

**МИНОБРНАУКИ РОССИИ**  
**САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ**  
**ЭЛЕКТРОТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**  
**«ЛЭТИ» ИМ. В.И. УЛЬЯНОВА (ЛЕНИНА)**  
**Кафедра вычислительной техники**

**ОТЧЕТ**  
**по лабораторной работе №4**  
**по дисциплине «Машинное обучение»**  
**Тема: «Ассоциативный анализ»**

Студенты гр. 6307

\_\_\_\_\_

Ходос А.А.

Преподаватель

\_\_\_\_\_

Жангиров Т.Р.

Санкт-Петербург

2020

## Цель работы

Ознакомиться с методами ассоциативного анализа из библиотеки MLxtend.

## Ход работы

### 1. Загрузка данных

Был загружен датасет, данные в котором представляют собой информацию о купленных вместе товарах. Датасет представлен на рисунках 1-2. Видно, что датасет содержит значения nan.

	Item(s)	Item 1	Item 2	Item 3	Item 4	Item 5
0	4	citrus fruit	semi-finished b...	margarine	ready soups	nan
1	3	tropical fruit	yogurt	coffee	nan	nan
2	1	whole milk	nan	nan	nan	nan
3	4	pip fruit	yogurt	cream cheese	meat spreads	nan
4	4	other vegetables	whole milk	condensed milk	long life bakery...	nan
5	5	whole milk	butter	yogurt	rice	abrasive cleaner
6	1	rolls/buns	nan	nan	nan	nan
7	5	other vegetables	UHT-milk	rolls/buns	bottled beer	liquor (appetizer)
8	1	potted plants	nan	nan	nan	nan
9	2	whole milk	cereals	nan	nan	nan
10	5	tropical fruit	other vegetables	white bread	bottled water	chocolate

Рисунок 1 – Загруженный датасет (1)

	Item(s)	Item 1	Item 2	Item 3	Item 4	Item 5
9824	4	ice cream	long life bakery...	specialty choco...	specialty bar	nan
9825	10	chicken	hamburger meat	citrus fruit	root vegetables	other vegetables
9826	6	citrus fruit	herbs	other vegetables	dessert	sugar
9827	9	frankfurter	tropical fruit	other vegetables	whole milk	frozen meals
9828	7	sausage	butter	rolls/buns	pickled vegeta...	soda
9829	7	tropical fruit	other vegetables	domestic eggs	zwieback	ketchup
9830	17	sausage	chicken	beef	hamburger meat	citrus fruit
9831	1	cooking chocol...	nan	nan	nan	nan
9832	10	chicken	citrus fruit	other vegetables	butter	yogurt
9833	4	semi-finished b...	bottled water	soda	bottled beer	nan
9834	5	chicken	tropical fruit	other vegetables	vinegar	shopping bags

Рисунок 2 – Загруженный датасет (2)

Были удалены значения nap в нем. Список всех 169 товаров представлен далее:

```
{'grapes', 'misc. beverages', 'curd cheese', 'sparkling wine', 'organic products',  
'zwieback', 'fish', 'preservation products', 'beverages', 'canned beer', 'curd',  
'pork', 'coffee', 'sweet spreads', 'frozen dessert', 'onions', 'cream', 'potted  
plants', 'napkins', 'soda', 'cocoa drinks', 'other vegetables', 'roll products',  
'hamburger meat', 'liquor', 'berries', 'turkey', 'frozen meals', 'finished products',  
'rubbing alcohol', 'frankfurter', 'dishes', 'UHT-milk', 'abrasive cleaner',  
'fruit/vegetable juice', 'pip fruit', 'male cosmetics', 'chewing gum', 'cleaner',  
'toilet cleaner', 'flower (seeds)', 'sound storage medium', 'canned fish', 'rum',  
'butter', 'bottled beer', 'cat food', 'processed cheese', 'shopping bags', 'cereals',  
'nut snack', 'frozen chicken', 'pasta', 'liquor (appetizer)', 'dessert', 'sauces',  
'pudding powder', 'mustard', 'newspapers', 'organic sausage', 'beef', 'salty snack',  
'instant coffee', 'yogurt', 'brown bread', 'cling film/bags', 'mayonnaise', 'light  
bulbs', 'syrup', 'photo/film', 'tea', 'specialty vegetables', 'ice cream', 'Instant  
food products', 'hygiene articles', 'prosecco', 'liver loaf', 'bathroom cleaner',  
'pickled vegetables', 'ham', 'skin care', 'spices', 'tidbits', 'flower  
soil/fertilizer', 'meat', 'sugar', 'hair spray', 'butter milk', 'citrus fruit',  
'popcorn', 'herbs', 'seasonal products', 'cake bar', 'brandy', 'dog food', 'condensed  
milk', 'potato products', 'white wine', 'whisky', 'female sanitary products', 'semi-  
finished bread', 'specialty fat', 'specialty bar', 'domestic eggs', 'salt', 'soap',  
'frozen fish', 'jam', 'kitchen utensil', 'specialty chocolate', 'rice', 'bottled  
water', 'red/blush wine', 'baby food', 'softener', 'honey', 'make up remover', 'flour',  
'packaged fruit/vegetables', 'waffles', 'canned vegetables', 'cream cheese', 'pastry',  
'snack products', 'chocolate marshmallow', 'soft cheese', 'cookware', 'chicken',  
'artif. sweetener', 'whole milk', 'ketchup', 'bags', 'margarine', 'nuts/prunes', 'white  
bread', 'detergent', 'liqueur', 'spread cheese', 'candy', 'rolls/buns', 'root  
vegetables', 'frozen fruits', 'vinegar', 'frozen potato products', 'hard cheese',  
'whipped/sour cream', 'dish cleaner', 'long life bakery product', 'dental care', 'baby  
cosmetics', 'kitchen towels', 'canned fruit', 'ready soups', 'baking powder',  
'candles', 'sliced cheese', 'sausage', 'house keeping products', 'oil', 'chocolate',  
'tropical fruit', 'decalcifier', 'cooking chocolate', 'pet care', 'soups', 'salad  
dressing', 'frozen vegetables', 'meat spreads', 'specialty cheese'}
```

## 2. FPGrowth и FPMaх

Преобразовав данные к виду, удобному для анализа, был проведен ассоциативный анализ с помощью алгоритма FPGrowth при уровне поддержки 0.03. Результаты представлены на рисунках 3-4. Как видно, получилось 63 правило с максимальной длиной 2.

	support	itemsets	length
0	0.08276563294356888	frozenset({'citrus fruit'})	1
1	0.05856634468734113	frozenset({'margarine'})	1
2	0.13950177935943062	frozenset({'yogurt'})	1
3	0.10493136756481952	frozenset({'tropical fruit'})	1
4	0.05805795627859685	frozenset({'coffee'})	1
5	0.25551601423487547	frozenset({'whole milk'})	1
6	0.07564819522114896	frozenset({'pip fruit'})	1
7	0.03965429588205389	frozenset({'cream cheese'})	1
8	0.1934926283680732	frozenset({'other vegetables'})	1
9	0.037417386883579054	frozenset({'long life bakery produc...	1
10	0.05541433655312659	frozenset({'butter'})	1

Рисунок 3— Результат FPGrowth (1)

	support	itemsets	length
52	0.042602948652770714	frozenset({'rolls/buns', 'other vege...	2
53	0.05663446873411286	frozenset({'rolls/buns', 'whole milk'})	2
54	0.03436705643111337	frozenset({'bottled water', 'whole ...	2
55	0.03833248601931876	frozenset({'rolls/buns', 'soda'})	2
56	0.040061006609049314	frozenset({'whole milk', 'soda'})	2
57	0.032740213523131674	frozenset({'other vegetables', 'sod...	2
58	0.033248601931875954	frozenset({'pastry', 'whole milk'})	2
59	0.047381799694966954	frozenset({'other vegetables', 'root...	2
60	0.048906964921199794	frozenset({'whole milk', 'root vege...	2
61	0.030604982206405694	frozenset({'rolls/buns', 'sausage'})	2
62	0.032231825114387394	frozenset({'whipped/sour cream', '...	2

Рисунок 4— Результат FPGrowth (2)

Границы поддержки для набора из 1 элемента: [ 0.0304 ; 0.2555 ]

Границы поддержки для набора из 2 элементов: [ 0.0301 ; 0.0748 ]

Проведем аналогичный анализ алгоритма FPMaх. Результаты представлены на рисунках 5-6. Были получены 50 правил.

	support	itemsets	length
0	0.03040162684290798	frozenset({'specialty chocolate'})	1
1	0.031011692933401117	frozenset({'onions'})	1
2	0.03294356888662939	frozenset({'hygiene articles'})	1
3	0.033248601931875954	frozenset({'berries'})	1
4	0.033248601931875954	frozenset({'hamburger meat'})	1
5	0.03345195729537367	frozenset({'UHT-milk'})	1
6	0.03385866802236909	frozenset({'sugar'})	1
7	0.03711235383833249	frozenset({'dessert'})	1
8	0.037417386883579054	frozenset({'long life bakery produc...	1
9	0.03782409761057448	frozenset({'salty snack'})	1
10	0.038434163701067614	frozenset({'waffles'})	1

Рисунок 5 – Результат FPMaх (1)

	support	itemsets	length
39	0.048906964921199794	frozenset({'whole milk', 'root vege...	2
40	0.03436705643111337	frozenset({'bottled water', 'whole ...	2
41	0.03436705643111337	frozenset({'rolls/buns', 'yogurt'})	2
42	0.04341637010676157	frozenset({'other vegetables', 'yog...	2
43	0.05602440264361973	frozenset({'yogurt', 'whole milk'})	2
44	0.032740213523131674	frozenset({'other vegetables', 'sod...	2
45	0.03833248601931876	frozenset({'rolls/buns', 'soda'})	2
46	0.040061006609049314	frozenset({'whole milk', 'soda'})	2
47	0.042602948652770714	frozenset({'rolls/buns', 'other vege...	2
48	0.05663446873411286	frozenset({'rolls/buns', 'whole milk'})	2
49	0.07483477376715811	frozenset({'other vegetables', 'who...	2

Рисунок 6 – Результат FPMaх (2)

Границы поддержки для набора из 1 элемента: [ 0.0304 ; 0.0985 ]

Границы поддержки для набора из 2 элементов: [ 0.0301 ; 0.0748 ]

FPGrowth извлекает все возможные наборы, поддержка которых удовлетворяет заданному минимуму, в то время как FPMax извлекает лишь максимальные наборы.

Была построена гистограмма, на которой товары упорядочены по уменьшению частоты. Результат, на котором отображены 10 самых встречаемых товаров представлен на рисунке 7. Полученная гистограмма коррелирует с результатами частотного анализа – если взять все отсортированные наборы длиной 1, то получится в точности этот же набор данных.

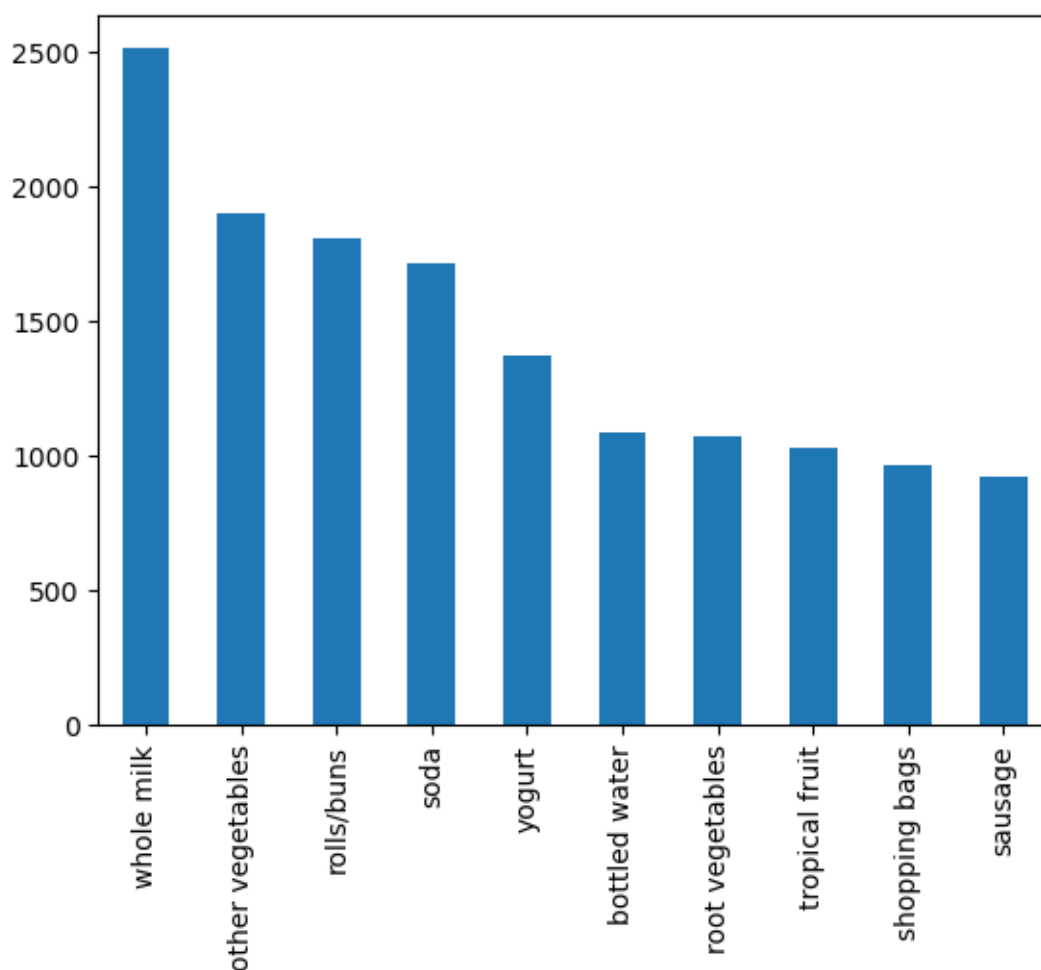


Рисунок 7 – гистограмма часто встречающихся товаров

Преобразуем набор данных, чтобы он содержал ограниченный набор товаров. Проведя анализ FPGrowth и FPMax для нового набора данных, получили результаты, представленные на рисунках 8-9 и 10-11, соответственно.

	support	itemsets	length
0	0.08276563294356888	frozenset({'citrus fruit'})	1
1	0.13950177935943062	frozenset({'yogurt'})	1
2	0.10493136756481952	frozenset({'tropical fruit'})	1
3	0.25551601423487547	frozenset({'whole milk'})	1
4	0.1934926283680732	frozenset({'other vegetables'})	1
5	0.18393492628368074	frozenset({'rolls/buns'})	1
6	0.08052872394509406	frozenset({'bottled beer'})	1
7	0.11052364006100661	frozenset({'bottled water'})	1
8	0.17437722419928825	frozenset({'soda'})	1
9	0.08896797153024912	frozenset({'pastry'})	1
10	0.10899847483477376	frozenset({'root vegetables'})	1

Рисунок 8 – Результат FPGrowth (1)

	support	itemsets	length
23	0.042602948652770714	frozenset({'rolls/buns', 'other vege...	2
24	0.05663446873411286	frozenset({'rolls/buns', 'whole milk'})	2
25	0.03436705643111337	frozenset({'bottled water', 'whole ...	2
26	0.03833248601931876	frozenset({'rolls/buns', 'soda'})	2
27	0.040061006609049314	frozenset({'soda', 'whole milk'})	2
28	0.032740213523131674	frozenset({'soda', 'other vegetable...	2
29	0.033248601931875954	frozenset({'pastry', 'whole milk'})	2
30	0.047381799694966954	frozenset({'root vegetables', 'other...	2
31	0.048906964921199794	frozenset({'root vegetables', 'whol...	2
32	0.030604982206405694	frozenset({'rolls/buns', 'sausage'})	2
33	0.032231825114387394	frozenset({'whipped/sour cream', '...	2

Рисунок 9 – Результат FPGrowth (2)

	support	itemsets	length
0	0.05765124555160142	frozenset({'pork'})	1
1	0.032231825114387394	frozenset({'whipped/sour cream', '...	2
2	0.07768174885612608	frozenset({'canned beer'})	1
3	0.08052872394509406	frozenset({'bottled beer'})	1
4	0.030503304524656837	frozenset({'citrus fruit', 'whole milk'})	2
5	0.033248601931875954	frozenset({'pastry', 'whole milk'})	2
6	0.030604982206405694	frozenset({'rolls/buns', 'sausage'})	2
7	0.09852567361464158	frozenset({'shopping bags'})	1
8	0.035892221657346214	frozenset({'other vegetables', 'trop...	2
9	0.04229791560752415	frozenset({'tropical fruit', 'whole m...	2
10	0.047381799694966954	frozenset({'root vegetables', 'other...	2

Рисунок 10 – Результат FPMax (1)

	support	itemsets	length
11	0.048906964921199794	frozenset({'root vegetables', 'whol...	2
12	0.03436705643111337	frozenset({'bottled water', 'whole ...	2
13	0.03436705643111337	frozenset({'yogurt', 'rolls/buns'})	2
14	0.04341637010676157	frozenset({'yogurt', 'other vegetab...	2
15	0.05602440264361973	frozenset({'yogurt', 'whole milk'})	2
16	0.032740213523131674	frozenset({'soda', 'other vegetable...	2
17	0.03833248601931876	frozenset({'rolls/buns', 'soda'})	2
18	0.040061006609049314	frozenset({'soda', 'whole milk'})	2
19	0.042602948652770714	frozenset({'rolls/buns', 'other vege...	2
20	0.05663446873411286	frozenset({'rolls/buns', 'whole milk'})	2
21	0.07483477376715811	frozenset({'other vegetables', 'who...	2

Рисунок 11 – Результат FPMax (2)

Были построены графики изменения количества получаемых правил от уровня поддержки, представленные на рисунках 12-13. По графикам видно, что при уменьшении максимальной длины наборов, график начинает убывать медленнее.



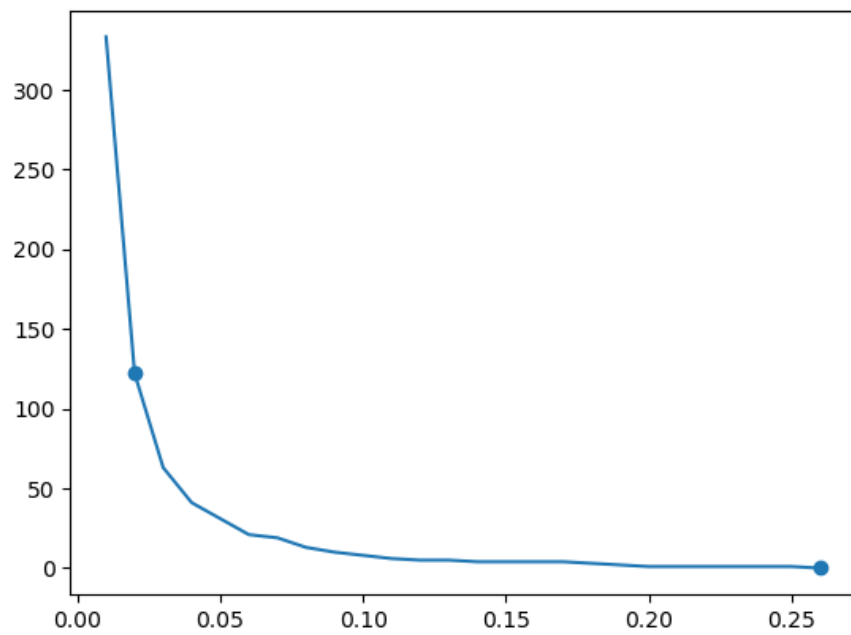


Рисунок 13 – График графики изменения количества получаемых правил в FPGrowth от уровня поддержки

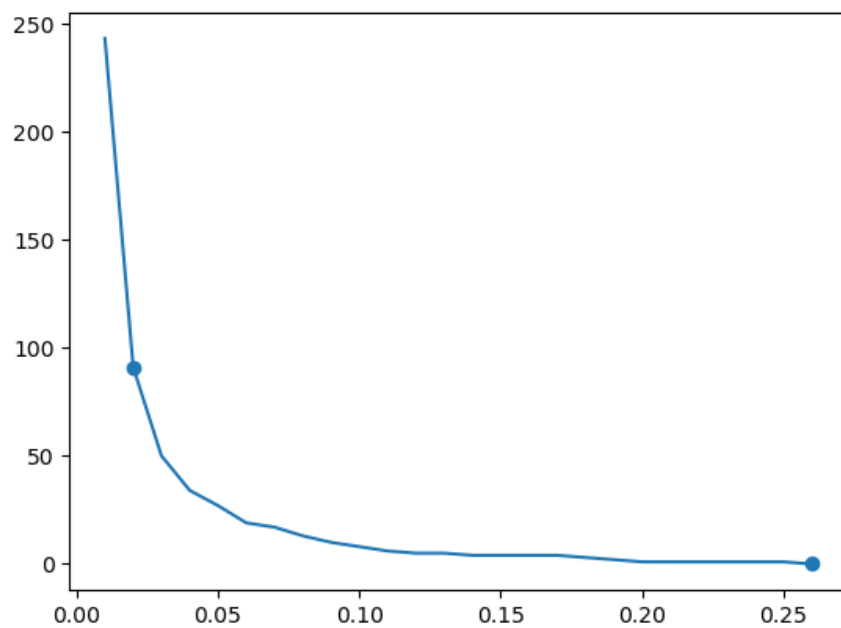


Рисунок 14 – График графики изменения количества получаемых правил в FPMax от уровня поддержки

### 3. Ассоциативные правила

Был сформированный набор данных из определенных товаров так, чтобы размер транзакции был 2 и более. Для сформированного набора данных получим частоты наборов, используя алгоритм FPGrowth. Результат представлен на рисунках 15-16.

	support	itemsets
0	0.24124043495771244	frozenset({'yogurt'})
1	0.18586387434554974	frozenset({'tropical fruit'})
2	0.42186870720902137	frozenset({'whole milk'})
3	0.33507853403141363	frozenset({'other vegetables'})
4	0.29621425694724124	frozenset({'rolls/buns'})
5	0.11337092227144584	frozenset({'bottled beer'})
6	0.18546113572291584	frozenset({'bottled water'})
7	0.1463954893274265	frozenset({'citrus fruit'})
8	0.26721707611759965	frozenset({'soda'})
9	0.19633507853403143	frozenset({'root vegetables'})
10	0.08276278695126862	frozenset({'canned beer'})

Рисунок 15 – Результат работы FPGrowth (1)

	support	itemsets
31	0.07933950865888038	frozenset({'soda', 'whole milk'})
32	0.0648409182440596	frozenset({'soda', 'other vegetable...
33	0.09383809907370116	frozenset({'root vegetables', 'other...
34	0.0968586387434555	frozenset({'root vegetables', 'whol...
35	0.05114780507450665	frozenset({'yogurt', 'root vegetabl...
36	0.060612162706403544	frozenset({'rolls/buns', 'sausage'})
37	0.05920257752718486	frozenset({'sausage', 'whole milk'})
38	0.053362867498993154	frozenset({'sausage', 'other vegeta...
39	0.06383407168747483	frozenset({'whipped/sour cream', '...
40	0.057188884414015304	frozenset({'whipped/sour cream', '...
41	0.06584776480064439	frozenset({'pastry', 'whole milk'})

Рисунок 16 – Результат работы FPGrowth (2)

Проведем ассоциативный анализ. Результаты представлены на рисунке 17, где

- ‘support’ – показывает насколько часто встречается набор, рассчитывается по формуле:

$$\text{support}(A \rightarrow C) = \text{support}(A \cup C), \text{ range: } [0,1]$$

- 'confidence' – показывает как часто консеквент встречается в транзакции при условии, что оно также содержит антецедент, рассчитывается по формуле:

$$\text{confidence}(A \rightarrow C) = \text{support}(A \rightarrow C) / \text{support}(A), \text{ range: } [0,1]$$

- 'leverage' – разница между наблюдаемой частотой появления консеквента и антецедент вместе и частотой, которую можно было бы ожидать, если бы они были независимыми.

$$\text{leverage}(A \rightarrow C) = \text{support}(A \rightarrow C) - \text{support}(A) \times \text{support}(C), \text{ range: } [-1,1]$$

- 'conviction' – показывает насколько консеквент зависит от антецедента.

$$\text{conviction}(A \rightarrow C) = (1 - \text{support}(C)) / (1 - \text{confidence}(A \rightarrow C)), \text{ range: } [0, \infty]$$

	antecedents	consequents	antecedent sup...	consequent su...	support	confidence	lift	leverage	conviction	
0	frozenset(['yog...	frozenset(['who...	0.24124043495...	0.42186870720...	0.11095449053...	0.45993322203...	1.09022834397...	0.00918270011...	1.07048112637...	
1	frozenset(['yog...	frozenset(['oth...	0.24124043495...	0.33507853403...	0.08598469593...	0.35642737896...	1.06371295909...	0.00515020463...	1.03317239459...	
2	frozenset(['trop...	frozenset(['yog...	0.18586387434...	0.24124043495...	0.05799436165...	0.31202600216...	1.29342331115...	0.01315647976...	1.10288988745...	
3	frozenset(['trop...	frozenset(['oth...	0.18586387434...	0.33507853403...	0.07108336689...	0.38244853737...	1.14136985373...	0.00880437234...	1.07670616331...	
4	frozenset(['who...	frozenset(['trop...	0.18586387434...	0.42186870720...	0.08376963350...	0.45070422535...	1.06835187737...	0.00535948112...	1.05249543046...	
5	frozenset(['who...	frozenset(['oth...	0.42186870720...	0.33507853403...	0.14820781312...	0.35131264916...	1.04844868735...	0.00684866516...	1.02502610095...	
6	frozenset(['oth...	frozenset(['who...	0.33507853403...	0.42186870720...	0.14820781312...	0.44230769230...	1.04844868735...	0.00684866516...	1.03664921465...	
7	frozenset(['rolls...	frozenset(['who...	0.29621425694...	0.42186870720...	0.11216270640...	0.37865397688...	0.89756355571...	-0.0128008192...	0.93044981585...	
8	frozenset(['bott...	frozenset(['who...	0.18546113572...	0.42186870720...	0.06806282722...	0.36699239956...	0.86992088603...	-0.0101774223...	0.91330861176...	
9	frozenset(['bott...	frozenset(['sod...	0.18546113572...	0.26721707611...	0.05739025372...	0.30944625407...	1.15803323113...	0.00783187130...	1.06115263033...	
10	frozenset(['citr...	frozenset(['who...	0.14639548932...	0.42186870720...	0.06041079339...	0.41265474552...	0.97815917245...	-0.0013488824...	0.98431252894...	

Рисунок 17 – результат ассоциативного анализа

Если не задавать метрику вручную, то будет выбрана метрика 'confidence'. Были рассчитаны средние значения для каждой метрики:

Antecedent support	0.250056
consequent support	0.250056
support	0.068443
confidence	0.309774
lift	1.240814

leverage	0.013108
conviction	1.096450

Проведем анализ со следующим правилом:  $\text{min\_confidence} = 0.4$ . Построим, ориентированный от антецедента к консеквенту, граф для полученных правил, где каждая вершина отображает набор товаров, ширина ребра отображает уровень support, а подпись на ребре отображает confidence. Результат представлен на рисунке 18.

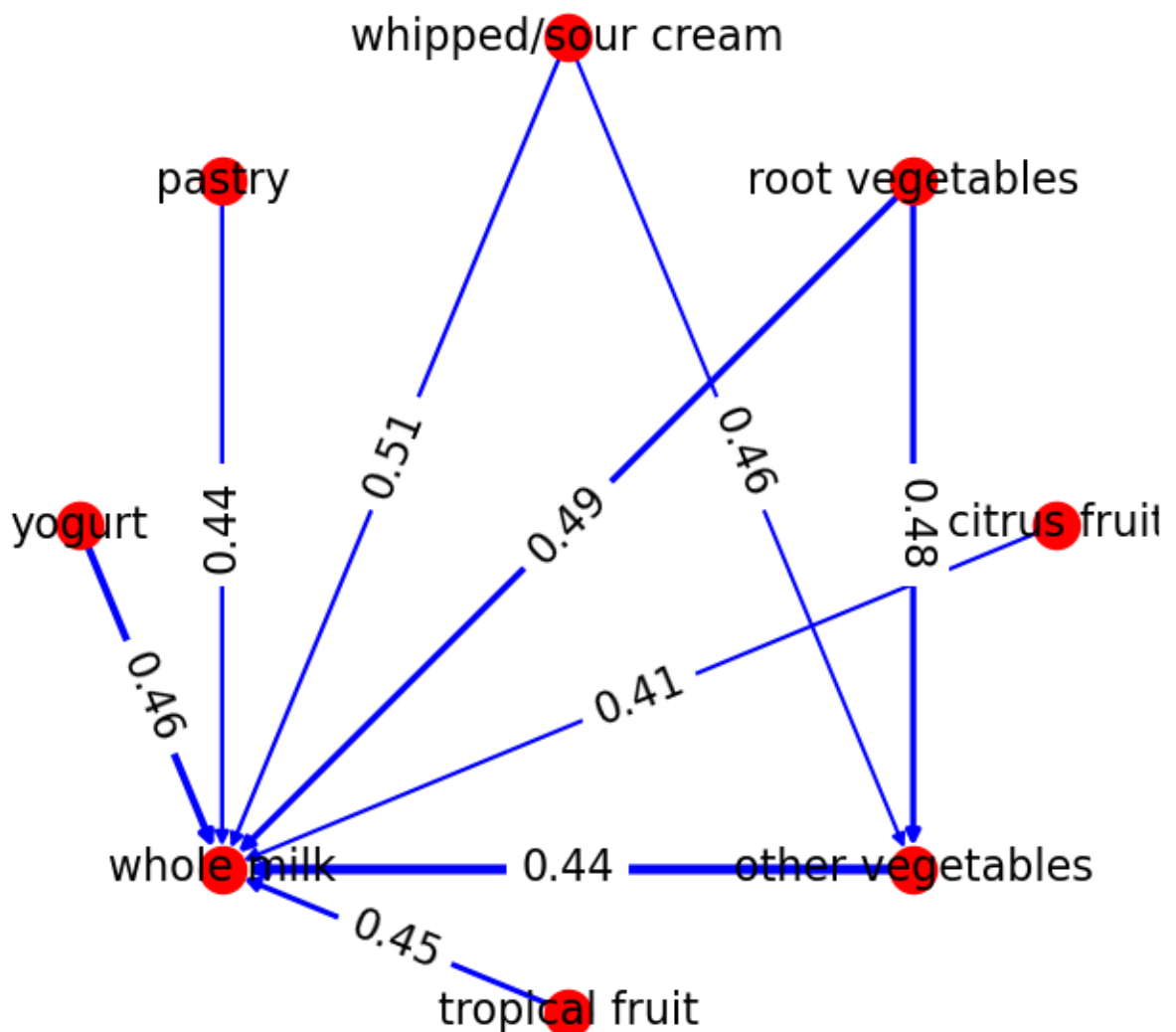


Рисунок 18 – Граф, отображающий правила ассоциативного анализа

Также был предложен свой способ визуализации полученных правил. На графике каждая точка отображает набор товаров, размер точки отображает support, а цвет отображает confidence.

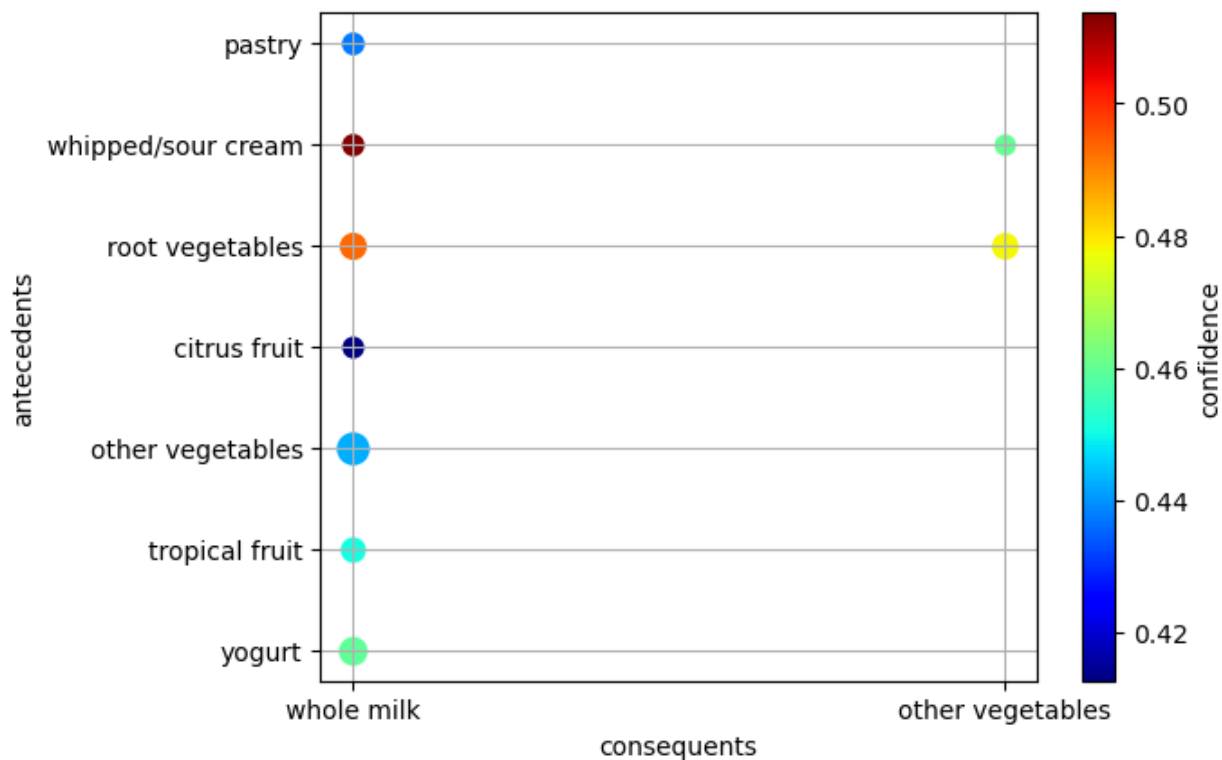


Рисунок 19 – Визуализация правил ассоциативного анализа

## Вывод

Были получены навыки работы с алгоритмами FPGrowth, FPMax, а также с ассоциативным анализом.