

Chapter 6 - Other Popular Machine Learning Methods

Segment 1 - Association Rule Mining Using Apriori Algorithm

▼ Import the required libraries

```
! pip install mlxtend
```

```
Collecting mlxtend
```

```
  Downloading https://files.pythonhosted.org/packages/52/04/c362f34f666f0ddc7cf593805e64d64fa670ed96fd9302e68549dd48287d/mlxtend-
Requirement already satisfied: matplotlib>=3.0.0 in c:\users\lillian\anaconda3\lib\site-packages (from mlxtend) (3.1.0)
Requirement already satisfied: numpy>=1.16.2 in c:\users\lillian\anaconda3\lib\site-packages (from mlxtend) (1.16.4)
Requirement already satisfied: joblib>=0.13.2 in c:\users\lillian\anaconda3\lib\site-packages (from mlxtend) (0.13.2)
Requirement already satisfied: setuptools in c:\users\lillian\anaconda3\lib\site-packages (from mlxtend) (41.0.1)
Requirement already satisfied: pandas>=0.24.2 in c:\users\lillian\anaconda3\lib\site-packages (from mlxtend) (0.24.2)
Requirement already satisfied: scipy>=1.2.1 in c:\users\lillian\anaconda3\lib\site-packages (from mlxtend) (1.2.1)
Requirement already satisfied: scikit-learn>=0.20.3 in c:\users\lillian\anaconda3\lib\site-packages (from mlxtend) (0.21.2)
Requirement already satisfied: cycycler>=0.10 in c:\users\lillian\anaconda3\lib\site-packages (from matplotlib>=3.0.0->mlxtend) (0.
Requirement already satisfied: kiwisolver>=1.0.1 in c:\users\lillian\anaconda3\lib\site-packages (from matplotlib>=3.0.0->mlxtend)
Requirement already satisfied: pyparsing!=2.0.4,!=2.1.2,!=2.1.6,>=2.0.1 in c:\users\lillian\anaconda3\lib\site-packages (from matplotlib>=3.0.0->mlxtend)
Requirement already satisfied: python-dateutil>=2.1 in c:\users\lillian\anaconda3\lib\site-packages (from matplotlib>=3.0.0->mlxtend)
Requirement already satisfied: pytz>=2011k in c:\users\lillian\anaconda3\lib\site-packages (from pandas>=0.24.2->mlxtend) (2019.1
Requirement already satisfied: six in c:\users\lillian\anaconda3\lib\site-packages (from cycycler>=0.10->matplotlib>=3.0.0->mlxtend)
Installing collected packages: mlxtend
Successfully installed mlxtend-0.17.0
```

```
import pandas as pd
from mlxtend.frequent_patterns import apriori
from mlxtend.frequent_patterns import association_rules
```

▼ Data Format

```
address = 'C:/Users/Lillian/Desktop/ExerciseFiles/Data/groceries.csv'
data = pd.read_csv(address)
```

```
data.head()
```

	1	2	3	4	5	6	7	8	9
0	citrus fruit	semi-finished bread	margarine	ready soups	NaN	NaN	NaN	NaN	NaN
1	tropical fruit	yogurt	coffee	NaN	NaN	NaN	NaN	NaN	NaN
2	whole milk	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	pip fruit	yogurt	cream cheese	meat spreads	NaN	NaN	NaN	NaN	NaN
4	other vegetables	whole milk	condensed milk	long life bakery product	NaN	NaN	NaN	NaN	NaN

▼ Data Coversion

```
basket_sets = pd.get_dummies(data)
```

```
basket_sets.head()
```

	1_Instant food products	1_UHT- milk	1_artif. sweetener	1_baby cosmetics	1_bags	1_baking powder	1_bathroom cleaner	1_beef	1_berries	1_beverage
0	0	0	0	0	0	0	0	0	0	
1	0	0	0	0	0	0	0	0	0	
2	0	0	0	0	0	0	0	0	0	
3	0	0	0	0	0	0	0	0	0	
4	0	0	0	0	0	0	0	0	0	

5 rows × 1113 columns

▼ Support Calculation

```
apriori(basket_sets, min_support=0.02)
```

	support	itemsets
0	0.030421	(7)
1	0.034951	(17)
2	0.029126	(23)
3	0.049191	(26)
4	0.064401	(47)
5	0.044660	(83)
6	0.024272	(90)
7	0.040453	(92)
8	0.038835	(99)
9	0.033981	(100)
10	0.076052	(105)
11	0.028803	(111)
12	0.044984	(123)
13	0.073463	(130)
14	0.022977	(131)

```
apriori(basket_sets, min_support=0.02, use_colnames=True)
```

	support	itemsets
0	0.030421	(1_beef)
1	0.034951	(1_canned beer)
2	0.029126	(1_chicken)
3	0.049191	(1_citrus fruit)
4	0.064401	(1_frankfurter)
5	0.044660	(1_other vegetables)
6	0.024272	(1_pip fruit)
7	0.040453	(1_pork)
8	0.038835	(1_rolls/buns)
9	0.033981	(1_root vegetables)
10	0.076052	(1_sausage)
11	0.028803	(1_soda)
12	0.044984	(1_tropical fruit)
13	0.073463	(1_whole milk)
14	0.022977	(1_yogurt)
15	0.028803	(2_citrus fruit)
16	0.058900	(2_other vegetables)
17	0.022977	(2_pip fruit)
18	0.040129	(2_rolls/buns)
19	0.036893	(2_root vegetables)
20	0.031068	(2_soda)
21	0.034628	(2_tropical fruit)
22	0.062136	(2_whole milk)
23	0.028479	(2_yogurt)
24	0.045955	(3_other vegetables)

25	0.033010	(3_rolls/buns)
26	0.024272	(3_soda)
27	0.057929	(3_whole milk)

```
df = basket_sets
```

```
frequent_itemsets = apriori(basket_sets, min_support=0.002, use_colnames=True)
```

```
frequent_itemsets['length'] = frequent_itemsets['itemsets'].apply(lambda x: len(x))
```

```
frequent_itemsets
```

	support	itemsets	length
0	0.006472	(1_UHT-milk)	1
1	0.030421	(1_beef)	1
2	0.011974	(1_berries)	1
3	0.008414	(1_beverages)	1
4	0.014887	(1_bottled beer)	1
5	0.019094	(1_bottled water)	1
6	0.006149	(1_brown bread)	1
7	0.009385	(1_butter)	1
8	0.004854	(1_butter milk)	1
9	0.034951	(1_canned beer)	1
10	0.002589	(1_cat food)	1
11	0.002265	(1_chewing gum)	1
12	0.029126	(1_chicken)	1
13	0.004854	(1_chocolate)	1
14	0.049191	(1_citrus fruit)	1
15	0.009709	(1_coffee)	1
16	0.004207	(1_cream cheese)	1
17	0.011650	(1_curd)	1
18	0.007767	(1_dessert)	1
19	0.002265	(1_detergent)	1
20	0.006149	(1_domestic eggs)	1
21	0.004531	(1_finished products)	1
22	0.064401	(1_frankfurter)	1
23	0.002265	(1_frozen fish)	1
24	0.003883	(1_frozen meals)	1

25	0.004207	(1_frozen vegetables)	1
26	0.003883	(1_fruit/vegetable juice)	1
27	0.006796	(1_grapes)	1
28	0.019741	(1_ham)	1
29	0.016828	(1_hamburger meat)	1
...
819	0.002265	(8_yogurt, 9_whipped/sour cream)	2
820	0.002589	(1_beef, 3_other vegetables, 2_root vegetables)	3
821	0.002589	(3_whole milk, 2_other vegetables, 1_chicken)	3
822	0.002589	(3_whole milk, 2_other vegetables, 1_citrus fr...	3
823	0.003236	(2_tropical fruit, 3_pip fruit, 1_citrus fruit)	3
824	0.002589	(3_other vegetables, 4_whole milk, 1_citrus fr...	3
825	0.002265	(5_other vegetables, 1_frankfurter, 6_whole milk)	3
826	0.002265	(3_other vegetables, 4_whole milk, 1_pork)	3
827	0.003560	(3_whole milk, 2_other vegetables, 1_root vege...	3
828	0.002589	(2_rolls/buns, 1_sausage, 3_soda)	3
829	0.002265	(1_sausage, 3_other vegetables, 4_whole milk)	3
830	0.002265	(4_other vegetables, 1_sausage, 5_whole milk)	3
831	0.002913	(1_tropical fruit, 3_whole milk, 2_other veget...	3
832	0.002265	(4 other vegetables, 2 citrus fruit, 5 whole m...	3

frequent_itemsets[frequent_itemsets['length'] >= 3]



	support	itemsets	length
820	0.002589	(1_beef, 3_other vegetables, 2_root vegetables)	3
821	0.002589	(3_whole milk, 2_other vegetables, 1_chicken)	3
822	0.002589	(3_whole milk, 2_other vegetables, 1_citrus fr...	3
823	0.003236	(2_tropical fruit, 3_pip fruit, 1_citrus fruit)	3
824	0.002589	(3_other vegetables, 4_whole milk, 1_citrus fr...	3
825	0.002265	(5_other vegetables, 1_frankfurter, 6_whole milk)	3
826	0.002265	(3_other vegetables, 4_whole milk, 1_pork)	3
827	0.003560	(3_whole milk, 2_other vegetables, 1_root vege...	3
828	0.002589	(2_rolls/buns, 1_sausage, 3_soda)	3
829	0.002265	(1_sausage, 3_other vegetables, 4_whole milk)	3
830	0.002265	(4_other vegetables, 1_sausage, 5_whole milk)	3
831	0.002913	(1_tropical fruit, 3_whole milk, 2_other veget...	3
832	0.002265	(4_other vegetables, 2_citrus fruit, 5_whole m...	3
833	0.002265	(4_butter, 3_whole milk, 2_other vegetables)	3
834	0.003560	(4_curd, 3_whole milk, 2_other vegetables)	3
835	0.003883	(4_yogurt, 3_whole milk, 2_other vegetables)	3
836	0.002265	(6_rolls/buns, 3_whole milk, 2_other vegetables)	3
837	0.003236	(3_other vegetables, 2_pip fruit, 4_whole milk)	3
838	0.005825	(3_other vegetables, 4_whole milk, 2_root vege...	3
839	0.002265	(2_tropical fruit, 4_other vegetables, 3_pip f...	3
840	0.003560	(5_butter, 3_other vegetables, 4_whole milk)	3
841	0.002913	(3_other vegetables, 5_yogurt, 4_whole milk)	3
842	0.003560	(6_yogurt, 3_other vegetables, 4_whole milk)	3
843	0.002265	(5_other vegetables, 3_pip fruit, 4_root vege...	3
844	0.002265	(5_other vegetables, 3_pip fruit, 6_whole milk)	3

▼ Association Rules

```
047 0.000200 (0_other vegetables, 0_whole milk, 4_root vege...
```

▼ Confidence

```
rules = association_rules(frequent_itemsets, metric="confidence", min_threshold=0.5)
rules.head()
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(2_sausage)	(1_frankfurter)	0.011327	0.064401	0.011327	1.000000	15.527638	0.010597	inf
1	(7_pastry)	(1_frankfurter)	0.005178	0.064401	0.002589	0.500000	7.763819	0.002256	1.871197
2	(2_ham)	(1_sausage)	0.007120	0.076052	0.004531	0.636364	8.367505	0.003989	2.540858
3	(2_meat)	(1_sausage)	0.006796	0.076052	0.004854	0.714286	9.392097	0.004338	3.233819
4	(3_beef)	(1_sausage)	0.004854	0.076052	0.002589	0.533333	7.012766	0.002220	1.979889

▼ Lift

```
rules = association_rules(frequent_itemsets, metric="lift", min_threshold=1)
rules.head()
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conviction
0	(1_beef)	(2_citrus fruit)	0.030421	0.028803	0.005502	0.180851	6.278986	0.004625	1.18561
1	(2_citrus fruit)	(1_beef)	0.028803	0.030421	0.005502	0.191011	6.278986	0.004625	1.19850
2	(1_beef)	(2_other vegetables)	0.030421	0.058900	0.003236	0.106383	1.806173	0.001444	1.05313

▼ Lift and Confidence

```
rules[(rules['lift'] >= 5) & (rules['confidence']>= 0.5)]
```

	antecedents	consequents	antecedent support	consequent support	support	confidence	lift	leverage	conv
93	(2_sausage)	(1_frankfurter)	0.011327	0.064401	0.011327	1.000000	15.527638	0.010597	
136	(7_pastry)	(1_frankfurter)	0.005178	0.064401	0.002589	0.500000	7.763819	0.002256	1.
238	(2_ham)	(1_sausage)	0.007120	0.076052	0.004531	0.636364	8.367505	0.003989	2.
243	(2_meat)	(1_sausage)	0.006796	0.076052	0.004854	0.714286	9.392097	0.004338	3.
258	(3_beef)	(1_sausage)	0.004854	0.076052	0.002589	0.533333	7.012766	0.002220	1.
351	(2_frozen vegetables)	(1_whole milk)	0.005825	0.073463	0.002913	0.500000	6.806167	0.002485	1.
478	(3_beef)	(2_pork)	0.004854	0.015534	0.002913	0.600000	38.625000	0.002837	2.
541	(3_curd)	(2_whole milk)	0.009709	0.062136	0.005825	0.600000	9.656250	0.005222	2.
575	(4_tropical fruit)	(3_citrus fruit)	0.002913	0.010356	0.002265	0.777778	75.104167	0.002235	4.
583	(4_whole milk)	(3_other vegetables)	0.033333	0.045955	0.018123	0.543689	11.830986	0.016591	2.
606	(4_grapes)	(3_pip fruit)	0.003560	0.016828	0.002265	0.636364	37.814685	0.002205	2.
618	(8_yogurt)	(3_pip fruit)	0.004854	0.016828	0.002589	0.533333	31.692308	0.002507	2.
626	(4_onions)	(3_root vegetables)	0.003883	0.018770	0.002265	0.583333	31.077586	0.002192	2.
641	(4_pip fruit)	(3_tropical fruit)	0.005825	0.014239	0.003560	0.611111	42.916667	0.003477	2.
650	(4_butter)	(3_whole milk)	0.008414	0.057929	0.005825	0.692308	11.951010	0.005338	3.
652	(4_curd)	(3_whole milk)	0.013592	0.057929	0.008414	0.619048	10.686353	0.007627	2.
693	(5_whole milk)	(4_other vegetables)	0.014239	0.020712	0.009061	0.636364	30.724432	0.008767	2.
714	(5_butter)	(4_whole milk)	0.008414	0.033333	0.005825	0.692308	20.769231	0.005545	3.
718	(5_pip fruit)	(4_whole milk)	0.005825	0.033333	0.003560	0.611111	12.000000	0.003333	2.

716	(5_curd)	(4_curd, 5_milk)	0.005825	0.0333333	0.003560	0.611111	18.333333	0.003366	2.0
753	(5_onions)	(6_other vegetables)	0.002913	0.007120	0.002265	0.777778	109.242424	0.002245	4.0
754	(5_other vegetables)	(6_whole milk)	0.012621	0.009385	0.007443	0.589744	62.838196	0.007325	2.0
755	(6_whole milk)	(5_other vegetables)	0.009385	0.012621	0.007443	0.793103	62.838196	0.007325	4.0
757	(7_butter)	(5_other vegetables)	0.004207	0.012621	0.002589	0.615385	48.757396	0.002536	2.0
759	(7_yogurt)	(5_other vegetables)	0.004531	0.012621	0.002913	0.642857	50.934066	0.002855	2.0
763	(5_root vegetables)	(6_other vegetables)	0.004531	0.007120	0.002589	0.571429	80.259740	0.002557	2.0
776	(6_domestic eggs)	(7_rolls/buns)	0.006149	0.011327	0.003236	0.526316	46.466165	0.003167	2.0
778	(6_other vegetables)	(7_whole milk)	0.007120	0.004207	0.003883	0.545455	129.650350	0.003854	2.0
779	(7_whole milk)	(6_other vegetables)	0.004207	0.007120	0.003883	0.923077	129.650350	0.003854	12.0
784	(7_brown rolls/buns)	(6_rolls/buns)	0.002560	0.016184	0.002013	0.818182	50.562626	0.002855	5.0
785	(7_butter)	(6_whole milk)	0.004207	0.009385	0.002913	0.692308	73.766578	0.002873	3.0
...
870	(2_citrus fruit, 5_whole milk)	(4_other vegetables)	0.003236	0.020712	0.002265	0.700000	33.796875	0.002198	3.0
875	(4_butter, 2_other vegetables)	(3_whole milk)	0.002265	0.057929	0.002265	1.000000	17.262570	0.002134	...
881	(4_curd, 2_other vegetables)	(3_whole milk)	0.003883	0.057929	0.003560	0.916667	15.824022	0.003335	11.0
...	(4_yogurt, 2_citrus fruit)	(6_citrus fruit)	0.002560	0.016184	0.002013	0.818182	50.562626	0.002855	5.0

886	(1_yogurt, 3_whole milk)	(2_other vegetables)	0.007767	0.058900	0.003883	0.500000	8.489011	0.003426	1.0
887	(4_yogurt, 2_other vegetables)	(3_whole milk)	0.005502	0.057929	0.003883	0.705882	12.185343	0.003565	3.0
892	(6_rolls/buns, 3_whole milk)	(2_other vegetables)	0.003560	0.058900	0.002265	0.636364	10.804196	0.002056	2.0