实验二 Apriori算法

**学号：**

**姓名：**

1. 实验目的

掌握Apriori算法的原理

2. 实验要求

通过本实验应达到如下要求：

（1）理解频繁项集和关联规则

（2）熟练使用Python或其他工具实现Apriori算法

3. 实验器材

（1）计算机一台

（2）Python或其他编程工具

4. 实验内容

（1）关联分析相关基本概念有：

* 关联分析：关联分析是在大规模数据集中有目的的寻找关系的任务。
* 关联分析要寻找的关系：频繁项集、关联规则。
* 支持度：数据集中包含该项集的记录所占的比例。
* 置信度或可信度：定义为条件概率。
* 频繁项集：经常一起出现的项目的集合,定义为支持度大于某一阈值的集合。
* 关联规则：置信度大于一定阈值的关系。

**（2）简述Apriori算法的原理：**

（3）Apriori算法的两个输入参数是数据集和最小支持度（阈值）。其流程如下：

* 生成单个项的所有项集，遍历所有交易事件，筛选出单个项的频繁项集。
* 对于包含k个项的频繁项集，两两组合生成k+1项项集，删除非频繁项集，获得k+1频繁项集直到算法收敛。
* 返回频繁项集。

（3）生成单项候选集，**代码：**

（4）计算候选集的支持度，选出k项频繁集，**代码：**

（5）由k项频繁集生成k+1项候选集，**代码：**

（6）**生成频繁项集代码：**

（7）**输入数据集（交易事件）：**

**输入最小支持度阈值：**

**输出频繁项集结果：**

**5. 实验心得**