

DeepSeek在数据挖掘中的应用

—— 从商业经典到智能前沿的全场景实践

汇报人：廖小青



目录

01

引言：数据挖掘的
价值与DeepSeek
破局

02

行业实战：三大领
域的深度数据掘金

03

实操秘籍：精准提
问与高效分析的七
大心法

04

DeepSeek的优势
与挑战

01

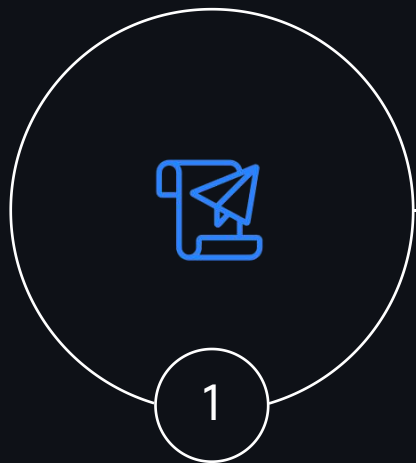
引言：数据挖掘的价值与 DeepSeek破局



啤酒与尿布有什么关系？



数据挖掘的重要性 —— 从啤酒尿布到数据金矿



数据时代，数据总量爆发式增长

据IDC 2023年数据，全球数据总量2025年将达175 ZB，但企业数据利用率仅15%，大量数据处于闲置状态，形成“数据沉默”困境。



传统分析仅能看到表面数据，难以发现“父亲买尿布时高频购啤酒”这一隐性关联。

DeepSeek介绍 —— 重新定义数据挖掘的智能引擎

01

DeepSeek定位于零代码智能平台

用户仅需3步即可完成数据挖掘全流程：上传数据，以自然语言提问，获取可视化报告，大大降低了数据挖掘的技术门槛。

02

在企业端，可助力电商精准选品

通过分析市场数据和消费者偏好，选出具有潜力的商品；有效防控金融风险，实时监测交易数据，识别潜在风险；还能对客户进行精细分群，实现精准营销。

03

在学术端，能深度分析医疗数据

辅助疾病诊断和药物研发；对教育学情进行精准诊断，帮助教师制定个性化教学方案。

DeepSeek vs 传统工具 —— 效率革命的三维对比



数据处理深度

传统搜索引擎和爬虫只能进行网页级抓取，获取表面数据；传统数据挖掘工具主要针对结构化数据进行统计分析。而DeepSeek具备全模态语义挖掘能力，不仅能处理文本，还能对情感分析、图像特征、时间序列预测等多模态数据进行深度挖掘，发现数据间的潜在联系。



操作门槛

使用传统工具，需掌握Python等编程语言以及精通复杂的算法调参，对使用者的技术要求高。DeepSeek采用自然语言对话式交互，用户只需像日常交流一样提问，如“分析某课程差评中的教学问题”，即可轻松获取分析结果。



输出价值

传统工具的输出多为原始数据堆砌或技术指标报表，难以直接应用于业务决策。DeepSeek则能提供业务可落地建议，例如根据数据分析结果，直接给出“建议增加XX实验课时以提升通过率”等切实可行的方案。

02

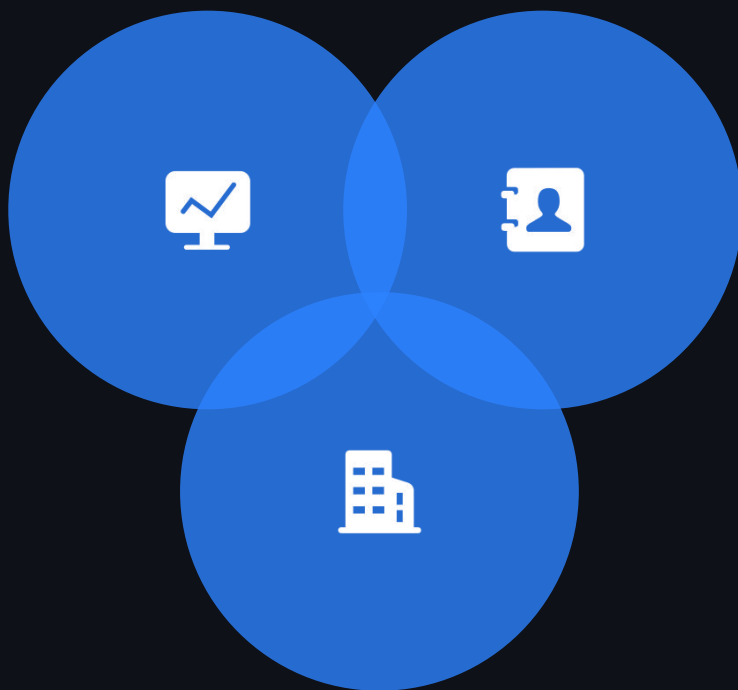
行业实战：三大领域的深度 数据掘金



金融行业 —— 风险防控与客户价值双引擎

商行信用卡欺诈检测案例

传统规则引擎在检测信用卡欺诈时，漏检率高达18%，且人工复核需耗费大量人力和时间，效率低下。



特征提取

DeepSeek方案通过自动提取200 + 交易特征，如设备指纹、地理位置偏移量、夜间交易频次等，全面刻画交易行为；并动态生成集成模型，融合随机森林和LSTM算法，实时计算欺诈概率。

最终成果显著

漏检率降至5%，2023年成功拦截欺诈交易超3000万元，风险成本下降40%，有效保障了银行资金安全和客户利益。

电商行业 —— 从「人找货」到「货找人」的智能进化



某跨境电商构建精准选品系统

操作流程为：首先上传10万 + 亚马逊商品评论，评论中包含文本和图片等多模态数据；DeepSeek自动识别高频痛点，如发现“充电接口接触不良”出现频次达28%。



根据分析结果，DeepSeek输出选品建议

如开发“磁吸式接口 + 可调节尺码”改良款产品。



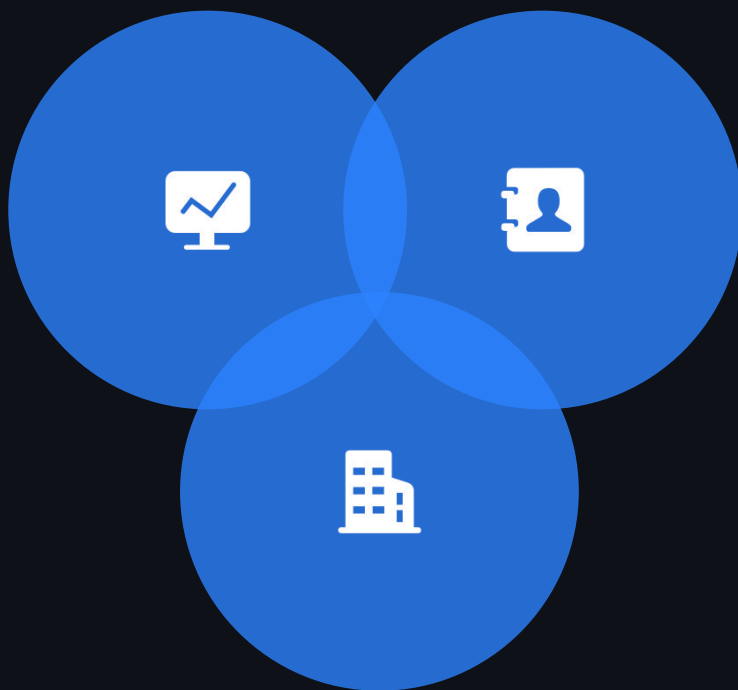
方案带来巨大商业价值

新品上市3个月销量破百万，退货率下降45%，同时选品周期从4周大幅缩短至3天，快速响应市场需求，提升了电商企业的竞争力。

医疗行业 —— 数据驱动精准医疗实践

某三甲医院

在分析糖尿病并发症时，面临电子病历数据分散、传统分析难以发现隐性关联的痛点，影响医生对病情的准确判断和治疗方案的制定。



DeepSeek方案自动提取

自动提取10万+病历中的诊断关键词，构建并发症知识图谱，以直观的方式展示疾病与症状、治疗方法等之间的关系；通过关联分析，发现“餐后血糖波动 $> 2\text{mmol/L}$ ”与视网膜病变存在强关联，关联强度OR值 = 3.2。

成果具有重要临床价值

医生科研效率提升80%，能更快地发现研究方向和突破点；患者并发症发生率下降22%，有效改善了患者的治疗效果和生活质量。

03

实操秘籍：精准提问与高效 分析的七大心法



精准提问七大心法 —— 让 DeepSeek秒懂需求

角色扮演法

明确自身在分析中的身份定位，例如“作为班主任，分析班级数学成绩波动原因”，使DeepSeek能从特定角度出发，提供更具针对性的分析结果。

知识蒸馏法

从复杂繁多的数据中提炼核心需求，如“从300页会议纪要提取客户投诉TOP3”，帮助用户快速获取关键信息，避免被大量无关数据干扰。

颗粒度调节法

根据分析目的，灵活控制分析精度，如“按「城市 - 商圈 - 门店」三级维度分析销售额”，可从宏观到微观全面了解业务情况。

时间轴推演法

加入时间维度，分析数据随时间的变化趋势，如“对比近三年Q3促销活动的ROI变化趋势”，有助于发现业务发展规律，为决策提供历史数据参考。



精准提问七大心法 —— 让DeepSeek秒懂需求



极端测试法

设定边界条件，挖掘特殊情况下的数据特征，如“找出客单价 > 2000元且复购率 < 5%的客户特征”，帮助企业关注高价值且特殊的客户群体。



逆向思维法

反向思考问题，挖掘隐藏在数据背后的信息，如“哪些客户特征与「从未购买」行为强相关？”，为企业拓展业务提供新的思路。



情景模拟法

预设业务场景，提前规划应对策略，如“假设预算减少20%，推荐最优促销组合方案”，帮助企业在面对不同情况时做好充分准备。

用批判性思考深入挖掘复杂问题

01

数据可靠性

仔细检查数据中的缺失值和异常值占比，例如当发现“用户年龄字段缺失20%”时，思考其对分析结果可能产生的影响，判断是否需要进行数据补充或处理，以确保分析基于可靠的数据基础。

02

逻辑严谨性

严格区分相关性与因果性，不能仅仅因为“促销活动与销量增长”在数据上呈现相关关系，就简单认定它们存在因果关系，需深入分析其他可能影响销量的因素，避免错误决策。

03

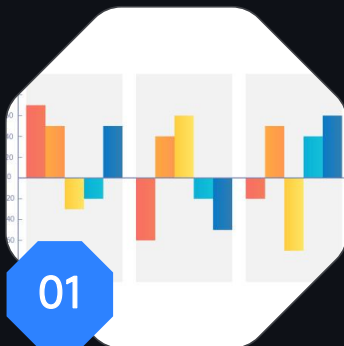
业务落地性

将技术指标转化为通俗易懂的业务语言，例如把“模型准确率90%”解释为“每100个预测仅10个误判”，使业务人员能够更好地理解和应用数据分析结果，推动业务发展。

可视化输出 —— 让数据「会讲故事」的三大技巧

图表选择黄金法则

对于不同类的数据采用不同的输出方式
(例：趋势类数据→折线图，占比类→南丁格尔玫瑰图)



图表选择黄金法则

z



动态交互设计

利用DeepSeek算法，公司能够识别出社交网络中具有较大影响力的用户，以优化营销策略。

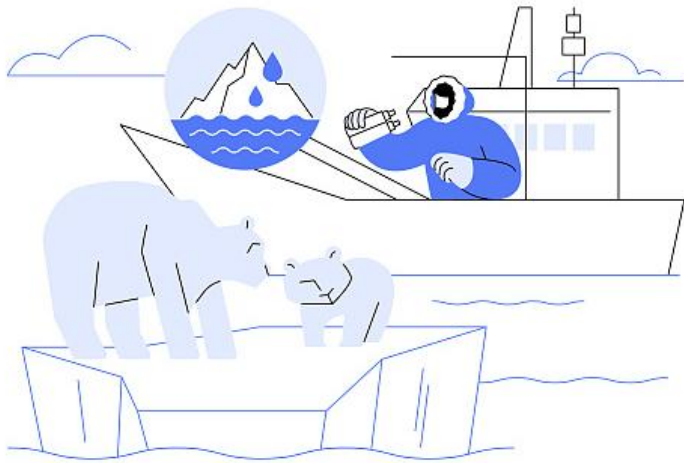
04

DeepSeek的优势与挑战





CORAL REEF BLEACHING



ICE MELTING



BIODIVERSITY PROTECTION

技术优势分析

高效的数据处理能力

DeepSeek通过先进的算法优化，能够快速处理大量数据，提高数据挖掘的效率。

精准的模式识别

利用深度学习技术，DeepSeek在模式识别方面表现出色，能够准确识别复杂数据中的关键信息。

成本优势

DeepSeek具备自适应学习能力，能够根据数据变化自我调整，持续优化挖掘结果。

面临的主要挑战

数据质量“垃圾进垃圾出”



建立严格的数据质检清单，要求数据完整性 $\geq 90\%$ ，异常值占比 $\leq 8\%$ ，从源头保证数据质量；利用 DeepSeek 自身的数据校验功能，自动标记可疑数据，并支持批量修复，提高数据的准确性和可用性。



隐私安全与合规风险

对敏感数据进行自动脱敏处理，如将手机号、身份证号模糊化为“138****1234”，保护用户隐私；针对医疗、金融等对数据安全要求高的行业，提供本地化部署方案，确保数据不出域，满足合规要求。

给未来数据分析师的寄语

数据挖掘的本质是将业务需求转化为技术语言

感谢聆听！

期待与同学们共同开启数据挖掘的智能之旅！



扫码下载本课件和

《DeepSeek：从入门到精通》

操作手册