

订阅DeepL Pro以编辑此演示文稿。  
访问[www.DeepL.com/pro](https://www.deepl.com/pro?cta=edit-document)，了解更多信息。

**作业 1：新闻源推荐**

**目标：**

鉴于您目前所了解的情况，您将如何端到端地实施您的第一个推荐系统？具体来说，考虑通过电子邮件推荐每日新闻。您有一份用户及其印象的列表，您的任务是设计一个简单的新闻源排名系统。

**数据集：**

**MIND 新闻数据集**[：https://www.kaggle.com/datasets/arashnic/mind-news-dataset/data](https://www.kaggle.com/datasets/arashnic/mind-news-dataset/data)

**方法：**

根据我们在课堂上所讲的内容，应用你的协同过滤方法。你可以使用基于内容的方法、基于用户的方法、矩阵因式分解（MF）方法，甚至可以提出自己的想法。你可以采用传统方法、聚类、SVD 等。请注意，这不一定要涉及深度学习。

**重要提示：**不接受抄袭 Kaggle 现有解决方案的行为，也不会获得评分。

**评估：**

使用正确的指标评估您的方法，并提交一份两页的短文（最多两页）。报告中不得包含代码。

**关于 MIND 数据集的论文样本：**

* <https://aclanthology.org/2020.acl-main.331.pdf>
* <https://arxiv.org/pdf/2209.06131>

**注：**报告应为 PDF 格式。可以是双栏或单栏。如果愿意，您可以使用 [Overleaf](https://www.overleaf.com/) 进行排版。

**评分标准：**

* **方法（40 分）**
  + 示例 1：基于内容
  + 示例 2：基于用户
  + 例 3：聚类或 SVD 等
* **成果与评估（40 分）**
  + 实验设置（10 分）：你是如何将数据分为训练集和测试集的？你是否遵循了给定的设置，为什么？
  + 指标（5 分）您选择这些指标的原因是什么？是否有足以反映推荐模型强度的指标？
  + 结果与比较（25 分） 请讨论结果，而不仅仅是数字。
* **可执行笔记本（20 分）**

**提交：**

提交一个 ZIP 文件。确保 ZIP 文件在提交前可以解压缩。提交的文件应包括

* 一份包含报告的 PDF 文件。
* 项目使用的代码。