



# 肖智元

求职目标：数据分析师

男 21岁

1786032122@QQ.com

18527850256



## 教育背景

2020-09 ~ 2024-06

西华师范大学

人工智能（本科）

主修课程：科学计算与数据分析、数据库原理及应用、线性代数、概率论与数理统计、数据结构、离散数学、操作系统。



## 技能栈

- 熟练使用 MySQL，熟悉分组聚合、嵌套查询、连接查询等查询方式以及窗口函数的使用。
- 熟悉 Numpy、Pandas、Matplotlib 能够进行数据清洗、数据分析及可视化操作。
- 熟悉Requests、Selenium、BeautifulSoup4、Xpath、正则表达式，可以熟练运用于网络数据爬取。
- 熟悉 Excel 基本函数，Power Query 数据清洗工具以及数据分析工具。
- 熟练使用 PowerBI，熟悉报表和可视化面板制作流程。熟悉
- 了解 逻辑回归、随机森林、朴素贝叶斯等机器学习算法。



## 项目经验

**和鲸：基于 ECLAT 算法的超市篮子商品关联规则挖掘**

**相关工具：** ECLAT、Apriori、Pandas、Numpy、Matplotlib

**项目背景：** 超市和人们的生活息息相关，此项目是针对超市购物篮数据的关联规则挖掘，目的是找出顾客购买商品搭配关系，以便超市进行优化促销和商品搭配策略。数据集一共有 7501 个购物篮中的商品清单，共有 20 种商品，用于训练关联规则模型，代码中是用来 Apriori 算法对手护具集进行关联规则挖掘，并使用了 Pyecharts 库可视化展示了商品之间的关系图谱。

**项目过程：**

1. 读入数据：Market\_Basket\_Optimisation.csv (7501 条)
2. 数据预处理：使用循环遍历数据集中的每个交易记录，并将该记录中的每个商品作为字符串添加到新列表中，这个列表可以用于关联规则挖掘算法来找出频繁商品的组合。



### 3. 对数据集训练 Eclat 模型:

使用 Python 中的 apyori 库来执行算法 Apriori 来发现数据集中的频繁项集合关联规则

相关参数:

(1) min\_support 表示频繁项集的最小支持度, 即项集在所有交易记录中的出现频率, 设置值为 0.003 (即至少在所有交易记录中的 0.3% 出现)

(2) min\_confidence 表示规则的最小置信度, 设置值为 0.2, 表示值考虑置信度不低于 0.2 的频繁项集

(3) min\_lift 表示规则的最小提升度, 设置值为 3, 表示只考虑提升度不低于 3 的规则

(4) min\_length 和 max\_length 表示规则左右侧的限制长度, 设置值都为 2, 表示只考虑包含至少 2 个商品的规则。

4. 将结果组织到 Pandas 数据框架: 定义函数 inspect, 接受参数 results, 该参数包含从 Apriori 算法中生成的所有规则的结果对象, 将它们打包成一个列表, 最后返回该列表。并按支持度降序显示记录。

5. 可视化: 将每个商品看作一个节点, 节点 (商品) 之间建立不同粗细的关联线, 生成商品之间的关联规则图, 线越粗代表节点 (商品) 之间的关联性越强。

### 项目成果:

分析结果显示商品之间的关联性的强弱, 我们可以在结果中找到更优的营销方案, 比如胡椒和碎牛肉的关联度高, 我们就可以将两个商品配套销售, 提高两个商品的购买量的同时还可以增强客户满意度 (不用买了牛肉又去找胡椒)。在此方案中, 我们可以举一反三的运用不同行业的相关业务中。

## 毕业设计: 家居电商平台 RFM 模型的 K-means 顾客聚类分析

**相关工具:** RFM、K-means、Tableau Prep Builder、Pandas、Numpy、Matplotlib

### 项目背景:

在互联网的快速发展下, 电商行业客户消费体验和精准性营销是行业需要关注的问题, 此项目通过运用 RFM 模型原理, 针对家居产品电商平台 2021 年部分订单数据通过商业数据分析工具 Tableau 进行数据清洗和异常处理, 得到 RFM 模型的参数, 再通过 K-means 聚类算法对其进行客户分类, 在完成客户分类之后对不同类别的客户群提出相对应的营销策略和个性化服务, 最终为企业增加实际性收益和提升打造企业知名度和口碑。(数据来源: 天池)

### 项目过程:

1. 数据清洗: 运用 Tableau Prep Builder 分析工具针对订单数据进行异常处理和清洗得到有效数据, 并实现 RFM 模型, 基于 RFM 模型的现有条件进行改进并得到用户画像的基础操作。数据处理过程如下:

(1) 筛选出 2021 年的订单明细数据 (以下单为标准)

(2) 处理业务规则不吻合的数据 (异常数据)

(3) 平台字段统一格式 (App、Web、支付宝、微信)

(4) 渠道字段有空值 - 填空空值 (众数)

2. 确定 K 值: 综合运用 K 值的确定方法来确定 K 值, 在此项目中运用的方法有肘部法, 肘部法显示的 K 值为 3 或 4 都是不错的选择, 继续用轮廓系数法再对 K 值进行判断得到的值为 4, 那么结合两个方法最终的 K 取值为 4。

3. 算法分析: 在建设好的 RFM 模型基础上, 运用 K-means 算法 (K=4) 将用户分为 4 类, 之后我们根据不同类别的客户进行优化有针对性的营销, 为客户提供个性化服务, 不仅能促成订单量的产生, 还能提高客户的满意度和品牌忠诚度。



## 项目经验

### 项目成果：

优化营销策略：在得到的客户画像的条件后，我们将电商客户分成了 4 类，A 类客户为高价值客户，针对这类客户我们可以实施 VIP 服务和个性化服务；B 类客户为一般价值客户，针对他们的消费行为和习惯，找到能完全触发他们心动情境的商品或是营销策略，促进他们的下单量；C 类客户为低价值客户，我们可以采用优惠券或是礼品方式吸引他们购买商品，慢慢提高他们的订单量；D 类客户为流失客户，我们可以采取挽留措施，提高他们对企业品牌的印象，激发他们的购买欲望，从而能达到较好的订单数量。



## 校园经验

- 数学建模大赛省一等奖
- 蓝桥杯B组省二等奖
- 计算机设计大赛校二等奖



## 自我评价

具有学习能力强且勤于思考的特点。我积极总结并吸取学习和工作中的经验教训，能够深入思考问题并提出自己独特的见解。

工作中认真负责且富有创造力。我对于生活和工作中出现的现象，善于提出问题并进行深入思考，并在解决问题的过程中运用自己的创新方法。

在性格方面，我坚韧不拔且真诚友善。面对困难，我总是能够全力以赴寻找解决办法，并善于与他人进行交流和沟通。

我相信这些特点将使我在数据分析领域中取得良好的成绩。