****

**大数据推荐系统实验报告**

**院（系、所） 电子信息与通信学院**

**班 级 提高2101班**

**专 业 电子信息工程（提高班）**

**姓 名 杨筠松**

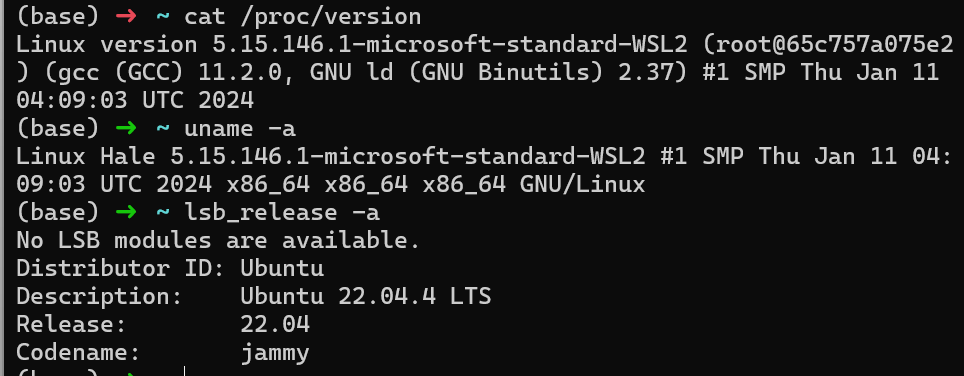
**指 导 老 师 陈建文**

**2024年 6 月 7 日**

1. 实验环境

软件环境：**wsl2** (**Windows Subsystem for Linux**)

Linux版本：Ubuntu 22.04.4LTS



1. 实验内容

#### 实验背景

本实验基于阿里巴巴集团主办的天池大数据竞赛中的一项赛题，即天猫推荐算法大挑战。该赛题要求参赛者根据用户在天猫三个月内的行为日志，预测他们在未来一个月内对品牌商品的购买行为。这是一个典型的推荐系统问题，涉及到大数据处理、用户行为分析和预测模型构建。

#### 实验目标

实验的主要目标是使学生掌握推荐系统的基本工作流程，包括数据加载、特征工程、模型训练和推荐效果评估。通过实践，学生将学习如何实现和优化推荐算法，提高推荐系统的准确率和召回率。

#### 实验方法和步骤

1. **数据准备和加载**：
   * 使用的数据集包含约100,000条天猫用户的行为记录，涉及点击、购买、收藏和加入购物车等行为。
   * 数据加载部分涉及从CSV文件中读取数据，并对数据进行初步处理。
2. **推荐方法**：
   * **基于规则的推荐**：此方法直接根据用户的行为日志生成推荐，如推荐用户最近一个月内互动次数较多的品牌。
   * **基于逻辑回归的推荐**：使用逻辑回归模型基于用户的历史行为预测其对品牌的购买概率。
3. **特征工程**：
   * 开发和选择影响用户购买决策的特征，如用户对品牌的点击次数、购买次数、收藏次数、加入购物车次数以及最近的互动日期等。
4. **模型训练与推荐实施**：
   * 使用训练数据集构建模型，并应用模型于测试集生成购买推荐。
   * 推荐结果的格式为每个用户推荐的品牌列表。
5. **推荐效果评估**：
   * 使用准确率、召回率和F1得分来评估推荐效果。
6. 实验过程说明及问题
7. 解决方案
8. 实验结果
9. 程序源代码