

# 大数据的数据挖掘课程设计

2025-春季

## 一、概述

围绕以下话题，完成课程报告：

基于大模型的文本数据分析  
基于大模型的合规性审计  
基于大模型的 SQL/Python 代码生成  
基于大模型的可视化与数据洞察  
基于多模态大模型的数据分析  
基于 AI 智能体的自动数据分析  
基于大模型的 AI4Science 研究  
大模型与知识图谱的融合数据分析

**报告要求：**限定选择 3 年以内发表在顶级会议、期刊的相关论文，阅读并撰写报告，报告需体现自己对该内容的理解，如参考互联网上已有的博客、文章需清晰标注出处，提交的笔记会进行查重及 AI 生成审查（疑似 AI 创作会在原得分基础上大幅减分）。

## 二、详细要求

技术报告/阅读笔记的内容至少应包括：

- ✓ 所选话题研究背景？（Introduction）
- ✓ 相关研究工作综述？（Related Work）
- ✓ 该话题的重要的数据集？（Data）

- ✓ 所选话题的主要技术路线是什么？（Approach）

针对技术报告而言，还应该包括：

- ✓ 你打算如何评估实验结果？定性的结果还是定量的结果（例如，哪些性能指标或使用什么样的统计检验？）（Evaluation）
- ✓ 所选论文对你的研究生研究方向有什么启发？你有哪些通过独立思考得到的见解？你打算提出什么方法或算法或者进行什么样的改进？（Future Work）

## 格式要求

大于 8 页，图文并茂；

下面是报告的建议结构以及评分标准。

- ✓ 标题， 作者
- ✓ 摘要：简要描述研究问题、方法和关键结果，应该小于 300 字。
- ✓ 简介（10%）：描述试图解决的问题、为什么这个问题重要以及结果概述
- ✓ 相关工作（10%）：讨论相关的已发表论文。你的方法有何相似或不同之处？
- ✓ 数据（10%）：描述所使用的数据。使用的是什么类型的数据？来自哪里？是否进行了一些预处理步骤？
- ✓ 方法（30%）：讨论解决问题使用的具体方法。为什么方法是正确的做法？你应该证明已将本学期积累的想法和技能应用于解决所选择的问题。
- ✓ 实验（30%）：介绍做的实验来证明你的方法确实解决了问题。采取的实验可能会因具体地选题而异，可以和先前的方法进行比较，进行消融实验以确定系统各种组件的作用，尝试不同的超参数来了解参数选择对结果的影响，使用可视化技术深入了解模型的工作原理，讨论算法的局限性等。这部分应该包含图片、表格来展示实验结果。
- ✓ 结论和未来工作（5%）：总结主要实验结果与可能的方法扩展。

## 二、提交方式

提交内容：(1)汇报 PPT(ppt 格式)和(2)结课报告(pdf 格式)

压缩为一个文件，命名为 2025 结课材料-学号-姓名.RAR

压缩文件发送至 zongchao\_dai@tju.edu.cn

邮件名为：2025 大数据的数据挖掘结课材料

材料提交截至时间为 2025 年 6 月 20 日

## 四、学术道德规范

只要在文章中清楚地引用来源，可以查阅任何论文、书籍、在线参考资料或公开可用的代码。如果你使用了其他课程中做过/正在做的项目，必须清楚的标注出工作的增量部分。只有以下情况会被认定为抄袭：课程报告未通过查重，使用大语言模型撰写。