# 餐饮智能服务项目分析报告

目录

[一、 背景介绍 1](#_Toc1553)

[1、 行业背景 1](#_Toc30429)

[2、 分析目标 2](#_Toc22944)

[二、 数据说明 2](#_Toc4750)

[1、 数据来源 2](#_Toc529)

[2、 属性名称说明 2](#_Toc27349)

[三、 数据分析 3](#_Toc14355)

[1、 数据获取与探索分析 3](#_Toc29035)

[1) 读取订单数据并进行探索 3](#_Toc31518)

[2) 处理菜品名称字符数据 3](#_Toc12803)

[3) 构建热销度评分指标 3](#_Toc29132)

[4) 绘制条形图展示热销Top10 4](#_Toc9145)

[2、 数据预处理 4](#_Toc7688)

[1) 统计每个订单状态的占比 4](#_Toc4871)

[2) 选取有效的订单数据 5](#_Toc3003)

[3) 选取主要特征 5](#_Toc20536)

[3、 构建模型准备 5](#_Toc5600)

[1) 将数据划分为训练集和测试集 5](#_Toc28349)

[2) 构建训练集数据客户——菜品二元矩阵 5](#_Toc32029)

[4、 模型构建 6](#_Toc966)

[1) 计算相似度矩阵 6](#_Toc9309)

[2) 生成推荐列表 7](#_Toc1810)

[5、 模型评价 7](#_Toc11003)

[1) 构建测试集客户ID——菜品字典 7](#_Toc20179)

[2) 构建评价指标，并分析推荐效果 7](#_Toc6526)

[四、 小结 8](#_Toc10262)

1. **背景介绍**
2. **行业背景**

在大数据和人工智能发展的时代，餐饮业的智能化已成为现代食客讨论的热门话题，如何运用智能技术来传达其产品理念，使其产品在众多产品中得到更深层次的关注，并通过网络营销手段，以便更好地服务于客户，提高业务以成为现代餐饮业关注的焦点。

目前，对于没有大数据支持的餐饮业来说，针对所有人一切都是一样的，包括菜品价格、会员打折信息、赠送奖品等。例如，今天的餐厅为会员提供打折菜品，但是有的会员恰好不爱吃，这样就达不到促销的效果，消费者会觉得这样的活动是可有可无的。所谓个性化营销方案是指基于对餐厅消费者用户的行为特征数据进行分析，然后根据用户的爱好与需求制定自己的营销方案，因为该方案是通过对用户数据进行有针对性的分析，用户有很大的可能会喜爱推荐的菜品，这就意味着我们的促销是有效的，这些都是大数据和人工智能发展的成果。

面对新环境下的市场需求，在保证餐厅品味和价格适中的前提下，企业需要更加关注用户的个性化需求，应该分析每个用户的饮食偏好，并最终使用人工智能和大数据的营销方法创建个性化餐厅是互联网的必然选择。随着餐饮业人工智能和大数据的发展，一家餐厅是否会使用基于大数据和人工智能技术的互联网解决方案，这将成为餐厅未来发展的趋势指标。

1. **分析目标**

对某餐饮企业的订单数据进行分析，利用协同过滤算法对就餐的客户进行菜品智能推荐。

1. **数据说明**
2. **数据来源**

数据来源：某餐饮企业的系统数据库

数据时间：2016年8月1日——2016年8月31日

1. **属性名称说明**

|  |  |  |  |
| --- | --- | --- | --- |
| **名称** | **含义** | **名称** | **含义** |
| detail\_id | 订单详情ID | place\_orser\_time | 用餐时间 |
| order\_id | 订单ID | discount\_amt | 折扣额度 |
| dishes\_id | 菜品ID | discount\_reason | 折扣说明 |
| logicpm\_name | 类别名称 | kick\_back | 回扣 |
| parent\_class\_name | 父类名称 | add\_inprice | 添加价格 |
| dishes\_name | 菜品名称 | add\_info | 添加信息 |
| itemis\_add | 是否为添加菜 | bar\_code | 条形码 |
| counts | 数量 | picture\_file | 图片 |
| amounts | 销售金额 | emp\_id | 客户ID |
| cost | 成本 |  |  |

**表格 1 订单详情表(meal\_order\_detail)说明**

|  |  |  |  |
| --- | --- | --- | --- |
| **名称** | **含义** | **名称** | **含义** |
| info\_id | 订单ID | lock\_time | 锁单时间 |
| emp\_id | 客户ID | cashier\_id | 收银ID |
| number\_consumers | 消费人数 | pc\_id | 终端ID |
| mode | 消费方式 | order\_number | 订单号 |
| dining\_table\_id | 桌子ID | org\_id | 门店ID |
| dining\_table\_name | 桌子名称 | print\_doc\_bill\_num | 打印doc账单的编程 |
| expenditure | 消费金额 | lock\_table\_info | 桌子关闭信息 |
| dishes\_count | 总菜品数 | order\_status | 0:未结算;1:结算;2:已锁单 |
| accounts\_payble | 付费金额 | phone | 电话 |
| use\_start\_time | 开始时间 | name | 名字 |
| check\_closed | 支付结束 |  |  |

**表格 2 订单表(meal\_order\_info)说明**

1. **数据分析**
2. **数据获取与探索分析**
3. **读取订单数据并进行探索**

读取数据的方法有很多中，在这里是用pandas读取的数据，因为用pandas读取的数据是dataframe类型的，便于后面的数据预处理。

1. **处理菜品名称字符数据**

在本次项目中，我们主要是对菜品进行推荐，因此我们首先打印出菜品名单，人后我们发现由于系统问题使得打印出的菜品名单有一些特殊符号，例如空格符、换行符，因此我们对菜品名称进行处理，得出规整的菜品名单。

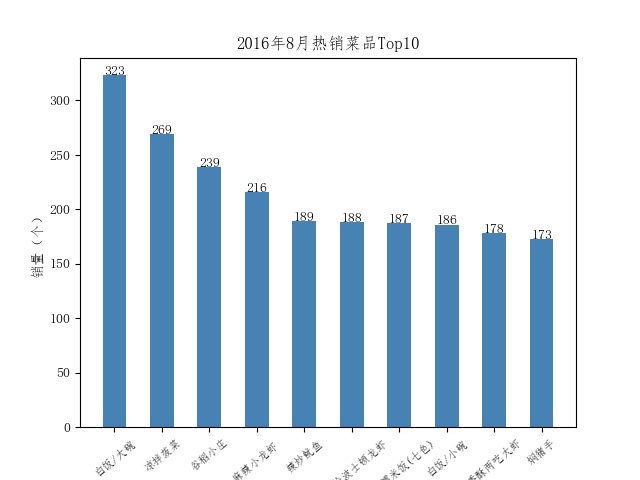
1. **构建热销度评分指标**

热销度即在一定时间内产品销量的程度。根据餐饮企业金30天（即2016年8月1日到2016年8月31日）的菜品销售统计每个菜品的热销度，其计算公式如式（1）所示。经Min-Max标准化后计算得分，最高分为1分，最低分为0分。

 （1）

其中，热销度评分为某项菜品的热销度评分，其值范围在0到1之间；为某项菜品的销售分数；为该餐饮企业最近30天内有销售记录的菜品中的最大销售份数；为该餐饮企业最近30天内有销售记录的菜品中的最小销售份数。

1. **绘制条形图展示热销Top10**



**图表 1 2016年度8月份热销菜品Top10**

由图表1可知，2016年度8月份热销菜品Top10分别为：白饭/大碗、 凉拌菠菜、 谷稻小庄、 麻辣小龙虾 、辣炒鱿鱼、芝士烩波士顿龙虾、五色糯米饭(七色)、白饭/小碗、香酥两吃大虾、焖猪手，销量均为170以上。

1. **数据预处理**
2. **统计每个订单状态的占比**

当客户看到菜单中没有喜欢的菜品而离开，或者是点餐过后突然有事离开而导致订单未完成或者未结算，但此时订餐系统中已经出现了这份订单，此时就会在订单表(meal\_order\_info)中的订单状态（order\_status）一列显示0或者2。

|  |  |  |  |
| --- | --- | --- | --- |
| **订单状态** | **0** | **1** | **2** |
| **占比** | 0.0095 | 0.9873 | 0.0032 |

**表格 2 各订单状态的占比**

1. **选取有效的订单数据**

由表格2可知，无效订单即未完成或者未结算的订单占比很少，而我们使用的数据量很大，因此可以直接在订单表（meal\_order\_info）中删除这部分无效订单。

而当我们完成上述操作之后，就会发现有部分订单在订单详情表（meal\_order\_detail）中有，而在订单表（meal\_order\_info）中没有出现，这也属于无效数据，会影响后期的分析，因此需要删除。

且由于我们的分析目标在于菜品推荐，而白饭是主食，大部分客户都会下单，因此并不需要向客户推荐白饭，因此对于只点了白饭的订单数据也可以直接删除。

1. **选取主要特征**

根据我们的分析目标，即对就餐的客户进行菜品智能推荐，选取菜品名称（dishes\_name）和客户ID（emp\_id）作为构建模型的主要特征属性。

1. **构建模型准备**
2. **将数据划分为训练集和测试集**

由于sklearn中的train\_test\_split函数是随即划分，可能会导致同一个客户的订单信息被划分到两个数据集中，因此我们不采用此函数。

首先将用户ID单独提取出来作为一个数据集，然后将用户ID按照2：8的比例划分为测试集和训练集，最后在全部遍历，将测试集名单中的用户点过的菜放进测试集菜品名单，训练集名单中的用户点过的菜放进训练集菜品名单，这样就避免了划分数据集时，同一个用户的多个订单重复出现在训练集和测试集中。

1. **构建训练集数据客户——菜品二元矩阵**

为方便后期模型构建中的一些计算，我们将训练集数据转化为客户——菜品二元矩阵。此矩阵可以用Python中的透视表实现，以菜品名称为列索引，用户ID为行索引，用户点过的菜用数字1表示，为点过的菜用数字0表示，形成了一个0-1矩阵，大大降低了计算难度。

1. **模型构建**

在本次项目中我们运用到的是协同过滤算法。协同过滤算法的原理是找到相似的用户或物品进行推荐，最常用的是基于用户的协同过滤算法（user-based collaboratlve filtering）和基于物品的协同过滤算法（item-based collaborative filtering）。

基于用户的协同过滤算法（user-based collaboratlve filtering）是通过用户的历史行为数据发现用户对商品或内容的喜欢（如商品购买，收藏，内容评论或分享），并对这些喜好进行度量和打分。根据不同用户对相同商品或内容的态度和偏好程度计算用户之间的关系。在有相同喜好的用户间进行商品推荐。

基于物品的协同过滤算法（item-based collaborative filtering）是通过计算不同用户对不同物品的评分获得物品间的关系。基于物品间的关系对用户进行相似物品的推荐。这里的评分代表用户对商品的态度和偏好。

根据项目分析目标可知我们应选用基于物品的协同过滤算法（简称ItemCF算法）原理，个性化推荐主要分为两部：

·计算菜品与菜品之间的相似度

·根据菜品的相似度和客户的历史行为给客户生成推荐列表

1. **计算相似度矩阵**

相似度就是比较两个事物的相似性。一般通过计算事物的特征之间的距离，如果距离小，那么相似度大；如果距离大，那么相似度小。

·欧几里得距离（Eucledian Distance）:

是最常用的距离计算公式，衡量的是多维空间中各个点之间的绝对距离，当数据很稠密并且连续时，这是一种很好的计算方式。

·皮尔逊相关系数（Pearson Correlation Coefficient）:



又称相关相似性，通过Peason相关系数来度量两个用户的相似性。计算时，首先找到两个用户共同评分过的项目集，然后计算这两个向量的相关系数。

·Jaccard相似度：



Jaccard系数主要用于计算符号度量或布尔值度量的个体间的相似度，因为个体的特征属性都是由符号度量或者布尔值标识，因此无法衡量差异具体值的大小，只能获得“是否相同”这个结果，所以Jaccard系数只关心个体间共同具有的特征是否一致这个问题。

由于我们上一步得到的客户——菜品二元矩阵是0-1矩阵，因此这里我们选择用Jaccard系数来做相似度矩阵。

1. **生成推荐列表**

采用下式计算推荐算法中目标客户对所有菜品的感兴趣程度，并根据感兴趣程度给目标客户生成推荐列表。



其中P为目标客户对所有菜品的感兴趣程度，SIM为所有菜品之间的相似度，R代表了客户对物品的兴趣。在原始数据中只记录了客户用餐之后的订单，说明客户的行为是用餐与否，并没有类似电子商务网站上的购买、评分和评论等客户行为，因此R只存在0和1。

由于经过数据预处理后，剩下的数据都是完成订单的的顾客，因此R均为1。因此我们可以直接对菜品相似矩阵进行操作，当客户点了一道菜后，可以推荐与这道菜相似度最高的一道或者几道其他的菜品。

1. **模型评价**

为了验证我们构建的模型是否准确，即我们推荐的菜品，客户是否真的点过，构建以下评价指标。

1. **构建测试集客户ID——菜品字典**

构建测试集客户ID——菜品字典，若我们推荐给客户的菜品出现在这个字典相对应的value中，则表示我们的推荐菜品是正确的，否则则为错误的。

1. **构建评价指标，并分析推荐效果**

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：



从而得出我们构建的模型的效果是否良好。

1. **小结**

最终得到的菜品推荐准确率P=0.387，这个准确率并不算高，可能是由以下原因造成的:

1. 生成的推荐列表中，对应每个菜品推荐的菜品只有一道，降低了正确推荐的概率；
2. 数据收集不完整，因为本项目是基于类似的用户对菜品的推荐，而不考虑菜品本身的结构属性及烹饪方法，若添加这些因素可能会使推荐更加准确；
3. 推荐时没有考虑用户本身的一些差异，比如地域差异性导致的口味差异。