**项目分析报告**

项目名称：餐饮智能推荐服务

项目成员： 吴鋆

项目时间： 2020年2月

目录

[一、 项目背景 3](#_Toc3350)

[二、 项目流程及具体实现 3](#_Toc7740)

[1、项目准备 3](#_Toc774)

[1.1 数据准备 3](#_Toc31075)

[1.2 环境准备 3](#_Toc8141)

[1.3 算法准备 3](#_Toc26874)

[2、 项目思路 4](#_Toc3425)

[2.1 数据预处理 4](#_Toc16598)

[2.2 划分训练集和测试集 4](#_Toc9007)

[2.3 计算菜品的相似度 4](#_Toc10730)

[2.4 模型评估 5](#_Toc24722)

[3、 项目流程 5](#_Toc8060)

[3.1 读取数据 5](#_Toc12235)

[3.2 数据预处理 5](#_Toc30382)

[3.3 划分训练集和测试集 6](#_Toc24007)

[3.4 构建菜品的相似度矩阵 7](#_Toc19361)

[3.5 模型评估 7](#_Toc21838)

[三、 项目结果评价分析 8](#_Toc18570)

# 项目背景

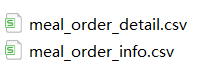
近年来，随着互联网的飞速发展，人们的日常生活中产生了越来越多的数据信息，如此大量的数据信息满足了用户在大数据时代对数据的需求，而基于这些数据的智能推荐技术也因此在人们的生活中占了越来越大的比重。智能推荐是一种个性化的信息过滤技术，通过用户的历史记录，归纳用户的兴趣特征，从而预测用户可能感兴趣的信息和商品。餐饮智能推荐服务是针对智能推荐技术的进一步细化，作为必不可少的衣食住行之一的餐饮行业，与人们的生活密不可分，而在就餐时，顾客经常在大量的菜品信息中，无法真正找到自己想要的菜品。因此，就应运产生了餐饮智能推荐服务技术。

餐饮智能推荐服务技术的任务是将菜品呈现在顾客面前，帮助顾客发现合适的菜品，通过分析顾客的订单信息并进行建模，从而预测顾客感兴趣的菜品，并给顾客作菜品推荐，一方面节约顾客点餐的时间，从而提高就餐效率，另一方面也可以提高顾客的就餐体验，回头率增加，对餐饮企业的好处也是不言而喻的。

# 项目流程及具体实现

**1、**项目准备

### 1.1 数据准备



meal\_order\_info.csv为订单表，meal\_order\_detail.csv为订单详情表

### 1.2 环境准备

本项目将在python环境下运行，需要的库有pandas、numpy、heapq，使用的软件为PyCharm2018.1，操作系统为Windows10

### 1.3 算法准备

基于物品的协同过滤算法(ItemCF)给用户推荐那些和他们之前喜欢的物品相似的物品。比如：该算法会因为你购买过《数据挖掘导论》而给你推荐《机器学习》。不过ItemCF算法不利用物品的内容属性计算物品之间的相似度，它主要通过分析用户的行为记录计算物品之间的相似度。该算法认为，物品A和物品B具有很大相似度的原因是因为喜欢物品A的用户大都也喜欢物品B。

## 项目思路

### 2.1 数据预处理

⑴订单详情表中会有多余的空格符、换行符等特殊符号，需要对数据先进行处理；

⑵由于订单表中订单的状态绝大多数都是1，只有少部分是0或2，故可以只保留订单状态为1的订单，其他订单清除；

⑶订单详情表中有些订单数据没有在订单表中出现，需要清除这类数据；

⑷由于给用户推荐白饭意义不大，故清除订单详情表中含有白饭的订单；

⑸在本项目中只需考虑菜品名称和用户id，故清除其他无关数据；

⑹由于有些用户点同一道菜点了多次，这不在本项目的考虑范围内，故清除重复数据；

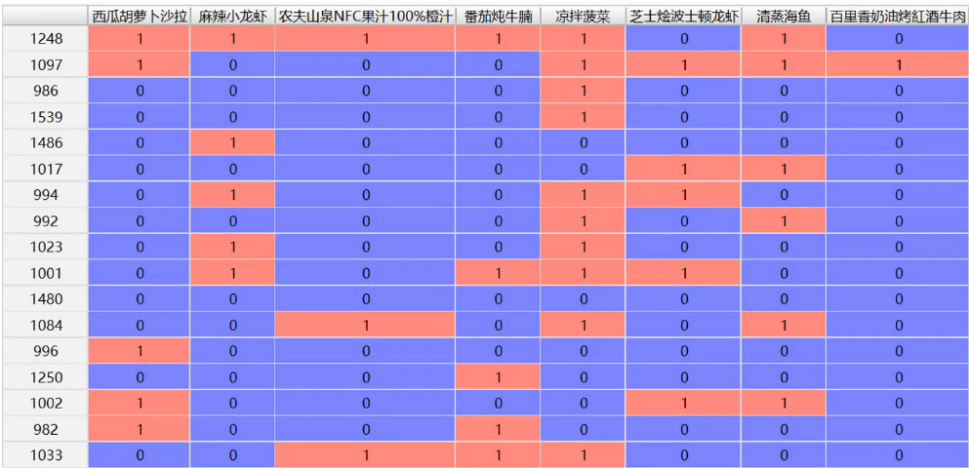
⑺保留点过三道菜以上的用户。

### 2.2 划分训练集和测试集

考虑到同一用户不能同时出现在训练集和测试集中，故对用户进行划分，然后再索引用户对应的菜品名称。

### 2.3 计算菜品的相似度





然后再构建训练集中菜品的同现矩阵，同现矩阵表示同时喜欢两个菜品的用户数，是根据菜品的二元矩阵计算出来的，此为；

接下来统计训练集中喜欢菜品的用户数矩阵，再运用上面的公式即可求出菜品与菜品之间的相似度，进而得到基于训练集的关于所有菜品的相似度矩阵SIM。

### 2.4 模型评估

与训练集操作相同，构建基于测试集的菜品的二元矩阵R，采用公式：



P表示测试集用户对每个菜品的感兴趣程度，由矩阵P可以给测试集的每个用户生成一个推荐列表，再根据正确推荐的菜品数和所有推荐的菜品数的比值，来计算菜品推荐准确率：



准确率越接近1，说明本模型的推荐效果越好

## 项目流程

### 3.1 读取数据





### 3.2 数据预处理

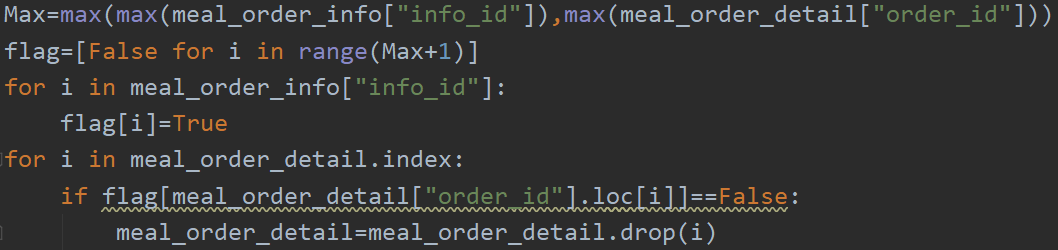
⑴处理订单详情表中的特殊符号



⑵保留订单表中订单状态为1的订单



⑶清除订单详情表中订单表没出现的订单



⑷清除订单详情表中含有白饭的订单



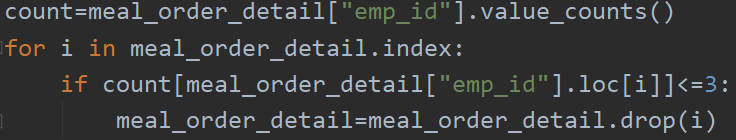
⑸只保留菜品名称和用户id



⑹清除重复数据



⑺保留点过三道菜以上的用户



⑻由于此时订单详情表的标签已经乱了，故重置数据的标签

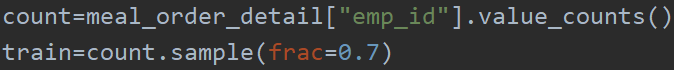


### 3.3 划分训练集和测试集

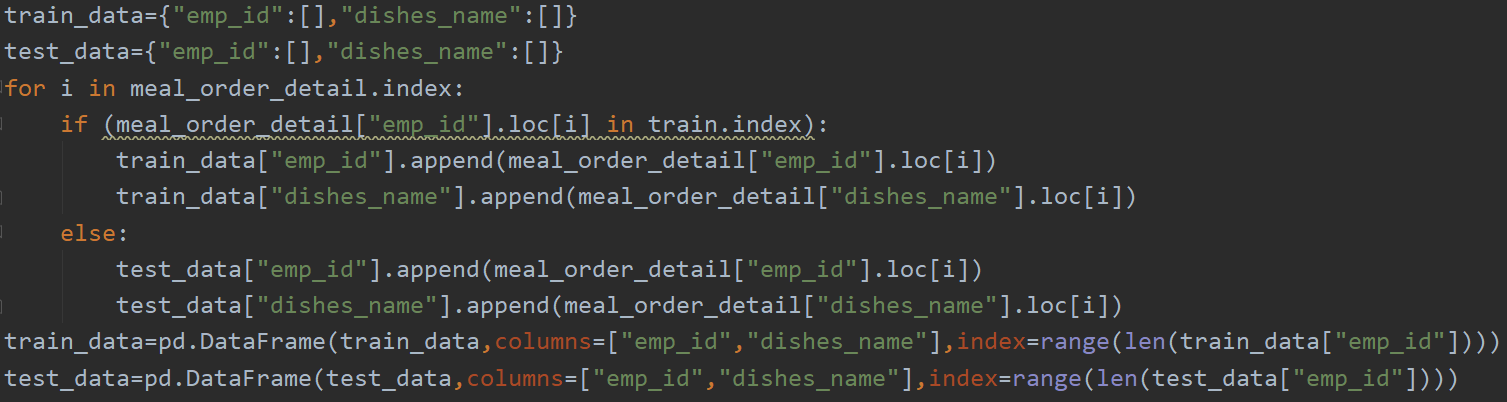
⑴在划分前先把所有菜名储存起来



⑵按用户进行划分

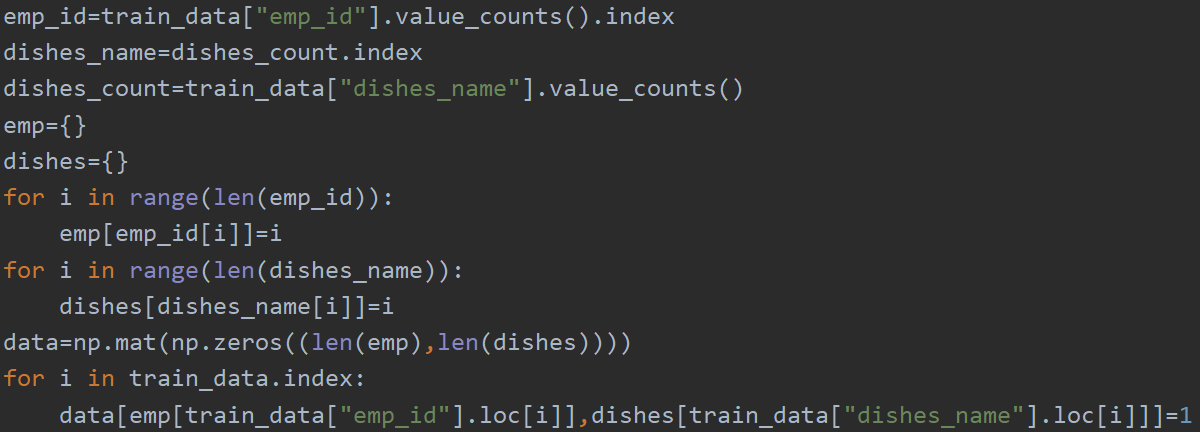


⑶把meal\_order\_index中按已经划分好的用户进行分组

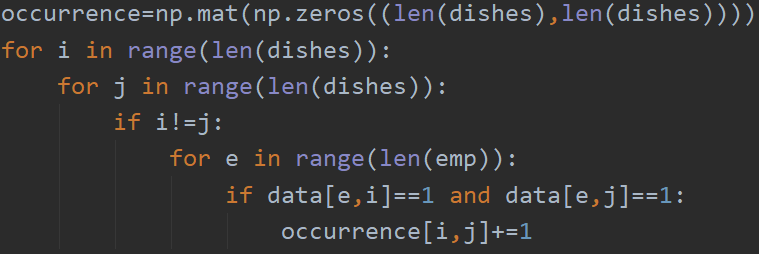


### 3.4 构建菜品的相似度矩阵

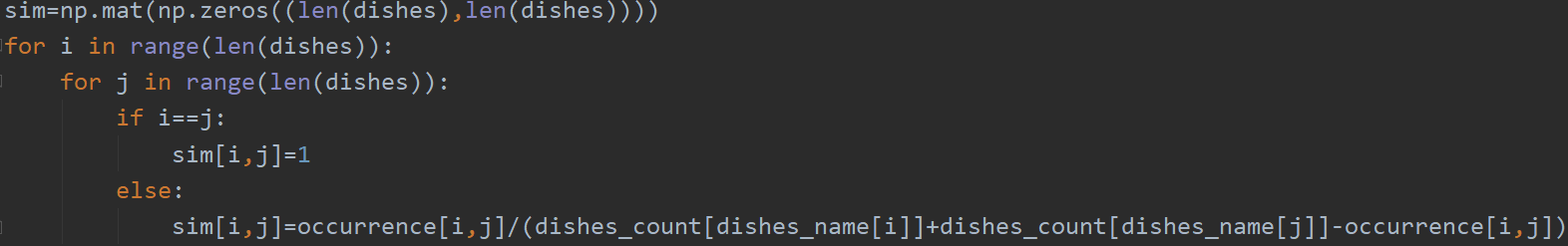
⑴通过建立用户id与标签、菜品名称与标签的反向索引，来构建基于训练集的所有菜品的二元矩阵data



⑵根据data矩阵构建同现矩阵occurrence

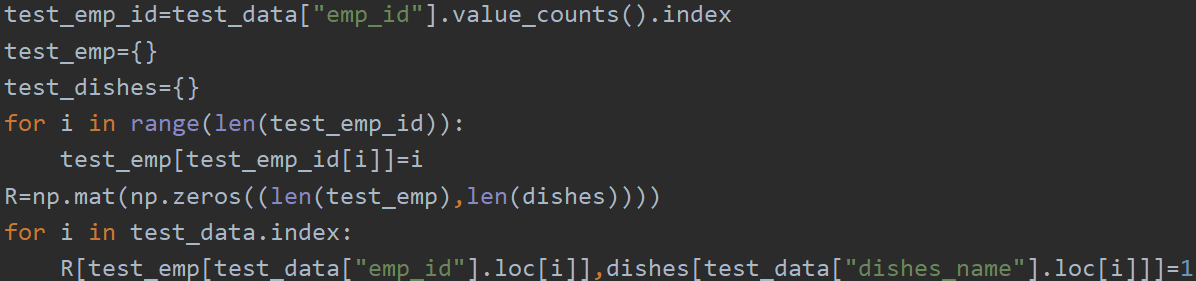


⑶由Jaccard相似度计算公式来构建所有菜品的相似度矩阵sim



### 3.5 模型评估

⑴构建基于测试集的菜品的二元矩阵R



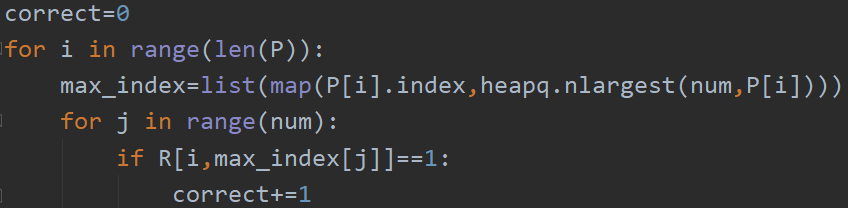
⑵计算测试集用户对每个菜品的感兴趣程度的矩阵P



⑶定义推荐列表长度



⑷生成推荐列表，并计算正确推荐的菜品数

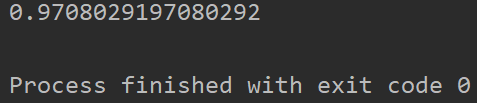


⑸计算准确率并输出



# 项目结果评价分析

本项目运行结果如下：



在推荐列表长度为5的情况下有97%的准确率，可以看出该模型的推荐效果还是不错的，而且随着推荐列表长度的增加，准确率在不断减少，因为推荐列表越长，就有越多的错误推荐出现，而推荐列表长度越短，就越能保证所推荐的是客户感兴趣的菜品。不过本项目有一个缺点，就是我们对于重复值采取的操作是去除重复值，但是对这些重复值而言，有重复则说明这个用户特别喜欢这道菜，所以对本模型的改进建议是不去除重复值，或者在去除重复值的时候考虑重复次数，并在计算感兴趣程度矩阵的时候对该菜品给予一定权重，这样更能推荐到深得人心的菜品。