

《大数据技术基础》课程

关于大模型助力大数据分析案例的实施细则

为鼓励选课学生积极参加大模型助力大数据分析案例分享，对完成研究实践及课堂分享的小组及个人予以加分认定，特制定本实施细则。

一、分享形式

提交研究实践报告以及源代码，并完成 **15 分钟以内** 的课堂分享（以PPT形式）。研究实践报告需严格按照 **附表2** 的模板书写，在报告或 PPT都需要配以适当截图辅助说明，**做到图文并茂**。

二、分享主题

本次案例分享聚焦于**医疗与金融两大领域**，同学们可以围绕这些领域自由发挥，探索具体的应用场景。通过使用Cursor、WindSurf等智能编程工具，以及与大模型交互得到对数据的文字和可视化分析，体验大模型助力大数据分析的过程。

三、分享内容 & 分值额度

在给定的分享主题下自拟题目，各部分内容以及对应分值额度参见下表，**总分为20分**。

各部分内容	各部分分值额度
爬虫数据采集	根据质量，分值在 5-7 分
数据分析	根据质量，分值在 1-3 分
数据可视化（鼓励使用第2种方式v0.dev平台实现可视化）	根据质量，分值在 4-6 分
整体分析	根据质量，分值在 2-4 分

给分细则：

1. 爬虫数据采集（5-7分）

评分维度：数据源（爬取的目标网站）质量、数据完整性、代码规范性、工具使用情况

● 7分：

- 数据源：数据源权威且丰富，覆盖多个相关领域或多个时间维度（如多年度数据、多地区数据等）。
- 数据量：采集的数据量大，能够满足后续分析和挖掘的需求。
- 代码规范性：代码逻辑清晰，注释详细，遵循编程规范，易于理解和维护。
- 工具使用：使用了智能编程工具（如Cursor、WindSurf等）进行代码撰写和调试，代码高效且无明显错误。

- 实践报告：在实践报告中明确给出了源网站链接，并对数据采集过程进行了详细说明，包括数据源的选择理由、爬取策略、数据清洗和预处理的具体步骤。

● 6分：

- 数据源较为丰富，但可能在权威性或多样性上略有不足。
- 数据采集基本完整，但可能存在少量缺失或需要进一步修正的地方。
- 使用了智能编程工具，代码较为规范，但可能在逻辑上有小瑕疵。
- 数据清洗和预处理基本完成，数据质量良好。
- 实践报告：在实践报告中给出了源网站链接，但对数据采集过程的说明不够详细，部分步骤可能需要进一步补充。

● 5分：

- 数据源单一或权威性不足。
- 数据采集存在明显缺失或错误，但进行了必要的清洗和预处理。
- 未使用智能编程工具，代码存在较多问题（如逻辑混乱、效率低）。
- 数据质量一般，可能需要进一步优化才能用于后续分析。
- 实践报告：源网站实际页面解析与代码不匹配，或对数据采集过程的说明过于简略，缺乏关键信息。

（注：缺少数据源网站、采集代码、数据结果、实践报告截图中任意一项，会在5分基础上依次递减）

2. 数据分析（1-3分）

评分维度：分析方法的多样性、分析结果的准确性、分析的深度

● 3分：

- 综合对数据进行多角度分析（如描述性统计、时间序列分析、相关性分析、聚类分析等），至少涵盖4种以上。
- 分析结果准确，能够清晰地展示数据特征和趋势。
- 分析具有深度，能够从数据中挖掘出有价值的见解或发现潜在问题。

● 2分：

- 使用了2-3种分析角度，但可能在某些角度的运用上不够深入或准确。
- 分析结果基本准确，能够展示部分数据特征，但缺乏深度。
- 分析角度之间缺乏逻辑连贯性，未形成完整的分析体系。

● 1分：

- 只使用了1-2种分析角度，且角度选择不合理或运用不准确。
- 分析结果存在明显错误或无法有效展示数据特征。
- 分析过程缺乏逻辑性，未对数据进行深入挖掘。

3. 数据可视化（4-6分）

评分维度：图表类型多样性、维度指标合理性、交互功能丰富性、可视化效果

● 6分：

- 包含了3-5种不同类型的图表（如柱状图、折线图、饼图、散点图、地图等），图表选择合理，能够清晰地展示数据特征。
- 维度指标设计合理，能够从多个角度展示数据，且逻辑清晰。
- 交互功能丰富，如支持筛选、缩放、悬停显示详细信息等，用户体验良好。
- 使用了v0.dev平台或类似工具实现可视化，效果美观且具有一定的创新性。

● 5分：

- 包含了3-4种图表类型，但可能在某些图表的选择上不够合理。
- 维度指标设计基本合理，但可能在某些地方不够完善。
- 交互功能较为丰富，但可能在用户体验上存在小瑕疵。
- 可视化效果良好，但缺乏一定的创新性。

● 4分：

- 只包含了2-3种图表类型，且图表选择可能较为单一。
- 维度指标设计简单，未能充分展示数据的多维度特征。
- 交互功能较少或交互设计不合理，用户体验一般。
- 可视化效果一般。

4. 整体分析（2-4分）

评分维度：应用方向的创新性、可行性分析的深度、整体价值的体现

● 4分：

- 应用方向：提出了至少2种创新且可落地的应用方向（例如：智能金融风险预警系统、基于大数据的智能投资组合优化、基于多源数据的疾病预测与早期干预），每个方向都具有较高的实用价值，能够紧密结合实际场景（如金融、医学等领域）。
- 可行性分析：可行性分析深入，考虑了技术可行性、成本效益、市场需求、数据可用性等多方面因素，论证充分，有具体数据或案例支持。
- 整体价值：对整个项目的潜在价值进行了全面分析，能够清晰展示从爬虫到可视化的全流程对目标领域的贡献和意义，具有较强的说服力。

● 3分：

- 提出了2种应用方向，但可能在创新性或实用性上略有不足。
- 可行性分析：可行性分析较为全面，但论证深度不够，缺乏与前期数据爬取和分析工作的联系，部分因素的考虑不够充分。
- 整体价值：对项目的潜在价值进行了分析，但分析不够深入，未能充分展示全流程对目标领域的贡献。

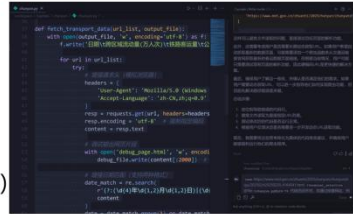
● 2分：

- 只提出了1种应用方向，或提出的2种方向都缺乏创新性和实用性。
- 可行性分析：可行性分析简单，未考虑关键因素或缺乏论证，整体论证不够合理。
- 整体价值：对项目的潜在价值分析不足，未能清晰展示其对目标领域的贡献，整体说服力较弱。

爬虫数据采集

- 目标网站要求：
 - 包含2-4个不同的网站
 - 数据总量：医疗和金融场景各不少于1000条有效数据
- 开发流程规范：
 - 需提交与大模型的完整对话记录截图（含错误调试过程）
 - 关键代码段需注释说明大模型的贡献点
- 输出要求：
 - 代码文件压缩包
 - 大模型对话截图（放入**案例分享报告**中）
 - CSV数据文件（UTF-8编码）
 - 字段说明（写成表格，插入**案例分享报告**中）

参考截图



日期	国际货币基金组织(中国)公共财政数据库	国际货币基金组织(中国)公共财政数据库	国际货币基金组织(中国)公共财政数据库	国际货币基金组织(中国)公共财政数据库
2023-12-31	1000.00	1000.00	1000.00	1000.00
2023-12-30	1000.00	1000.00	1000.00	1000.00
2023-12-29	1000.00	1000.00	1000.00	1000.00
2023-12-28	1000.00	1000.00	1000.00	1000.00
2023-12-27	1000.00	1000.00	1000.00	1000.00
2023-12-26	1000.00	1000.00	1000.00	1000.00
2023-12-25	1000.00	1000.00	1000.00	1000.00
2023-12-24	1000.00	1000.00	1000.00	1000.00
2023-12-23	1000.00	1000.00	1000.00	1000.00
2023-12-22	1000.00	1000.00	1000.00	1000.00
2023-12-21	1000.00	1000.00	1000.00	1000.00
2023-12-20	1000.00	1000.00	1000.00	1000.00
2023-12-19	1000.00	1000.00	1000.00	1000.00
2023-12-18	1000.00	1000.00	1000.00	1000.00
2023-12-17	1000.00	1000.00	1000.00	1000.00
2023-12-16	1000.00	1000.00	1000.00	1000.00
2023-12-15	1000.00	1000.00	1000.00	1000.00
2023-12-14	1000.00	1000.00	1000.00	1000.00
2023-12-13	1000.00	1000.00	1000.00	1000.00
2023-12-12	1000.00	1000.00	1000.00	1000.00
2023-12-11	1000.00	1000.00	1000.00	1000.00
2023-12-10	1000.00	1000.00	1000.00	1000.00
2023-12-09	1000.00	1000.00	1000.00	1000.00
2023-12-08	1000.00	1000.00	1000.00	1000.00
2023-12-07	1000.00	1000.00	1000.00	1000.00
2023-12-06	1000.00	1000.00	1000.00	1000.00
2023-12-05	1000.00	1000.00	1000.00	1000.00
2023-12-04	1000.00	1000.00	1000.00	1000.00
2023-12-03	1000.00	1000.00	1000.00	1000.00
2023-12-02	1000.00	1000.00	1000.00	1000.00
2023-12-01	1000.00	1000.00	1000.00	1000.00
2023-11-30	1000.00	1000.00	1000.00	1000.00
2023-11-29	1000.00	1000.00	1000.00	1000.00
2023-11-28	1000.00	1000.00	1000.00	1000.00
2023-11-27	1000.00	1000.00	1000.00	1000.00
2023-11-26	1000.00	1000.00	1000.00	1000.00
2023-11-25	1000.00	1000.00	1000.00	1000.00
2023-11-24	1000.00	1000.00	1000.00	1000.00
2023-11-23	1000.00	1000.00	1000.00	1000.00
2023-11-22	1000.00	1000.00	1000.00	1000.00
2023-11-21	1000.00	1000.00	1000.00	1000.00
2023-11-20	1000.00	1000.00	1000.00	1000.00
2023-11-19	1000.00	1000.00	1000.00	1000.00
2023-11-18	1000.00	1000.00	1000.00	1000.00
2023-11-17	1000.00	1000.00	1000.00	1000.00
2023-11-16	1000.00	1000.00	1000.00	1000.00
2023-11-15	1000.00	1000.00	1000.00	1000.00
2023-11-14	1000.00	1000.00	1000.00	1000.00
2023-11-13	1000.00	1000.00	1000.00	1000.00
2023-11-12	1000.00	1000.00	1000.00	1000.00
2023-11-11	1000.00	1000.00	1000.00	1000.00
2023-11-10	1000.00	1000.00	1000.00	1000.00
2023-11-09	1000.00	1000.00	1000.00	1000.00
2023-11-08	1000.00	1000.00	1000.00	1000.00
2023-11-07	1000.00	1000.00	1000.00	1000.00
2023-11-06	1000.00	1000.00	1000.00	1000.00
2023-11-05	1000.00	1000.00	1000.00	1000.00
2023-11-04	1000.00	1000.00	1000.00	1000.00
2023-11-03	1000.00	1000.00	1000.00	1000.00
2023-11-02	1000.00	1000.00	1000.00	1000.00
2023-11-01	1000.00	1000.00	1000.00	1000.00
2023-10-31	1000.00	1000.00	1000.00	1000.00
2023-10-30	1000.00	1000.00	1000.00	1000.00
2023-10-29	1000.00	1000.00	1000.00	1000.00
2023-10-28	1000.00	1000.00	1000.00	1000.00
2023-10-27	1000.00	1000.00	1000.00	1000.00
2023-10-26	1000.00	1000.00	1000.00	1000.00
2023-10-25	1000.00	1000.00	1000.00	1000.00
2023-10-24	1000.00	1000.00	1000.00	1000.00
2023-10-23	1000.00	1000.00	1000.00	1000.00
2023-10-22	1000.00	1000.00	1000.00	1000.00
2023-10-21	1000.00	1000.00	1000.00	1000.00
2023-10-20	1000.00	1000.00	1000.00	1000.00
2023-10-19	1000.00	1000.00	1000.00	1000.00
2023-10-18	1000.00	1000.00	1000.00	1000.00
2023-10-17	1000.00	1000.00	1000.00	1000.00
2023-10-16	1000.00	1000.00	1000.00	1000.00
2023-10-15	1000.00	1000.00	1000.00	1000.00
2023-10-14	1000.00	1000.00	1000.00	1000.00
2023-10-13	1000.00	1000.00	1000.00	1000.00
2023-10-12	1000.00	1000.00	1000.00	1000.00
2023-10-11	1000.00	1000.00	1000.00	1000.00
2023-10-10	1000.00	1000.00	1000.00	1000.00
2023-10-09	1000.00	1000.00	1000.00	1000.00
2023-10-08	1000.00	1000.00	1000.00	1000.00
2023-10-07	1000.00	1000.00	1000.00	1000.00
2023-10-06	1000.00	1000.00	1000.00	1000.00
2023-10-05	1000.00	1000.00	1000.00	1000.00
2023-10-04	1000.00	1000.00	1000.00	1000.00
2023-10-03	1000.00	1000.00	1000.00	1000.00
2023-10-02	1000.00	1000.00	1000.00	1000.00
2023-10-01	1000.00	1000.00	1000.00	1000.00
2023-09-30	1000.00	1000.00	1000.00	1000.00
2023-09-29	1000.00	1000.00	1000.00	1000.00
2023-09-28	1000.00	1000.00	1000.00	1000.00
2023-09-27	1000.00	1000.00	1000.00	1000.00
2023-09-26	1000.00	1000.00	1000.00	1000.00
2023-09-25	1000.00	1000.00	1000.00	1000.00
2023-09-24	1000.00	1000.00	1000.00	1000.00
2023-09-23	1000.00	1000.00	1000.00	1000.00
2023-09-22	1000.00	1000.00	1000.00	1000.00
2023-09-21	1000.00	1000.00	1000.00	1000.00
2023-09-20	1000.00	1000.00	1000.00	1000.00
2023-09-19	1000.00	1000.00	1000.00	1000.00
2023-09-18	1000.00	1000.00	1000.00	1000.00
2023-09-17	1000.00	1000.00	1000.00	1000.00
2023-09-16	1000.00	1000.00	1000.00	1000.00
2023-09-15	1000.00	1000.00	1000.00	1000.00
2023-09-14	1000.00	1000.00	1000.00	1000.00
2023-09-13	1000.00	1000.00	1000.00	1000.00
2023-09-12	1000.00	1000.00	1000.00	1000.00
2023-09-11	1000.00	1000.00	1000.00	1000.00
2023-09-10	1000.00	1000.00	1000.00	1000.00
2023-09-09	1000.00	1000.00	1000.00	1000.00
2023-09-08	1000.00	1000.00	1000.00	1000.00
2023-09-07	1000.00	1000.00	1000.00	1000.00
2023-09-06	1000.00	1000.00	1000.00	1000.00
2023-09-05	1000.00	1000.00	1000.00	1000.00
2023-09-04	1000.00	1000.00	1000.00	1000.00
2023-09-03	1000.00	1000.00	1000.00	1000.00
2023-09-02	1000.00	1000.00	1000.00	1000.00
2023-09-01	1000.00	1000.00	1000.00	1000.00
2023-08-31	1000.00	1000.00	1000.00	1000.00
2023-08-30	1000.00	1000.00	1000.00	1000.00
2023-08-29	1000.00	1000.00	1000.00	1000.00
2023-08-28	1000.00	1000.00	1000.00	1000.00
2023-08-27	1000.00	1000.00	1000.00	1000.00
2023-08-26	1000.00	1000.00	1000.00	1000.00
2023-08-25	1000.00	1000.00	1000.00	1000.00
2023-08-24	1000.00	1000.00	1000.00	1000.00
2023-08-23	1000.00	1000.00	1000.00	1000.00
2023-08-22	1000.00	1000.00	1000.00	1000.00
2023-08-21	1000.00	1000.00	1000.00	1000.00
2023-08-20	1000.00	1000.00	1000.00	1000.00
2023-08-19	1000.00	1000.00	1000.00	1000.00
2023-08-18	1000.00	1000.00	1000.00	1000.00
2023-08-17	1000.00	1000.00	1000.00	1000.00
2023-08-16	1000.00	1000.00	1000.00	1000.00
2023-08-15	1000.00	1000.00	1000.00	1000.00
2023-08-14	1000.00	1000.00	1000.00	1000.00
2023-08-13	1000.00	1000.00	1000.00	1000.00
2023-08-12	1000.00	1000.00	1000.00	1000.00
2023-08-11	1000.00	1000.00	1000.00	1000.00
2023-08-10	1000.00	1000.00	1000.00	1000.00
2023-08-09	1000.00	1000.00	1000.00	1000.00
2023-08-08	1000.00	1000.00	1000.00	1000.00
2023-08-07	1000.00	1000.00	1000.00	1000.00
2023-08-06	1000.00	1000.00	1000.00	1000.00
2023-08-05	1000.00	1000.00	1000.00	1000.00
2023-08-04	1000.00	1000.00	1000.00	1000.00
2023-08-03	1000.00	1000.00	1000.00	1000.00
2023-08-02	1000.00	1000.00	1000.00	1000.00
2023-08-01	1000.00	1000.00	1000.00	1000.00
2023-07-31	1000.00	1000.00	1000.00	1000.00
2023-07-30	1000.00	1000.00	1000.00	1000.00
2023-07-29	1000.00	1000.00	1000.00	1000.00
2023-07-28	1000.00	1000.00	1000.00	1000.00
2023-07-27	1000.00	1000.00	1000.00	1000.00
2023-07-26	1000.00	1000.00	1000.00	1000.00
2023-07-25	1000.00	1000.00	1000.00	1000.00
2023-07-24	1000.00	1000.00	1000.00	1000.00
2023-07-23	1000.00	1000.00	1000.00	1000.00
2023-07-22	1000.00	1000.00	1000.00	1000.00
2023-07-21	1000.00	1000.00	1000.00	1000.00
2023-07-20	1000.00	1000.00	1000.00	1000.00
2023-07-19	1000.00	1000.00	1000.00	1000.00
2023-07-18	1000.00	1000.00	1000.00	1000.00
2023-07-17	1000.00	1000.00	1000.00	1000.00
2023-07-16	1000.00	1000.00	1000.00	

四、完成方式

小组合作完成或个人独立完成。

五、额度分配及材料提交

小组合作完成：组长提交《研究实践报告+PPT》、数据文件（压缩包形式）、代码文件（压缩包形式），组长在研究实践报告给出组内成员分工（附表1）。

个人独立完成：个人提交《研究实践报告+PPT》、数据文件（压缩包形式）、代码文件（压缩包形式）。

六、其他

（一）本细则于 2025 年 3 月 4 日经《大数据技术基础》教学团队讨论通过，自 2024—2025 学年春季学期起施行。

（二）本细则由《大数据技术基础》教学团队负责解释。

附表 1：《组内分工表》

分享题目				
给分明细（此处不需要填写，由助教老师统一打分）				
爬虫数据采集 （5-7分）	数据分析 （1-3分）	数据可视化 （4-6分）	整体分析 （2-4分）	合计
分工明细				
序号	姓名	学号	负责工作	备注
1	张三	2022xxxxxx	爬虫数据采集和数据分析	组长
2	李四	2022xxxxxx	对采集数据进行清洗和预处理	
3	王五	2022xxxxxx	数据可视化	

附表 2：《研究实践报告》（示例模板）

一、爬虫数据采集

【大模型的完整对话记录截图（含错误调试过程）】

【爬取的数据文件各个字段的说明，写成表格样式，如下所示】

字段名称	单位	描述
跨区域患者流动量	万人次	当日跨行政区划就医、转诊或流行病学追踪的患者总量
三甲医院日接诊量	万人次	三级甲等医院当日接诊患者总量 (含门诊、急诊)
基层医疗转运总量	万人次	社区卫生服务中心/乡镇卫生院完成的 患者转运总量（含救护车与非急救转运）

二、数据分析

【四类分析说明+与大模型交互截图】

三、数据挖掘

【文字内容+与大模型交互截图】

四、数据可视化

【至少五类图表截图+说明文字】