

泰迪科技数据分析

项目实训

餐饮大数据智能推荐策略

广东泰迪智能科技股份有限公司

# 项目介绍

## 项目背景与目标

广大上班族已经习惯于在餐饮外卖平台上订餐。W餐饮外卖平台，向广大用户提供网上订餐服务。W平台订餐完成后，平台会引导用户对于品尝过的菜品进行评价打分。运营方发现老用户的下单率呈现下降态势，市场部门建议，希望针对老用户进行个性化的菜品推荐，包括用户的偏爱菜品及新菜品。

W餐饮平台的技术部门决定根据近期用户对菜品的评分历史数据，建立菜品推荐模型，向用户们提供菜品推荐。菜品智能推荐系统，作为原来的餐饮外卖平台系统的扩展与补充，主要负责对用户的历史评分数据进行处理，并生成推荐结果集。

本文提供了平台2个月的用户评分数据及菜品数据，通过数据分析，为平台制定综合的菜品推荐策略，以便更好地服务线上用户。

## 项目数据

1. 用户评分数据集

在网站的后台服务器就会以JSON格式保存这些用户评分数据。

表1‑1 MealRatings\_201705\_201706.json文件字段说明

|  |  |
| --- | --- |
| **字段名** | **描述** |
| UserID | 用户ID |
| Rating | 评分 |
| ReviewTime | 评分的时间戳 |
| Review | 评论 |
| MealID | 菜品ID |

1. 菜品数据集

网站的后台数据库（MySQL）中保存着菜品的数据集。（已提供了一份meal\_list.sql文件，可以通过指令方式直接执行该sql文件，实现数据导入MySQL数据库）

表1‑2 meal\_list.sql文件字段说明

|  |  |
| --- | --- |
| **字段名** | **描述** |
| mealno | 序号 |
| mealID | 菜品ID |
| meal\_name | 菜品名称 |

## 项目难度

★★★

# 任务清单

**TASK 1 ：数据读取/存储**

**TASK 2 ：数据探索与预处理**

结合最终业务目标实现相关探索与分析

缺失值、异常值、重复值处理

**TASK 3 ：智能推荐**

# 核心操作

* + 了解数据各字段信息与数据分析流程
  + 数据预处理
  + 智能推荐

# 实现技术及工具

结合目标及数据，小组自行探索确认合理的技术及工具，注意事项如下。

1、**必须使用**到本次实习实训所学的一些技术框架，如HDFS、Hive、Spark SQL、Spark MLlib。

2、可以使用以前学的专业知识，如python等辅助使用，实现可视化分析，或处理等环节。**（若使用其他编程语言或技术实现，其在案例实现的占比如果大于80%，小组的人数不能超过3人）**

# 项目提交成果及汇报要求

**【项目提交成果】**

小组材料文件夹命名：**“第？组+小组成员名称”**，文件夹包含如下内容，并将文件夹于9月23日前提交至邮箱：[yelifan@tipdm.com](mailto:yelifan@tipdm.com)。

1、**项目报告文档**——需用word完整呈现对应方案设计内容和成果（已提供文档模板：广告检测中的流量作弊识别【文档模板】）

2、**汇报PPT**——基于项目报告自行转换成PPT用于汇报

3、**其他材料，如代码、数据等文件**（若有则同步提交）

**【汇报要求】**

1、每组或每人进行答辩，时间5~8分钟。

2、答辩时每个项目均需提供辅助汇报用PPT，PPT 不得少于12页。

3、答辩汇报内容从需求分析、成员分工、实现过程、结论分析等方面进行阐述。

# 评分标准

**考勤及课堂表现30%+报告文档40%+汇报30%**

*注：根据小组的成员分工情况、实现的过程逻辑、最终效果等评定小组中每个人的报告文档分及汇报的分数。*