大作业内容与说明

总体说明：除去已经完成综述任务的章节，每个小组在剩下两个章节中挑选一个，完成对应的大作业任务。小组在完成大作业任务时最好有自己的创新点，不要只拘泥于复现别人的算法，并在此基础上尽可能提升模型的性能。

Section 5 Graph Data Mining

(1) 基础任务：在 Cora 与 Citeseer 数据集上完成节点分类任务。

1. Cora 数据集链接：https://relational.fit.cvut.cz/dataset/CORA
2. Citeseer 数据集链接：https://relational.fit.cvut.cz/dataset/CiteSeer
3. 此外，也可以调用 DGL 库的 API 来获取 Cora 数据集。官方文档：

Cora: https://docs.dgl.ai/en/1.1.x/generated/dgl.data.CoraGraphDataset.html

Citeseer: https://docs.dgl.ai/en/1.1.x/generated/dgl.data.CiteseerGraphDataset.html#dgl.data.CiteseerGraphDataset

(2) 拓展任务：在 Ogbn-products 数据集上完成节点分类任务。

1. Ogbn-products 数据集链接：https://ogb.stanford.edu/docs/nodeprop/#ogbn-products
2. OGB 提供了基于 Pytorch Geometric 和 DGL 的 DataLoader API 来下载数据集。API 的使用说明：https://ogb.stanford.edu/docs/nodeprop/#dgl
3. 测试方法：调用 OGB 的统一评估函数来测试模型的性能。评估函数的调用方法：https://ogb.stanford.edu/docs/nodeprop/#eval

Section 6 Recommendation System

(1) 基础任务：在 Kaggle 平台上完成基于 KKBox 历史数据的音乐推荐任务。

1. Kaggle 竞赛链接：https://www.kaggle.com/competitions/kkbox-music-recommendation-challenge/data
2. 各小组需要注册一个 Kaggle 账号，最终将模型在测试集上的结果提交至 Kaggle。
3. 测试方法：将规范化的测试集结果上传至 Kaggle 上，最终性能以排行榜上的性能为准。在 16 周之前，各小组可以提交任意次结果。

(2) 拓展任务：基于 Synerise 数据集，完成用户流失预测、产品倾向性预测、类别倾向性预测等任务。

1. Recsys 2025 竞赛链接：https://recsys.acm.org/recsys25/challenge/
2. 任务说明：https://www.recsyschallenge.com/2025/
3. 数据集说明以及下载方式：https://recsys.synerise.com/data-set
4. 测试方法：Recsys 提供了测试用的代码库：https://github.com/Synerise/recsys2025，各小组需按照代码库的测试脚本进行性能评估。

Section 7 Time Series Data Mining

任务：在 Kaggle 平台上完成股票走势预测任务。

1. Kaggle 竞赛链接：https://www.kaggle.com/competitions/the-winton-stock-market-challenge
2. 各小组需要注册一个 Kaggle 账号，最终将模型在测试集上的结果提交至 Kaggle。
3. 测试方法：将规范化的测试集结果上传至 Kaggle 上，最终性能以排行榜上的性能为准。在 16 周之前，各小组可以提交任意次结果。