

## 李晓颖

女 | 25 | 现居昆明

意向郑州 | 5天/周 | 3-6个月 | 2周内到岗

联系电话：+86 13526486887 | 联系邮箱：2812524174@qq.com



### 教育经历

2019/09 ~ 2022/06 | 昆明理工大学 | 管理科学与工程 | 硕士

绩点：3.6

主修课程：数据库与数据挖掘

2014/09 ~ 2018/06 | 中原工学院 | 信息管理与信息系统 | 本科

绩点：3.6

主修课程：数据库，数据结构，管理信息系统

### 学术经历

2020-10 ~ 2021-08 | 负责人 | PPP项目社会资本参与度预测分析

1.利用爬虫技术获取CPPPC项目库中的相关项目信息，共获取项目数据9800条；2.使用numpy, pandas对数据进行预处理，主要包括异常值处理（经验值+箱图），缺失值填充（类内均值或众数），数据标准化（Z-score），并根据文献以及相关经验进行特征构造，最终保留4258条数据，90维特征变量；3.使用ADASYN算法对非平衡数据进行处理，提高预测模型精确度。之后，对比分析随机森林、XGBoost、LightGBM三种模型的预测准确性，对比发现，ADASYN+XGBoost表现出了最佳预测性能，AUC值达到95.75%，但是不同模型之间的差别不大，仅存在2%-5%的差距。

2020-05 ~ 2020-10 | 文章收集、识别；关键字提取、聚类、组词 | 大数据驱动信息基础设施PPP可融资性影响因素获取及评价方法研究

1.通过公众号抓取软件获取PPP领域5个权威公众号近5年的所有发文，根据知网PPP相关文献的关键字以及PPP网站发布的PPP专业术语等构建用户字典，并根据经验完善停用词表。利用停用词表以及用户字典对文章题目进行分词，并利用分词结果识别与PPP相关的文章，最终识别出的PPP文章每年大概2000条左右；2.利用停用词表以及用户字典对筛选过后的文章进行分词，根据关键字“融资”、“落地”等在每篇文章中提取20个关键字，分年份合并当年的所有关键字。合并后的关键词去重并删除出现频次低的关键字，每年保留3000-5000的关键字；对保留下来的关键词进行聚类，簇数限定在10以内。对簇内关键字进行正反组词，组词后的结果在原文章中进行搜索并记录出现频次，删除频次低的词语，每年保留1000-3000的组词结果；3.最终结果：对组词结果进行进一步处理之后，得到包括政策环境、法律法规、政府监管等在内的20个核心指标作为PPP项目

### 获奖经历

2020/01 | 第十三届“认证杯”数学中国数学建模网络挑战赛

2020/12 | “华为杯”第十七届中国研究生数学建模竞赛

### 技能/语言

英语 CET-6

数据分析

熟练

数据挖掘

熟练

数据建模

熟练

SQL

熟练

Python

熟练

仅限本招聘账户所属公司内部职位使用，未经许可转载、转售、公开等，均需承担法律责任！