

Capstone Project Report

Name: Pooja Ramdas Kadam

Course: AI & ML (Batch - 4)

Problem Statement

Apriori is a statistical algorithm for implementing associate rule mining, that primarily relies on three components: Life, Support and Confidence. Using this algorithm try to find the rules that describe the relation between each of the products that were brought by the customers

Prerequisites

Along with Python below packages needed to be installed

Apyori

Pandas

Dataset Used

<https://drive.google.com/file/d/1y5DYn0dGoSbC22xowBq2d4po6h1JxcTQ/view>

Implementation

Import required libraries and load data

```
In [1]: import numpy as np
import pandas as pd
from apyori import apriori
```

Load data

```
In [2]: store_data = pd.read_csv('store_data.csv', header=None)
store_data.head()
```

Out[2]:

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	
0	shrimp	almonds	avocado	vegetables mix	green grapes	whole wheat flour	yams	cottage cheese	energy drink	tomato juice	low fat yogurt	green tea	honey	salad	mineral water	salmon	antioxydant juice	frozen smoothie	spinach
1	burgers	meatballs	eggs	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	chutney	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	turkey	avocado	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	mineral water	milk	energy bar	whole wheat rice	green tea	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

Apply Apriori

```
In [18]: results = list(apriori(store_data_list, min_support=0.005))
print(len(results))
```

1447

In [19]: results

```
Out[19]: [RelationRecord(items=frozenset({'almonds'}), support=0.020397280362618318, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'almonds'}), confidence=0.020397280362618318, lift=1.0)]),
RelationRecord(items=frozenset({'antioxydant juice'}), support=0.008932142381015865, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'antioxydant juice'}), confidence=0.008932142381015865, lift=1.0)]),
RelationRecord(items=frozenset({'avocado'}), support=0.03332888948140248, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'avocado'}), confidence=0.03332888948140248, lift=1.0)]),
RelationRecord(items=frozenset({'bacon'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'bacon'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'barbecue sauce'}), support=0.010798560191974404, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'barbecue sauce'}), confidence=0.010798560191974404, lift=1.0)]),
RelationRecord(items=frozenset({'black tea'}), support=0.014264764698040262, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'black tea'}), confidence=0.014264764698040262, lift=1.0)]),
RelationRecord(items=frozenset({'blueberries'}), support=0.009198773496867084, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'blueberries'}), confidence=0.009198773496867084, lift=1.0)]),
RelationRecord(items=frozenset({'body spray'}), support=0.011465137981602452, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'body spray'}), confidence=0.011465137981602452, lift=1.0)]),
RelationRecord(items=frozenset({'brownies'}), support=0.03372883615517931, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'brownies'}), confidence=0.03372883615517931, lift=1.0)]),
RelationRecord(items=frozenset({'cottage cheese'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'cottage cheese'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'energy drink'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'energy drink'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'green grapes'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'green grapes'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'green tea'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'green tea'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'honey'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'honey'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'low fat yogurt'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'low fat yogurt'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'meatballs'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'meatballs'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'mineral water'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'mineral water'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'milk'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'milk'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'salmon'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'salmon'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'shrimp'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'shrimp'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'spinach'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'spinach'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'turkey'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'turkey'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'whole wheat flour'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'whole wheat flour'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'whole wheat rice'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'whole wheat rice'}), confidence=0.008665511265164644, lift=1.0)]),
RelationRecord(items=frozenset({'yams'}), support=0.008665511265164644, ordered_statistics=[OrderedStatistic(items_base=frozenset(), items_add=frozenset({'yams'}), confidence=0.008665511265164644, lift=1.0)])]
```