数据湖，湖仓一体的有关知识内容。
5) 熟悉常用的大数据架构 Lambda，Kappa 架构，能够从多维度进行业务分析，提炼出相关的数据模型和数据指标，理解常见数仓架构体系对应的优缺点，熟悉离线数仓和实时数仓的数据同步，数仓开发建模，数据链路搭建方法。
6) 熟悉常用的机器学习算法 AI 模型以及相关原理（线性回归，逻辑回归，决策树，随机森林，GBDT，xgboost，时间序列等），能够根据相应的业务场景进行对应的分析应用（分类预测，回归，推荐等），了解深度学习相关的框架 TensorFlow，Pytorch 等，以及 NLP 文本挖掘算法。
7) 熟悉常用的 BI 工具如 tableau，帆软等。
8) 熟悉 AWS 云数据产品和解决方案，对 AWS 的数据产品以及 data lake 有深入理解（主要数据产品组件 Glue，S3，lambda，Stepfunction，SNS，redshift 云数仓，EMR 大数据，Athena，Cloudwatch，Jupyter，Sagemaker 等产品）。
9) 熟悉常用的 Scrum 模式，git 和 JIRA，Azure DevOps，CI/CD，具备一定的 devops 运维开发能力，熟悉有关的监控工具。
10) 熟悉网络 TCP/IP socket，负载均衡 load balance，掌握多线程，异步的开发编程等相关知识。
11) 熟悉云原生相关技术包括容器，微服务，DevOps，CI/CD，分布式和集群运行模式，涉及 docker，k8s，Prometheus 等常用工具组件。

自我评价

- 业务经验丰富，工作技能扎实，同时有很强的学习和理解能力，对数据驱动业务以及云计算，大数据，AI 技术应用和云原生技术有强烈的追求，能不断学习和实践解决问题，善于进行迭代创新。
- 金融，零售，汽车等多行业领先公司的 ToB，ToC 数据项目经验（涉及 BI，DS，ML 等多方面），知识面广泛，具有一定的架构规划能力，代码重构，性能优化，能及时根据业务需求和新环境业态下的数字化转型要求，不断进行自我驱动，快速应对新环境与行业变化的学习能力和适应能力。
- 严谨的职业操守，对数据质量，数据管理和治理，数据信息安全等有深刻的认识和良好的数据素养，认真负责踏实的专业工作态度，负责过项目团队中的各主要职责，拥有积极主动良好的团队协作沟通能力。
- 大型云原生数据项目落地经验，具备云平台经验，熟悉云平台的敏捷开发模式，能根据相关的业务场景和数据量情况，进行相关的数据业务流程和数据架构应用和分析能力（redshift，EMR，odps，hologres）。

项目经验

2022/8——2023/3 知名外资零售集团数据项目（VF BI）

使用技术：AWS + glue + pyspark + redshift + python + tableau

项目描述：知名 global 外资零售集团 APAC 云数据项目，满足向云端的统一迁移，主要是对已有的各系统数据平台进行统一的整合迁移，同时在 AWS 平台上进行数据需求的开发，完成整个数据迁移，数据开发，ETL 流程。

职责描述：

- 负责数据模型和数据流程工作，涉及整个数据湖 data lake 的 ETL 任务，核心架构是 S3 + Lambda 函数+glue + redshift 数仓，结合 eventbridge 和 Stepfunction 的调度配置，完成整个 data pipeline 的作业流程，为终端用户呈现 tableau dashboard 的 BI 应用落地。
- 在 CRM，CDP 数据项目中，根据 prototype 的需求和相关的页面链接埋点，来开发整体的 dashboard 报表，完成整个的 data pipeline 处理，在 tableau 上分析展示客户行为数据和用户画像，帮助业务团队根据用户网页访问行为来更好的分析和提升产品销量。
- 品牌销售报表的开发和展示，通过不同时间的维度（日，月度，季度，年）来分析和展示销量和销售趋势变化情况，在数仓模型中，进行数据处理开发，在 redshift 中进行千万级别大数据量的 sql 处理，对相关的分析查询从表结构和字段设计，指标聚合，逻辑层等多方面进行优化调整，同时配合 procedure，view 的使用，提供高质量的数据模型仓库，供前端 tableau hyper 高可用性的使用。
- 代码重构和性能优化，提升 data pipeline 中 lambda，glue，redshift 的整体稳定性，可维护性和拓展性。
- 参与 devops 的有关建设和维护，Codecommit 的自动化部署，不断完善 devops 和 CI/CD 工作。
- 对整个数据管道和 data pipeline 有比较清楚的认识，同时结合公司的业务需求，维护数据作业调度的正常运行，满足各项目品牌，部门的数据分析、数据应用 BI dashboard 报表的数据需求。

2022/6 知名外资车企数据项目（Volvo）

使用技术：AWS + glue pyspark + airflow + hive + python + hue

项目描述：知名外资车企 APAC 云数据项目，在 AWS 平台上进行数据管道链路的开发，完成数据 ETL 和数仓分层的管道流程。

职责描述：

- 根据公司的 data lake 架构，使用 AWS 的技术产品进行 data pipeline 的整体开发：S3 + Lambda 函数 + glue pyspark 进行 ETL 数据开发，同时配合 Airflow 的任务调度，搭建 data pipeline：ODS——>DW——>DM 的数据链路，在此基础上以供 sagemaker 的模型分析预测应用以及相关的 BI 数据应用。
- 负责核心代码的开发，开发语言为 Python，整合车辆信息和 Operation 数据，将相关数据进行整合包含数据 ETL 和数据分层，分层处理中涉及 udf 函数，主要组件为 Lambda，glue 等，产出的数据应用宽表结果存放于 DM 层供相关部门后续的数据应用和分析。
- 根据业务需求，梳理相关的处理步骤，配置 Airflow 的任务调度，T+1 离线更新 ODS，DW 层数据，搭建好链路，满足内部自动化流程需求。

2021/10 集团企业数据报表开发项目

项目描述：根据集团发展的要求，在财务领域进行数字化转型，监控相关子公司和部门的经营情况，从而提升公司的经营能力和效率。

- 使用 BI 工具对集团层面下的子公司运营情况进行报表系统开发，根据工业部门的运营要求和财务指标计算，对相关数据指标进行加工处理和计算，建立报表的展示，从而分析展示资源利用率，产出率和达成率情况。
- 主要使用 fine report 进行各计算指标的页面展示开发，涉及概览页，明细页等页面的计算指标，图表展现。

2019/7——2021/6 平安集团项目

项目描述：项目主体为集团智能风控系统项目，涉及参与的模块为风险监测板块，通过当前大数据技术的使用，同时结合 AI 模型，跟踪预测宏观经济以及政策趋势。

使用技术：SQL + Hive + Python +机器学习算法

(1) 使用 SQL 或者 Hive 进行数据提取（涉及主要数据库有 MySQL，Oracle，Postgresql）

(2) 使用 Python 进行数据处理和有关的变量处理，主要包括缺失值，异常值处理，变量转换和衍生

(3) 使用线性回归算法和时间序列算法进行预测，同时进行模型的参数调整和预测准确率评估

(4) 使用 Python 进行 NLP 自然语言处理和相关的程序应用

职责描述：

- 作为项目组成员，参与数据应用需求的讨论，落地方案设计，实现业务生产上线的实施。
- 负责核心代码的开发，主要使用 SQL（Hive）+ Python +机器学习算法技术实现 AI 预测的功能，在数据应用层面主要是在 PostgreSQL 中进行数据处理，转换和开发。
- 根据金融行业的特点以及有关合规监管要求，处理结构化和非结构化的数据，包含文书和有关文件文本内容，通过 wind 数据终端，账号接口获取数据源，根据业务规则和需求，设计数据处理，清洗，转换的步骤，建立一套完整的数据处理流程，对有关的数据倾斜情况进行分析和优化，提升 SQL 语句的运行效率，同时根据模型训练的要求和结果，对得到的宽表，建立一定的数据结果质量标准，从而达到较好的预测效果。
- 对 Python 脚本部署应用，生产上线时的 bug 分析和解决；同时对日常运行中的问题进行定位，对现有的模型进行跟踪和维护，及时根据变化进行模型参数调整，参与产品的各种上线测试。
- 根据业务模块从 2 方面（指标和政策内容）进行拆解和分析，结合公司业务，市场情况，金融固收产品收益，分析指标变化大体原因，从而为相关的提前分析预案作参考。
- 定期跟进数据源的采集处理和推送，按照相关的业务规则进行数据监控，维护数据质量，参与元数据管理。
- 撰写相关文档（业务需求方案，模型说明方案）和有关参考分析资料。
- 跨团队的流程沟通协作。

工作经历

公司名称：中软国际

任职时间：2022/8——2023/5

岗位：AWS 数据开发工程师

公司名称：埃森哲企云

任职时间：2021/9——2022/4

岗位：数据开发工程师

责任描述：

1.熟悉数据 ETL 处理流程，数据建模和数仓搭建，BI 报表应用以及机器学习的整体业务架构流程，主要涉及使用阿里云技术产品：dataWorks, Maxcompute odps, dataphin, QuickBI, DataV 等。

2.熟悉数据中台和集市有关的数仓开发方法和流程（离线 or 实时），能根据行业和相关案例进行分析，具备一定的方案设计和规划能力，涉及数据量估计，数据同步，数仓设计和分层，主题域分析和设计，集市应用，报表体系搭建，AI 模型应用部署分析预测等方案。

公司名称：拓保软件有限公司

任职时间：2019/7——2021/9

岗位：数据开发工程师

责任描述：

1.外派平安项目，参与智能风控系统项目的数据开发工作。

公司名称：快钱支付清算
任职时间：2017/11——2018/8
岗位：风险控制模型
责任描述：
1.通过电话或视频对借款人进行征信及反欺诈审核，识别、评估及控制调查风险，出具审核意见。
2.协助团队日常管理，保证作业质量，结合业务场景的模型需求，提供相关的风控数据，发现问题并优化相关操作流程与信贷政策。
3.使用 Python 进行数据分析和整理，开发有关的信贷数据报表。
工作技能：
根据互金产品的特点，熟悉整个信贷运营与操作风险，同时从数据运营的维度分析（作业处理量、通过率、回退率等），保证处理时效和运行机制的有效性。

信贷客户的违约风险预测
项目简介：根据用户画像以及业务的整体运营要求，推动较优客户的引入和风险客户的排查，对信贷客户使用模型进行相应的预测和分析。根据业务特点和要求，结合相应的特征和变量，使用决策树模型进行客户违约风险的分类，同时平衡好相应的响应度，进一步使用随机森林模型提升分类效果。
项目作用与启示：根据客户的风险情况，对客户进行用户画像和信贷风险等级分类，便于后续的进一步客群运营和精细化风险管控，以适应互联网金融行业特点以及合规监管要求。

公司名称：平安普惠
任职时间：2015/9——2017/11
岗位：风控反欺诈
责任描述：
1.对各渠道上报的疑似欺诈案件进行调查分析，根据不同贷款产品类别和流程操作规范，综合分析信息材料，判定案件风险等级，给予相应的欺诈结论。
2.对各类欺诈案件进行分析总结，整理相应业务数据，进行分析总结，辅助提供一定的作业方式思路。
3.深入挖掘欺诈信息，分析团体案件，关联案件以及其他特殊案件，为反欺诈策略和规则提供思路和信息。
关键成果：高质量的作业处理量 1000+，对调查结果负责。
工作技能：灵活应对各类欺诈手段和方法的能力，能够从数据的角度发 现欺诈特点，提供应对策略和建议，同时兼顾产品，营销等各环节，调查结果符合评判标准和依据。

欺诈客户分类
项目简介：针对不同客户群体，根据业务指标和地区情况，在作业环节对客户群体开展分析，提供合理的作业流程和方式建议。根据作业环节的数据汇总，通过分析各类欺诈案件和产品渠道特点，为运营提供数据支持和作业指导
项目作用与启示：通过业务分析和模型预测，使得欺诈客户的特征更加明显，为后续的作业和调查方向提供思路和数据支持

相关证书
- 2021/11　阿里云 ACP 云计算证书