

AI3602 数据挖掘 期末课程项目安排

概览

课程期末考核方式为 **开放式选题的小组课程项目**。同学们可以自由组队并自行选择感兴趣的、与数据挖掘相关的课题进行探索。

期末项目的评分方式为 **助教评分 + 组间互评（投票）**。我们计划在第十六周上课时组织课堂展示，展示形式为 **线下 Demo + 线上投票**。

具体的组队、选题、展示须知详见下文。

选题

- 课程期末项目为 **开放式选题**，原则上任何和数据挖掘相关的 Topic 均可作为项目选题。
- 我们鼓励同学们选择和课堂内容相关的题目，并结合人工智能专业和领域前沿知识进行拓展。

参考的选题范围如下（包括但不限于）

- **高维数据挖掘**. 探讨在向量、张量等高维数据上的高效算法。
- **图数据挖掘**. 在图结构的数据上的数据挖掘任务。例如社区检测、链路预测、节点分类等。
- **社交网络分析**. 对社交网络上的数据进行分析。例如谣言溯源等。
- **流数据算法**. 流数据的算法或应用。
- **源代码数据挖掘**. 在大量源代码数据上的数据挖掘任务。例如代码搜索、API 推荐等。
- **数据挖掘中的安全和隐私保护**. 探讨数据挖掘中的安全和隐私问题。例如对用户隐私信息的提取、重构攻击，差分隐私等。
- ***推荐系统**. 例如音乐推荐、电影推荐、论文推荐等。
- ***时空数据挖掘**. 在时间、空间数据上的任务。例如利用车辆移动轨迹做地点预测、轨迹相似度学习等。

项目的具体内容包括但不限于：使用现有算法在特定数据集上搭建系统解决问题，或尝试对已有算法进行改进等。请注意在展示时突出项目的主要内容、尝试解决问题的思考过程以及对结果的分析。

展示

- 可以通过 **录屏、现场互动展示、海报** 等形式展示课程项目，也可以采用其他形式。
 - 如需使用录屏、Demo 或 Slide，请自备电脑。如需制作海报，请使用 A1 大小的纸张。
 - 各大学术会议的网站上通常会公开录用文章的 Poster，可供参考（例如 [ICML 2023](#)）。
- 课堂展示时，展示材料和口头交流 **使用中文或英文均可**。
- 课堂展示的听众流动性较大，与一般的 Presentation 不同，解说时请注意控制时长、突出重点、吸引听众。
- 考虑到每位同学都需要参与投票，我们建议组内的同学轮换负责展示，组内其他同学可以与其他小组交流投票。

评分

小组项目的最终成绩将综合投票结果和助教评分，助教评分的参考指标如下。

指标	参考标准	分值 (%)
选题	选题是否与数据挖掘相关	10
工作量/完成度	工作量是否充分（包含必要的实验，或实现了可用的系统）	15
问题与创新性	问题场景或提供的解决方案是否具有创新性	20
技术正确性	项目的方法设计、实现和实验设置是否科学	15
展示材料	展示材料是否结构清晰、重点突出、有吸引力	20
汇报	汇报是否逻辑清晰、演讲流利、易于理解	20