

求职意向：大数据讲师

姓 名: 现 居: 北京
性 别: 男 电 话:
学 历: 本科 邮 箱:

- ✓ 熟悉 Hadoop、Spark 生态系统，掌握各组件工作原理 及 相关集群搭建
- ✓ 熟悉阿里云生态（ECS、MaxCompute、DataWorks 等），持有阿里云 ACP 大数据专业认证
- ✓ 熟练使用 MySQL 、HBase 等数据库 及 数据仓库工具 Hive
- ✓ 熟悉 Flume 、Kafka、Sqoop、Azkaban 等工具的使用
- ✓ 掌握 Java、熟悉 Java EE 开发，熟悉 Scala、Python 等编程语言
- ✓ 熟悉 Linux 操作系统及 Shell 脚本开发
- ✓ 熟悉常用数据结构和算法
- ✓ 熟悉机器学习经典算法 及 相关工具的使用，了解深度学习

| | | |
|---|--------------------|--------|
| 2018.05-至今 | 北京光环致成国际管理咨询股份有限公司 | 大数据讲师 |
| <ul style="list-style-type: none">● 面向全国合作院校的大数据项目实训● 面向全国 4 地的大数据直播班机器学习教学● 负责公司学员的阿里云 ACP 认证培训● 参与腾讯云大数据 TCP 认证的题库研发 | | |
| 2017.03-2018.04 | 成都哒哒哒信息技术有限公司 | 大数据开发 |
| <ul style="list-style-type: none">● 数据采集、ETL 清洗、Hadoop 和 Spark 计算等大数据开发工作 | | |
| 2012.09-2016.12 | 四川德阳久辉游戏工作室 | 脚本研发 |
| <ul style="list-style-type: none">● 研发热门游戏全自动挂机脚本● 开发并部署脚本售卖及验证系统 | | |
| 2010.08-2012.08 | 共青团资阳市委 | 青工部负责人 |
| <ul style="list-style-type: none">● 负责与下属各县市区团委工作对接 | | |

使用 Flume 抽取日志，下沉到 Kafka 中。Kafka 中一个消费组用来做实时统计，另外一个消

费组使用 Flume 下沉到 HDFS, 经 MR 预处理后存入 Hive 表中供后续离线批处理

- 实时统计:
累计用户、累计付费用户、累计付费金额、付费率、新增用户、活跃用户等
- 用户分析:
新增用户、付费用户、活跃用户、流失用户、启动次数、版本分布等
- 留存分析:
从日周月不同时间粒度分析留存率
- 关卡分析:
平均完成时间、进入用户数量、失败率、付费用户、付费金额
- 等级统计:
玩家等级分布、流失玩家等级
- 终端统计:
设备、网络、运营商、地域

项目名称: 客户流失预警

项目背景:

随着移动互联网和物联网时代的来临, 客户可以选择的应用提供商也越来越多。对于现代企业来说, 保障所有客户特别是高价值客户的个性化需求、维系相对稳定的客户群体, 是非常重要的工作。因此需要事先对客户的流失行为进行预测和报警, 从而提醒公司相关部门事先做好客户关怀工作, 减少客户流失。

技术选型:

sk-learn、pandas、numpy、matplotlib

项目职责:

- 特征工程
 - 1) 数据的加载: 使用 pandas 包的 read_csv 进行读取, 定义 特征处理函数
 - 2) 数据预处理: 对数据使用 info 和 describe 等进行整体查看, 考虑每一个特征列的情况, 针对不同的情况, 对于数据的缺失值、异常 值以及重复值进行处理
 - 3) 特征处理: 主要有 LabelEncoder 编码, OneHotEncoder 编码处理、归一化处理等
 - 4) 特征选择: 使用随机森林 (RF) 模型, 得到 feature_importances_, 剔除重要性低于均值的特征
 - 5) 划分训练集与测试集
- 参数优化
使用随机森林 (RF) 模型, 通过 GridSearchCV 方法验证随机森林 (RF) 的最佳 n_estimators 和 max_features 参数并且对模型进行调试得到 best_score_和 best_params_

教育背景

2006.09-2010.07

四川师范大学

物理学

学士

自我评价

师范专业出身, 热爱教师职业, 授课经验丰富, 多年程序开发经验, 善于分析解决问题, 有很强的自学能力、良好的团队精神及强烈的责任心