



年龄：21岁 民族：汉族 手机：13597704421 邮箱：feitanyuyida@126.com 现居地：江苏省南京市

教育背景

南京大学 **硕士** **图书情报** **2023-2026**

科研方向：自然语言处理与知识图谱

专业排名前 40%

华中师范大学 **本科** **信息管理与信息系统** **2019-2023**

专业成绩前 20%

主修课程：数据挖掘、人工智能、信息分析与预测、经济学、管理学、用户行为数据分析、数据结构、计算机网络、数据库原理与开发、操作系统、Java、信息组织与检索、信息系统建模

实习经历

算法组数据分析实习生 **满帮集团** **2023.11 – 2024.1**

- 数据分析** 算法模型效果 AB 实验观测分析；finB 看板核心指标监测；问题排查；进线分析；算法周报更新、解读
- 数据基建** AB 实验数据看板搭建、流程自动化；埋点设计、埋点解析、数仓落地；AB 实验配置；hive 调度优化

项目经历

科大讯飞比赛项目 **水泵状态监测与故障诊断** **2023.10**

- 探索性数据分析** 对数据集的基本特征进行分析，如样本个数、特征维度、特征含义与数据类型、缺失情况及基本统计量等；对全部特征进行数据分布可视化和不同 y 值上的分布可视化；使用箱型图进行异常值分析
- 特征工程** 通过 EDA 得到有效特征列并进行相应清洗；使用均值填充对缺失值进行处理；对有明显指向结果的数值特征进行数据分桶；对部分离散数值型特征进行十折分层目标编码
- 模型训练** 使用逻辑回归、支持向量机、随机森林、XGBoost、LightGBM、CatBoost 模型以及异常检测算法 IForest 进行对比实验，最后采用 XGBoost、LightGBM、CatBoost 进行模型加权融合，使用折外预测 oof 验证模型质量，评价标准采用 F1 score 指标，最终模型在测试集上效果达到 99.596%

毕业论文 **细粒度小说内容主题挖掘** **2023.02-2023.04**

- 数据采集与存储** 使用爬虫技术对图书网站上小说 100 条内容数据和 30000 条评论数据进行采集，并使用 mysql 数据库存储；
- 数据清洗与特征工程** 使用 python 对文本数据进行预处理，包括数据加工和清洗；
- 模型训练** 使用机器学习中的主题模型及情感分析技术对数据进行细粒度的挖掘和分类

专业技能

- 对风控和数据挖掘有强烈兴趣，对数据挖掘有一定了解，熟悉 Python 及 sql 等常见数据处理工具；
- 了解常用的机器学习算法和技术；会基本的自然语言处理，有一定的深度学习基础。
- 会简单使用 R 语言和 Java 及 linux 系统，会使用爬虫技术；
- 具备良好的英文书面表达能力，流利使用英文，无障碍阅读英文文本；
- 学习能力强，上手速度快，认真负责，沟通能力强；

相关证书及奖项

- 全国大学生英语竞赛三等奖
- 英语六级
- 数据库系统工程师证书
- 计算机二级
- 信息技术高级教师资格证