

姜博文

☎ 15776542286 ✉ 15776542286@163.com 📍 杭州
👤 实习中 📁 数据分析、数据挖掘



教育经历

浙江理工大学 - 计算机科学与技术 硕士 计算机学院 2020.09 - 2023.06
校一等奖学金 (专业前10%)、数模大赛三等奖、优秀班干部、剑桥大学暑期实践项目推荐信。

东北石油大学 - 资源勘查工程 本科 地球科学学院 2015.09 - 2019.06
专业GPA : 3.45 / 4.00、校优秀毕业生、Coursera 结业证书、CAD 高级绘图工程师。

实习经历

字节跳动 - 电商数据科学实习生 2022.04 - 2022.09
工作内容：流量侧商达指标体系建设、流量归因分析、推荐模型纠偏、推荐-广告人群画像、商家成长规划等。

滴滴出行 - 货运决策分析部实习生 2021.11 - 2022.04
工作内容：数据化运营业务和产品规则、司机侧数据指标设计与完善、活动策略评估分析、司机流失预警等。

上海准确技术有限公司 - 数据工程师 2020.09 - 2021.09
工作内容：数据开发、数据爬虫、数据可视化、统计分析、机器学习、深度学习、商业分析等

项目经历

抖音电商 推荐模型纠偏项目 2022.07 - 2022.09
项目描述：

当前推荐页短视频精排链路，用户触达到店铺广告后在自然侧相同的店铺7d内发生转化的模型预估值存在稳定低估现象，采用pava 算法纠正部分视频在“精排阶段”模型的低估，提升千川短视频团队关注的指标，使得投广商家在自然推荐侧得到合理的自然流量，营造对投广商家友好的自然经营环境。

项目内容：

1. 离线数据准备：构建广告触达用户在推荐侧发生支付场景下的店铺视频的预估、后验转化数据，构造出PAVA的训练底表。
2. PAVA算法实现：根据训练底表，使用Spark的定时任务训练模型，将纠偏系数透传到算法团队进行实时纠偏。

抖音电商 商家成长项目 2022.04 - 2022.07
项目描述：

针对抖音电商传统商家分型进行优化，探寻商家真实成长路径，弄清头部商家的流量破圈办法，给予中小商家成长建议，反推产品调整推荐模型优化思路，上线长周期成长任务及流量入口经营攻略，提升长期 GMV。

项目内容：

1. 数据建设：牵头数仓进行埋点治理，配合业务设计流量侧行业、商家、达人的指标体系与归因体系，搭建大盘流量归因看板(日均观看人数 6000+)。
2. 评估分析：完成抖音电商行业现状摸底，针对二奢行业直播间特性，配合算法与产品在大促前上线AB实验评估策略，GPM 提高 0.54%，GMV提升0.37%。
3. 建模分析：基于商家历史流量表现，构建聚类模型进行业务解读，完成商家成长阶段分型，摸清成长阶段的卡点问题，找出头部商家阶段跃迁周期与破圈办法，配合运营上线长周期商家成长任务与商家流量入口运营攻略。

滴滴货运 司机流失预警项目 2021.11 - 2022.04
项目描述：

针对货运线运力不足，调度效率较低的情况，构建司机服务水平、运力效率与留存情况核心指标，完善司机生命周期建设，配合业务进行实时监测和预警，实现司机精细化运营，提高司机留存率。

项目内容：

1. 数据建设:基于现有数据，优化表逻辑，固化业务口径，完善运力侧分级指标体系；搭建日报、周报 看板，配合运力、战略、产品同学实现运力指标的实时监测；
2. 评估分析：对司机生命周期进行探索性分析，定位 分层阈值，产出司机效率画像;设计 AB 实验评估策略影响，产出上海专项关于司机分级动调的评估报告，协助产 品调整优化。
3. 建模分析：搭建双向递归神经网络(BiLSTM)模型，预测周度流失司机，上线司机流失预警机制与挽 回策略，设计 PSM-DID实验评估策略效果，截至目前预警 3.3w、挽回 0.9w 人，提高 司机留存率 2.47%。

英国剑桥大学暑期实践项目

2020.12 - 2022.02

项目描述:

深度学习中各激活函数探索与研究项目。

项目内容:

比对不同激活函数的优缺点，探究适用条件及应用场景，作为 组长带领成员完成实验代码并撰写论文，获得个人 86 分的项目成绩，并获得项目推荐书。

项目开源地址: <https://github.com/jiangbowen-available/cu-Activation-Functions>

基于图自注意力机制的蛋白质扭转角预测

2020.09 - 2021.09

项目描述:

使用深度学习图神经网络与自我注意力机制预测蛋白质空间结构信息。

项目内容:

将 DNA 序列与蛋白质结构数据库进行多序列比对，计算位置特异性评分矩阵 pssm、隐马尔可夫状态 转移量 hhm、kl 散度、wasserstein 距离等统计势能，利用图神经网络与自注意力机制对蛋白质 二维结构的空 间表 示 进行训练与预测，测试 AUC 相较于最新文章提高 2.6%。

项目开源地址: https://github.com/jiangbowen-available/Turning_angle

闻道教育 离线数仓项目

2021.04 - 2021.06

项目描述:

针对网上教育平台的用户信息建立基于Spark的大数据分析平台，采集业务数据与用户点击日志数据，进行处理并存储ODS层数据;依据业务形式聚合成ADS层数据，提供业务查询与分析功能;对 接可视化平台进行离线数据展示。

项目内容:

1. 进行数据清洗与存储。将Kettle采集MySQL中的课程、订单、用户等业务数据、Flume收集Nginx采集的用户日志数据，使用SparkSQL处理后，存储至ODS数据库中。
2. 生成ADS业务指标。将事实表与核心维度表进行关联聚合，依据业务形态生成对应的DW与ADS数据表。
3. 提供可视化接口。使用Kylin对接可视化平台实现离线数据展示,并提供用户行为画像核心看板。

专业技能

1. 熟练掌握 python 数据处理与数据分析工具，熟练使用sql工具。
2. 熟练掌握机器学习监督与非监督算法、与深度学习 tensorflow、pytorch 框架与主流算法模型。
3. 熟悉Spark、hadoop、hive大数据开发框架，具备离线大数据开发经验。
4. 熟悉爬虫技术，并具备一定程度的，Django、Flask 的 web开发经验。
5. 了解 PowerBI 和 Tableau 等数据可视化工具，可以使用 spss、stata 等集成分析工具。