**李孝川** 计算机科学与技术

电话: 15021205204 出生年月:1993.12.2

邮箱: [942727923@qq.com](mailto:71215901067@stu.ecnu.edu.cn)  求职意向: 数据挖掘



**工作经历**

北京恒泰实达科技股份有限公司  数据挖掘工程师 2018.10—2022.06

涉及业务: 智能大屏可视化集成项目，产品包含系统硬件，VBI5和数智云图等软件系统产品。

主要职责:

1. 综合运用机器学习、数据挖掘算法结合项目需求和业务场景。

2. 参与电力系统业务指标建模预测，完成智能大屏可视化产品的功能模块开发。

3. 采集数据进行业务处理，挖掘探索人工智能在电力系统中的落地场景。



**专业技能**

1. 熟悉常见机器学习算法基本原理(LR、DT、SVM、GBDT、XGBoost、KNN、RF)

2. 熟悉Python、C/C++、sklearn库、Arcmap的常规使用，关系型数据库MySQL的基本使用

3. 了解Spark、Hadoop基本知识

4. 了解深度学习基本原理，用Pytorch做过简单的CV领域的demo案例



**项目经历**

**贷款违约预测（天池大赛）**  2022.10 — 2022.11

项目描述：

赛题以金融风控中的个人信贷为背景，要求选手根据贷款申请人的数据信息预测其是否有违约的可能，以此判断是否通过此项贷款。

负责内容：

1. 特征预处理：包括缺失值填充、特征编码（label encoding、mean encoding）等
2. 特征构造：比例特征（X1/X2）、用groupby将类别特征与数值特征的组合（如：median(X1)\_by(X2)）、交叉特征（X1 \* X2）、计数编码（count(label)）等
3. 特征筛选：根据Xgboost输出feature\_importances\_,结合方差等方法，过滤部分特征，提高模型泛化能力
4. 超参数调优：采用Optuna进行超参数调优

项目成果：

1. 提交测评分数，赛题的评估指标是AUC，排名在

**上海电力台区线损分析治理**  2020.10 — 2021.11

项目描述：

电网企业把线损率作为重要衡量指标，项目子任务的主要目的是分析建模判断台区线路损耗是否出现异常。为相关部门实地排查检修提供有效指导依据

负责内容：

1. 由PMS2.0系统采集台区线路脱敏数据30万条
2. 负责前期数据预处理（缺失值填充、特征编码等）、创造新特征、筛选提取数据集中的重要特征等
3. LR、XGBoost等多个算法分别建模训练， GridSearchCV接口进行参数调优，绘制模型评估PR、ROC曲线
4. 交叉验证，反复迭代优化特征和模型参数

项目成果：

1. 通过多次特征工程和交叉验证模型参数调优，最终以XGBoost算法建模，模型precision达0.93，recall为0.4
2. 综合权衡电网系统的其他指标，如供电可靠率、平均停电次数。选择较高精确率的阈值，保证了供电系统的供电可靠性不会因为频繁停电检修而有所降低，上线后经汇报被客户认可

**城市大脑智慧能源****浦电云**  2019.12 — 2020.07

项目描述：

浦东供电公司定制智慧能源浦电云项目，主要目的是从电力能源角度协助城市运管中心治理浦东，力求电力信息与政务信息的交融，提升城市治理水平

负责内容：

1. 负责与客户协调数据的获取，以及沟通评估指标，“群租风险指数”(F1 score)和“配电网健康指数”(Accuracy)
2. 采用多种方式进行数据合理填充和编码，提高数据质量、创造新特征、筛选提取数据集中的重要特征等
3. 对两个业务指数进行分析建模，电网故障类型数据可视化

项目成果：

1. 采用Random Forest算法对"配电网健康指数"进行多分类预测（5类），Accuracy最高达0.92，有效提升浦东供电公司在协助城市治理中的决策效率
2. 利用Xgboost算法得出概率数值，首次将浦电云应用于群租房治理，在数据的支撑下，大幅提升执法人员判定群租房的速度

**浦东供服指挥中心数据融合与治理** 2019.09 — 2020.05

项目描述：

数据融合治理项目的其中一项子任务是分析工业园区和普通居民存在窃漏电用户的关键特征，构建窃漏电用户的模型，利用智能电表收集的数据，对电力用户进行判别

负责内容：

1. 数据集近5w条数据，负责处理样本分布不均匀的问题，数据预处理、模型构建和训练
2. 跟踪实际上线后的效果反馈，优化迭代机器学习模型

项目成果：

1. 数据集使用过采样SMOTE算法和欠采样两种方式建模的recall指标对比，根据和客户沟通结果，确定采用人工增加小部分正样本的方案
2. 综合评估训练的多个二分类学习器，最终采用Random Forest算法建模
3. 识别用户窃电行为的recall达0.80，AUC在0.75左右
4. 模型能有效识别出存在窃漏电潜在风险的用户，顺利通过评估指标汇报和模型上线验收

**市区供服GIS数据分析**  2018.10 — 2019.02

项目描述：

项目目的是希望采用GIS地图的形式展示配电网设备、配电网架空线、数字化网格等数据，利用空间可视化数据分析，有效支撑市区供服指挥工作

负责内容：

1. 预处理8万多条用电地址数据，并采用高德地图API+Python脚本批量转换地址为WGS84坐标
2. 使用GooSeeker工具给用电地址数据打上地理位置标签，然后做标签频率统计
3. 在PMS2.0系统中采集原始数据，用Arcmap、LSV制作架空线、区域矢量图层

项目成果：

1. 总计得到600多个小区/街道标签，并制成20个shapefile文件
2. 最终聚合分析得到符合标准的31个“电网类故障”和19个“非电网类故障”的小区/街道及发生频次
3. 经转为WGS84坐标后可视化展示，大幅度提升市区供电公司的配网线路检修排查效率，同时有助于协助相关部门打击偷电、私接电线等存安全隐患的行为



**教育经历**

华东师范大学 硕士 计算机技术 （在读） 2021.09—2024.07

许昌学院 本科 计算机科学技术 2013.09—2017.07