



曹颖

13645535413 | bnu_ying124@163.com
23岁 | 170cm | 53kg | 汉族
数据科学 | 数据挖掘



北京师范大学
BEIJING NORMAL UNIVERSITY

教育经历

- 北京师范大学 | 统计学院 | 应用统计 (硕士) 985 2025.09 - 至今
- 核心课程：应用多元统计分析、多元回归分析、机器学习、数据挖掘、深度学习、R语言数据可视化
 - 研究方向：机器学习在社会科学和教育数据中的应用、因果机器学习
- 安徽大学 | 数学科学学院 | 数学与应用数学 (本科) 211 2021.09 - 2025.06
- 数学基础：数学基础：高等代数、数学分析、概率论与数理统计、数值分析
 - 荣誉：校级优秀团干、社会实践先进个人、校运动会女子100米决赛冠军/1500米决赛亚军、校级羽毛球女子单打比赛季军

专业技能

- 数据分析与建模**：熟练使用 R / Python 进行数据清洗、探索性分析、统计建模与结果可视化；掌握回归分析、多元统计分析、基础机器学习模型 (LR、RF、GBDT 等) 的建模逻辑与评估方法；能基于业务问题构建分析指标体系，并将模型结果转化为可解释结论。
- 数据挖掘与机器学习**：具备从非结构化或复杂数据中提取特征并建模的经验；理解模型训练、验证、误差分析与调优的基本流程；对可解释性分析 (如特征重要性、模型行为分析) 有实际应用经验。
- 语言水平**：CET-6 (560)，可熟练阅读英文技术文献，擅长技术文档与分析报告撰写，能够在算法、产品与业务之间进行有效沟通。

实习经历

- AI专业垂类实习生 | 北京字节跳动科技有限公司 | SFT-知识问答团队 2023.12 - 2024.03
- 深度参与字节旗下大模型产品「豆包」的数学知识问答系统迭代，负责数学领域数据构建与有监督微调 (SFT) 针对模型在复杂数学问题中存在的推理断裂与表达不准确问题，设计高质量标注样本，推动模型回答准确率提升约 **18%**；
 - 基于 SQL 提取用户反馈与模型输出数据，构建多维评估指标，辅助分析模型错误模式与能力边界；
 - 撰写中英文技术说明文档，与算法、产品团队协作完成 3 次版本迭代，推动模型在教育与知识问答场景中的落地应用。
- 行业研究实习生 | 国泰君安证券股份有限公司 | 计算机团队 2023.04 - 2023.10
- 围绕人工智能产业链，系统整理并分析算力、应用与底层技术相关数据，构建 AI 行业研究分析框架；
 - 负责企业财务与行业数据的收集、清洗与对比分析，参与多篇行业研究报告的撰写；
 - 独立编写 AI 研究框架与科普手册，发表于团队公众号，提升内部研究效率并辅助决策支持。

项目经历

- 大学生创新创业大赛 | 项目负责人 2023.06 - 2023.10
- 构建符合中国国情的农民数字素养评价框架，围绕农业数字化背景下的能力测度问题开展建模研究；
 - 负责数据整合、指标体系设计与模型构建，采用多元线性回归与中介效应分析探讨数字素养对农业行为的影响路径；
 - 结合实地调研数据对模型进行修正与验证，提升分析结果的现实解释力；
 - 项目顺利结题并获得国家级立项。
- 银行营销数据挖掘与客户转化预测
- 数据清洗与EDA：处理某银行电话营销数据集 (包含客户画像、经济指标等特征)，使用 R 语言进行缺失值插补与异常值处理；绘制云雨图与马赛克图揭示“通话时长”与“订阅成功率”的非线性关系；
 - 建模与调优：构建逻辑回归与随机森林模型预测客户是否订阅定期存款。通过网格搜索 (Grid Search) 优化超参数，利用 SMOTE 技术解决样本不平衡问题。

内容创作能力

- 独立运营小红书账号，围绕自我成长与专业知识输出，累计 **20w+** 浏览、**2w+** 点赞收藏；
- 具备较强的信息结构化、表达与传播能力，能够将复杂内容转化为易理解的形式。

个人画像

- 具备扎实的数学与统计学背景，能够从数据与模型视角理解复杂问题；
- 对数据分析与数据科学具有长期投入意愿，学习能力强，成长曲线明确；
- 兼具技术理解与表达能力，能够在多角色团队中承担连接与推进作用。