Experiment 1

# 数据预处理

# 数据来源

## 亚马逊评论

来自 6,643,669 位用户的 34,686,770 条亚马逊评论，涵盖 2,441,053 种产品

- https://www.kaggle.com/datasets/kritanjalijain/amazon-reviews

## LinkedIn 招聘信息（2023 - 2024 年)

该数据集包含 124,000 年和 2023 年列出的 2024+ 个职位发布的近乎全面的记录。每个单独的帖子都包含帖子和公司的数十个有价值的属性

- https://www.kaggle.com/datasets/arshkon/linkedin-job-postings

# 实验要求

**实验一：在提供的Amazon数据集上构建词向量表示**

- 了解什么是word2vec编码；
- word2vec作为简单的神经网络模型， 其架构图构建；
- word2vec常用的两种模式；

# 实验要求

**实验二： 在LinkedIn数据集上利用node2vec方法构建节点表示**

- 了解什么是节点表示；
- 了解node2vec的实现算法；

# 实验要求

## 实验三： 向量表示的相似度计算

- 计算word2vec的向量相似度；
- 计算node2vec的向量相似度；

# 实验要求

**实验四： 利用T-SNE进行数据可视化 或 任何你感兴趣的分析，参考Kaggle数据集网站上的 code**

- 了解什么是节点表示；
- 了解node2vec的实现算法；