



Московский институт электроники и
математики им. А.Н. Тихонова

Кафедра информационной
безопасности киберфизических
систем

Москва 2025

Анализ данных с использованием ассоциативных правил

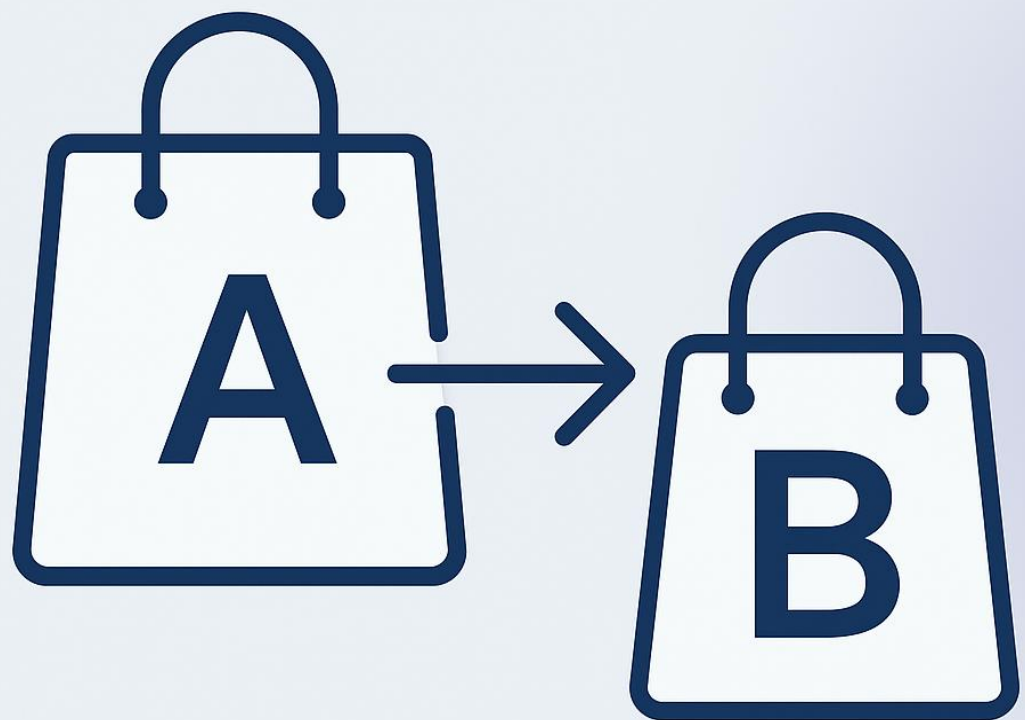
Анализ данных с использованием ассоциативных правил

Ассоциативные правила помогают выявить скрытые закономерности в транзакционных данных. Используются для формирования рекомендаций, оптимизации ассортимента и маркетинговых решений.





Сущность ассоциативных правил



Ассоциативные правила имеют вид: «Если товар А был куплен, то с заданной вероятностью будет куплен товар В».

Широко используются в ритейле и электронной коммерции.

Области применения



- Рекомендации товаров (например, «с этим товаром покупают»).
- Оптимизация полочного пространства магазинов.
- Создание промо-комплектов товаров.





Преимущества ассоциативного анализа

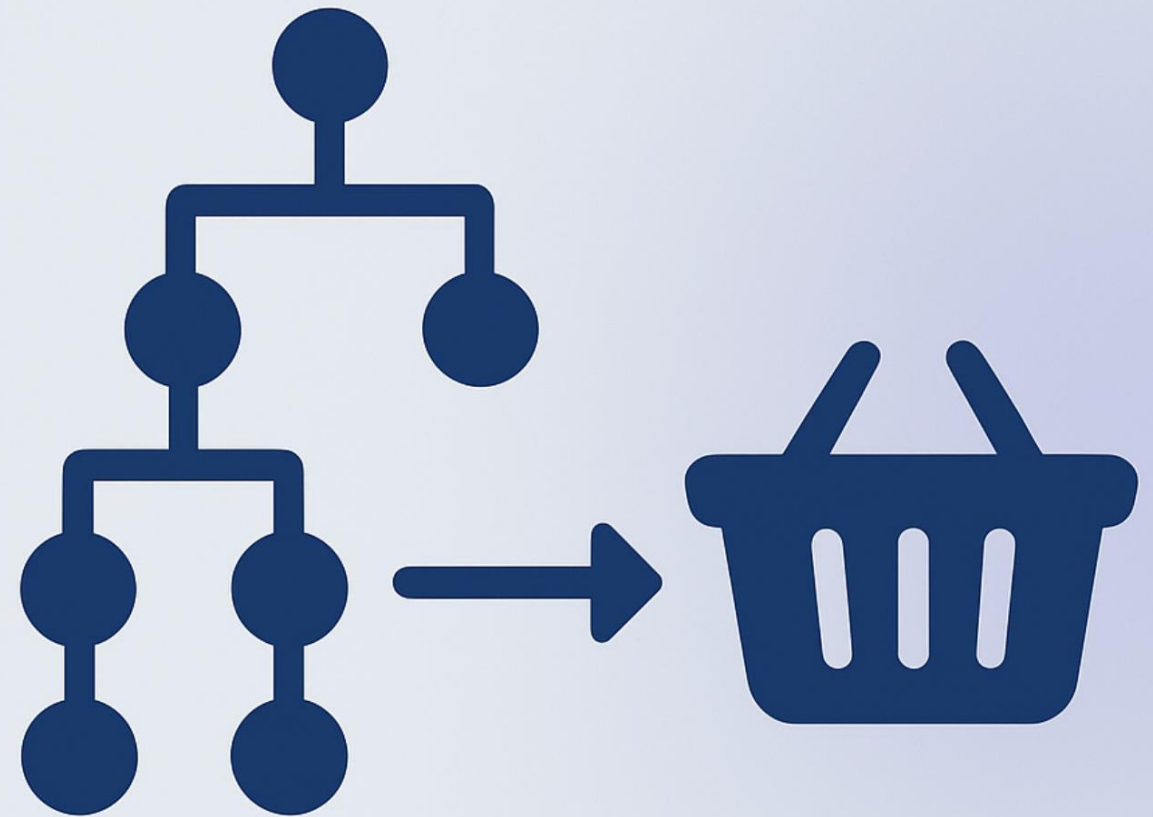
- Повышение продаж благодаря релевантным рекомендациям.
- Улучшение восприятия покупателями выкладки товара.
- Оптимизация закупок за счёт выявления популярных связок товаров.

Алгоритм FP-Growth



Алгоритм FP-Growth —
быстрый метод поиска часто
встречающихся товарных
наборов.

Использует древовидные
структуры (FP-tree),
значительно ускоряя анализ
по сравнению с другими
методами.





Описание используемых данных

Набор данных онлайн-продаж
включает номера транзакций
(InvoiceNo) и описания
товаров (Description).

Эти столбцы позволяют
выявить товары, которые
часто покупаются вместе.

Result History

ExampleSet (//Local Repository/pract4/Online Retail) X

Open in Turbo Prep Auto Model Interactive Analysis

| Row No. | InvoiceNo | StockCode | Description | Quantity | InvoiceDate | UnitPrice | CustomerID | Country |
|---------|-----------|-----------|---------------|----------|-----------------|-----------|------------|-----------------|
| 1 | 536365 | 85123A | WHITE HAN... | 6 | Dec 1, 2010 ... | 2.550 | 17850 | United Kingd... |
| 2 | 536365 | 71053 | WHITE MET... | 6 | Dec 1, 2010 ... | 3.390 | 17850 | United Kingd... |
| 3 | 536365 | 84406B | CREAM CU... | 8 | Dec 1, 2010 ... | 2.750 | 17850 | United Kingd... |
| 4 | 536365 | 84029G | KNITTED UN... | 6 | Dec 1, 2010 ... | 3.390 | 17850 | United Kingd... |
| 5 | 536365 | 84029E | RED WOOL... | 6 | Dec 1, 2010 ... | 3.390 | 17850 | United Kingd... |
| 6 | 536365 | 22752 | SET 7 BABU... | 2 | Dec 1, 2010 ... | 7.650 | 17850 | United Kingd... |
| 7 | 536365 | 21730 | GLASS STA... | 6 | Dec 1, 2010 ... | 4.250 | 17850 | United Kingd... |
| 8 | 536366 | 22633 | HAND WAR... | 6 | Dec 1, 2010 ... | 1.850 | 17850 | United Kingd... |
| 9 | 536366 | 22632 | HAND WAR... | 6 | Dec 1, 2010 ... | 1.850 | 17850 | United Kingd... |
| 10 | 536367 | 84879 | ASSORTED ... | 32 | Dec 1, 2010 ... | 1.690 | 13047 | United Kingd... |
| 11 | 536367 | 22745 | POPPY'S PL... | 6 | Dec 1, 2010 ... | 2.100 | 13047 | United Kingd... |
| 12 | 536367 | 22748 | POPPY'S PL... | 6 | Dec 1, 2010 ... | 2.100 | 13047 | United Kingd... |
| 13 | 536367 | 22749 | FELTCRAFT... | 8 | Dec 1, 2010 ... | 3.750 | 13047 | United Kingd... |
| 14 | 536367 | 22310 | IVORY KNIT... | 6 | Dec 1, 2010 ... | 1.650 | 13047 | United Kingd... |
| 15 | 536367 | 84969 | BOX OF 6 A... | 6 | Dec 1, 2010 ... | 4.250 | 13047 | United Kingd... |
| 16 | 536367 | 22623 | BOX OF VIN... | 3 | Dec 1, 2010 ... | 4.950 | 13047 | United Kingd... |
| 17 | 536367 | 22622 | BOX OF VIN... | 2 | Dec 1, 2010 ... | 9.950 | 13047 | United Kingd... |
| 18 | 536367 | 21754 | HOME BUIL... | 3 | Dec 1, 2010 ... | 5.950 | 13047 | United Kingd... |
| 19 | 536367 | 21755 | LOVE BUILD... | 3 | Dec 1, 2010 ... | 5.950 | 13047 | United Kingd... |
| 20 | 536367 | 21777 | RECIPE BO... | 4 | Dec 1, 2010 ... | 7.950 | 13047 | United Kingd... |
| 21 | 536367 | 48187 | DOORMAT ... | 4 | Dec 1, 2010 ... | 7.950 | 13047 | United Kingd... |

ExampleSet (541,909 examples,0 special attributes,8 regular attributes)

Цель анализа набора данных

Цель анализа — выявить товары, которые приобретаются совместно чаще всего, и определить сильные ассоциативные связи, полезные для маркетинговых и коммерческих решений.





Предварительная очистка данных

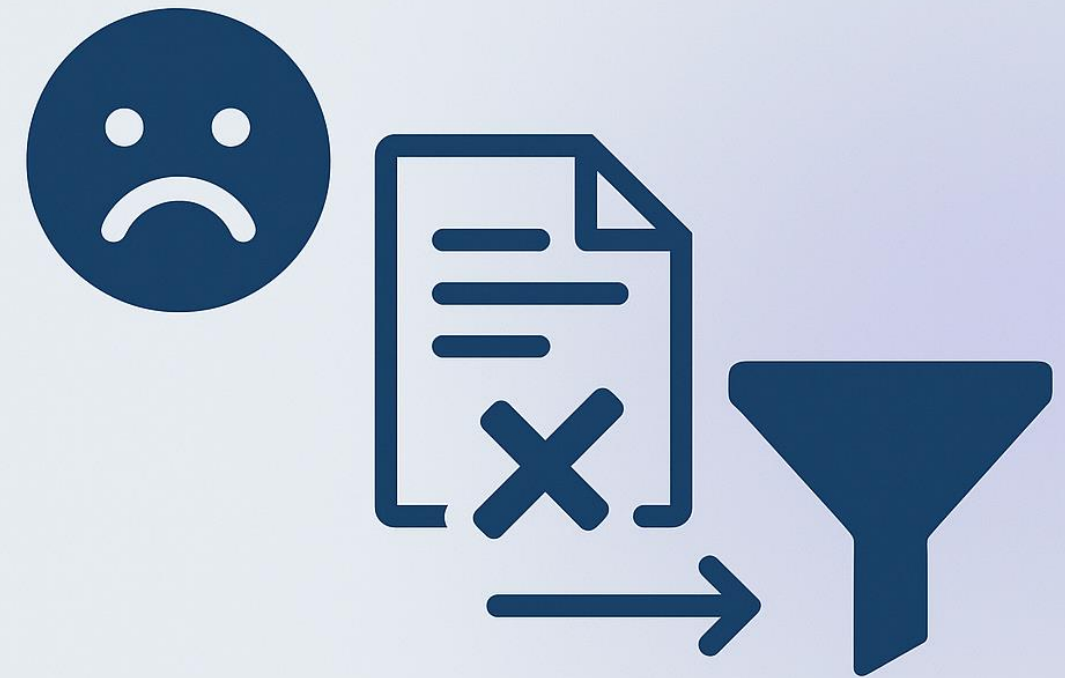
Перед началом анализа необходимо удалить некорректные записи, такие как возвраты ($\text{Quantity} < 0$), чтобы получить достоверные результаты анализа покупок.

The screenshot shows a dialog box titled "Create Filters: filters" with a close button (X) in the top right corner. Inside the dialog, there is a funnel icon and the text "Create Filters: filters" and "Defines the list of filters to apply." Below this, there is a filter entry field containing "Quantity" in a dropdown, a greater-than sign ">" in a dropdown, and the value "0" in a text input. To the right of the input field are two small icons: a star and a red X. At the bottom of the dialog, there are three radio buttons: "Match all" (selected), "Match any", and "Preselect comparators" (checked). To the right of these are three buttons: "Add Entry" (with a plus icon), "OK" (with a green checkmark), and "Cancel" (with a red X).

Проблемы данных и их решение

Типичные проблемы: возвраты, дубликаты, пустые строки.

Решаются с помощью фильтрации данных, обеспечивая корректность анализа и высокое качество ассоциативных правил.

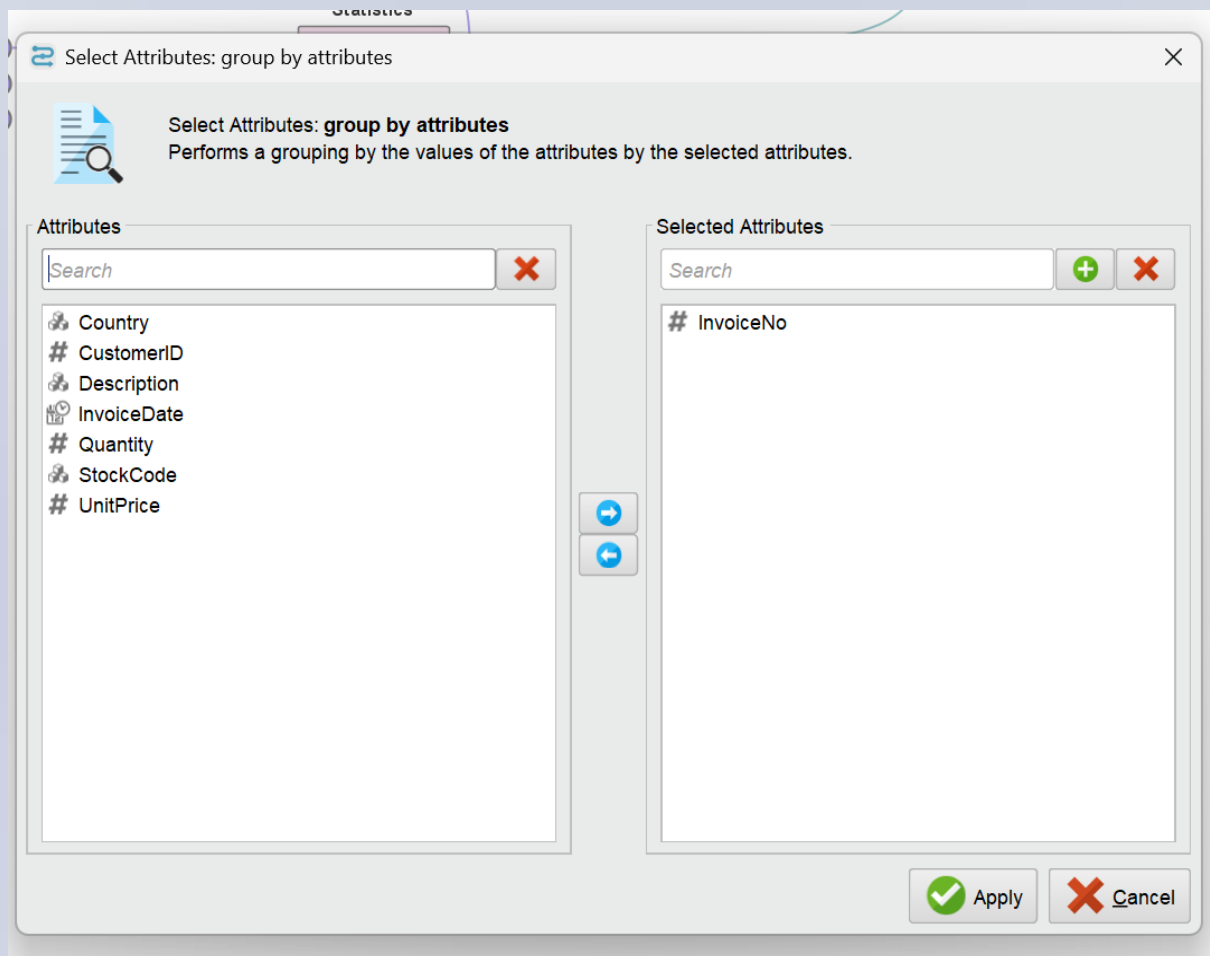




Агрегация транзакций

Данные объединяются по номеру транзакции (InvoiceNo).

Это позволяет получить списки товаров, купленных за один заказ, необходимые для выявления ассоциаций.



Конкатенация товаров в одной транзакции

Использование функции concatenation позволяет представить все товары одной транзакции в виде единой строки, удобной для дальнейшего анализа.

Edit Parameter List: aggregation attributes

Edit Parameter List: **aggregation attributes**
The attributes which should be aggregated.

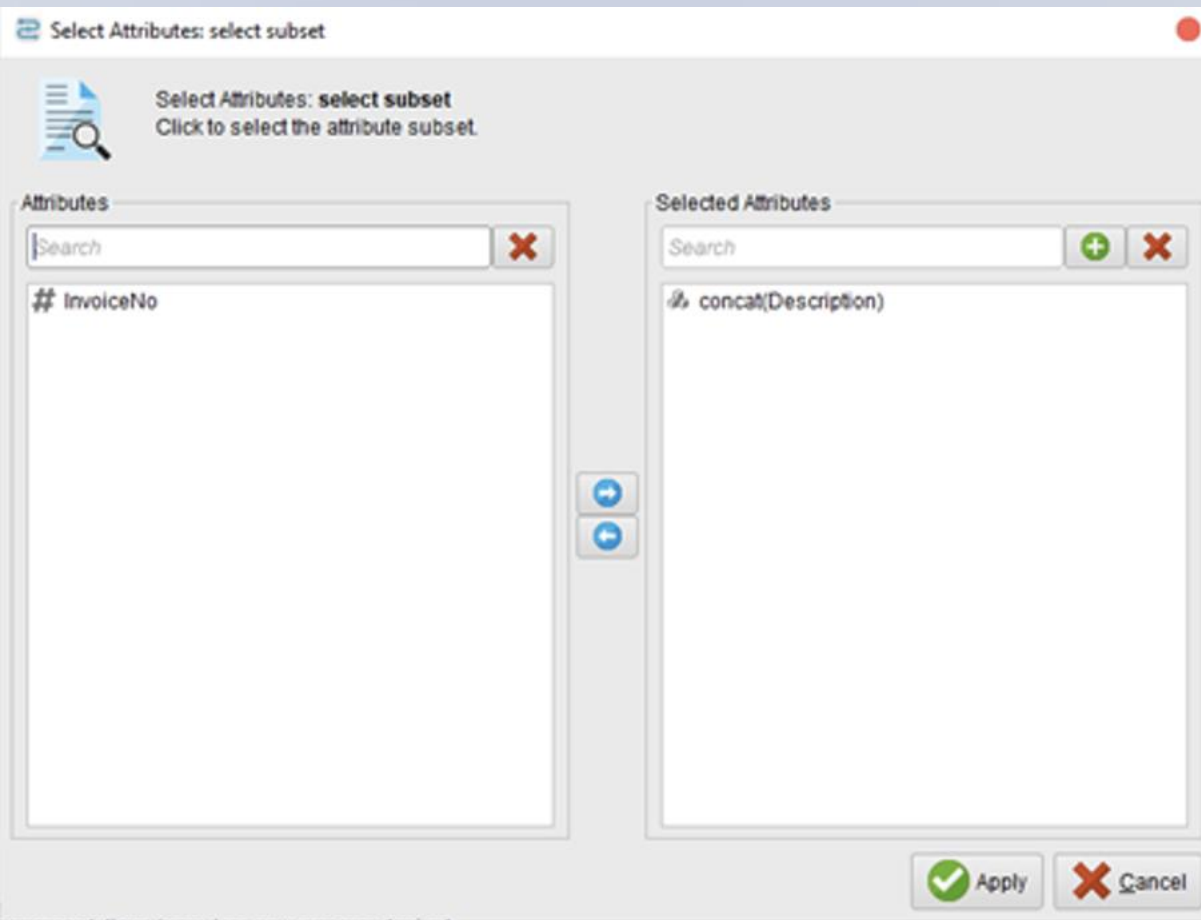
| aggregation attribute | aggregation functions |
|-----------------------|-----------------------|
| Description | concatenation |

Add Entry Remove Entry Apply Cancel



Выбор нужных атрибутов

На этапе анализа оставляют только два столбца: идентификатор транзакции (InvoiceNo) и объединённые названия товаров (concat(Description)), исключая лишнюю информацию.

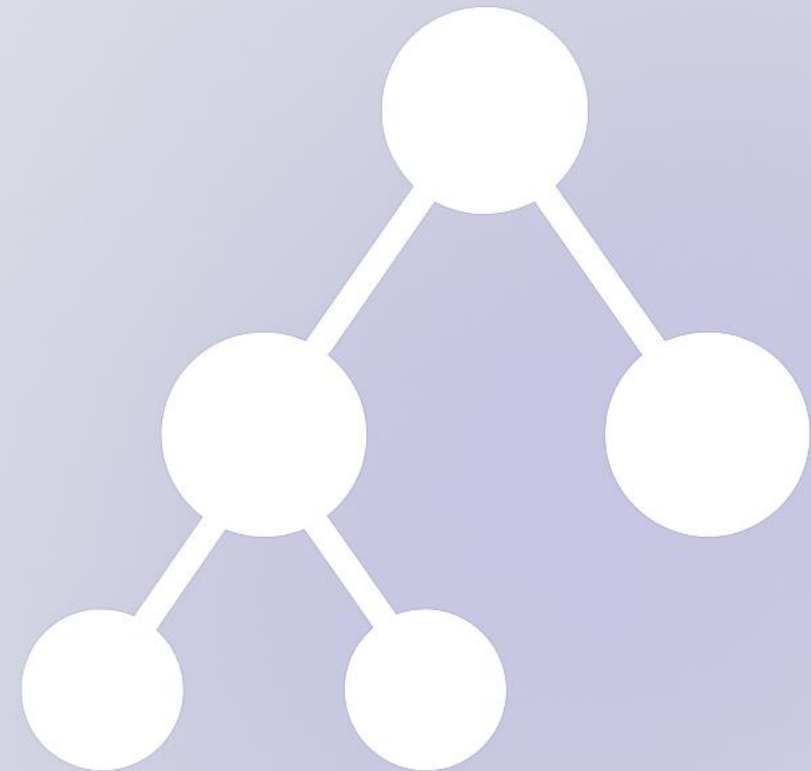


Поиск частых наборов: FP-Growth



Параметр min support
(минимальная поддержка)
задаётся на уровне 2%.

Это исключает редкие
наборы товаров, не
имеющие практической
значимости.





Настройка алгоритма FP-Growth

Parameters

FP-Growth

input format: item list in a column

item separators: |

☐ use quotes

☒ trim item names

min requirement: support

min support: 0.02

min items per itemset: 1

max items per itemset: 0

max number of itemsets: 1000000

☒ find min number of itemsets

min number of itemsets: 100

must contain list: [Edit Enumeration \(0\)...](#)

[Show advanced parameters](#)

[Change compatibility \(10.5.000\)](#)

- Формат данных: списки товаров в столбце.
- Разделитель товаров: символ "|".
- Удаление лишних пробелов в названиях товаров (trim item names).



Параметры ассоциативных правил

Уверенность (confidence) —
основной критерий создания
правил.

Значение выше 50% показывает
надёжные и значимые
ассоциации между товарами.

Parameters ×

Create Association Rules

| | |
|----------------|--|
| criterion | confidence ▼ ? |
| min confidence | 0.5 ? |



Создание ассоциативных правил



Правила формируются на основе часто встречающихся наборов товаров.

Например, «при покупке товара X покупают товар Y с уверенностью 56%».

Интерпретация ассоциативных правил

Правило ROSES REGENCY TEACUP → PINK REGENCY TEACUP (уверенность 56%) указывает на желание покупателей дополнять наборы товарами одной серии или похожего стиля.



Data



Graph



Description



Annotations

AssociationRules

Association Rules

```
[GREEN REGENCY TEACUP AND SAUCER] --> [REGENCY CAKESTAND 3 TIER] (confidence: 0.501)
[LUNCH BAG BLACK SKULL.] --> [LUNCH BAG RED RETROSPOT] (confidence: 0.504)
[RED RETROSPOT CHARLOTTE BAG] --> [CHARLOTTE BAG PINK POLKADOT] (confidence: 0.505)
[ROSES REGENCY TEACUP AND SAUCER] --> [GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER] (confidence: 0.508)
[LUNCH BAG SUKI DESIGN] --> [LUNCH BAG RED RETROSPOT] (confidence: 0.512)
[LUNCH BAG WOODLAND] --> [LUNCH BAG RED RETROSPOT] (confidence: 0.517)
[CHARLOTTE BAG SUKI DESIGN] --> [WOODLAND CHARLOTTE BAG] (confidence: 0.517)
[JUMBO BAG WOODLAND ANIMALS] --> [JUMBO STORAGE BAG SUKI] (confidence: 0.520)
[HEART OF WICKER LARGE] --> [HEART OF WICKER SMALL] (confidence: 0.532)
[ALARM CLOCK BAKELIKE PINK] --> [ALARM CLOCK BAKELIKE GREEN] (confidence: 0.534)
[GREEN REGENCY TEACUP AND SAUCER] --> [ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER] (confidence: 0.534)
[WOODLAND CHARLOTTE BAG] --> [CHARLOTTE BAG SUKI DESIGN] (confidence: 0.546)
[60 TEATIME FAIRY CAKE CASES] --> [PACK OF 72 RETROSPOT CAKE CASES] (confidence: 0.547)
[RECYCLING BAG RETROSPOT] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.555)
[WOODEN FRAME ANTIQUE WHITE] --> [WOODEN PICTURE FRAME WHITE FINISH] (confidence: 0.555)
[LUNCH BAG PINK POLKADOT] --> [LUNCH BAG RED RETROSPOT] (confidence: 0.556)
[JUMBO BAG APPLES] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.558)
[ROSES REGENCY TEACUP AND SAUCER] --> [PINK REGENCY TEACUP AND SAUCER] (confidence: 0.562)
[STRAWBERRY CHARLOTTE BAG] --> [CHARLOTTE BAG SUKI DESIGN] (confidence: 0.575)
[CHARLOTTE BAG SUKI DESIGN] --> [RED RETROSPOT CHARLOTTE BAG] (confidence: 0.575)
[JUMBO SHOPPER VINTAGE RED PAISLEY] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.579)
[SPACEBOY LUNCH BOX] --> [DOLLY GIRL LUNCH BOX] (confidence: 0.598)
[GARDENERS KNEELING PAD KEEP CALM] --> [GARDENERS KNEELING PAD CUP OF TEA] (confidence: 0.598)
[JUMBO BAG PINK VINTAGE PAISLEY] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.599)
[ALARM CLOCK BAKELIKE PINK] --> [ALARM CLOCK BAKELIKE RED] (confidence: 0.600)
[WOODLAND CHARLOTTE BAG] --> [RED RETROSPOT CHARLOTTE BAG] (confidence: 0.606)
[ALARM CLOCK BAKELIKE RED] --> [ALARM CLOCK BAKELIKE GREEN] (confidence: 0.609)
[JUMBO STORAGE BAG SUKI] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.611)
[JUMBO BAG WOODLAND ANIMALS] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.620)
[GREEN REGENCY TEACUP AND SAUCER] --> [PINK REGENCY TEACUP AND SAUCER] (confidence: 0.624)
[JUMBO BAG SPACEBOY DESIGN] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.625)
[JUMBO BAG BAROQUE BLACK WHITE] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.627)
[DOLLY GIRL LUNCH BOX] --> [SPACEBOY LUNCH BOX] (confidence: 0.633)
[JUMBO BAG STRAWBERRY] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.651)
[ALARM CLOCK BAKELIKE GREEN] --> [ALARM CLOCK BAKELIKE RED] (confidence: 0.653)
[RED HANGING HEART T-LIGHT HOLDER] --> [WHITE HANGING HEART T-LIGHT HOLDER] (confidence: 0.668)
[STRAWBERRY CHARLOTTE BAG] --> [RED RETROSPOT CHARLOTTE BAG] (confidence: 0.673)
[PAPER CHAIN KIT VINTAGE CHRISTMAS] --> [PAPER CHAIN KIT 50's CHRISTMAS] (confidence: 0.673)
[JUMBO BAG SCANDINAVIAN BLUE PAISLEY] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.677)
[JUMBO BAG PINK POLKADOT] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.677)
[DOTCOM POSTAGE] --> [JUMBO BAG RED RETROSPOT] (confidence: 0.685)
[CHARLOTTE BAG PINK POLKADOT] --> [RED RETROSPOT CHARLOTTE BAG] (confidence: 0.703)
[ROSES REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER] --> [PINK REGENCY TEACUP AND SAUCER] (confidence: 0.706)
[PINK REGENCY TEACUP AND SAUCER] --> [ROSES REGENCY TEACUP AND SAUCER, GREEN REGENCY TEACUP AND SAUCER] (confidence: 0.708)
[GARDENERS KNEELING PAD CUP OF TEA] --> [GARDENERS KNEELING PAD KEEP CALM] (confidence: 0.720)
[ROSES REGENCY TEACUP AND SAUCER] --> [GREEN REGENCY TEACUP AND SAUCER] (confidence: 0.720)
[GREEN REGENCY TEACUP AND SAUCER] --> [ROSES REGENCY TEACUP AND SAUCER] (confidence: 0.757)
[PINK REGENCY TEACUP AND SAUCER] --> [ROSES REGENCY TEACUP AND SAUCER] (confidence: 0.782)
[PINK REGENCY TEACUP AND SAUCER] --> [GREEN REGENCY TEACUP AND SAUCER] (confidence: 0.826)
[GREEN REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER] --> [ROSES REGENCY TEACUP AND SAUCER] (confidence: 0.856)
[ROSES REGENCY TEACUP AND SAUCER, PINK REGENCY TEACUP AND SAUCER] --> [GREEN REGENCY TEACUP AND SAUCER] (confidence: 0.905)
```



Значение ассоциативных правил для бизнеса

Правила помогают эффективно планировать ассортимент, улучшать выкладку товаров, создавать целевые рекламные кампании и акционные предложения.



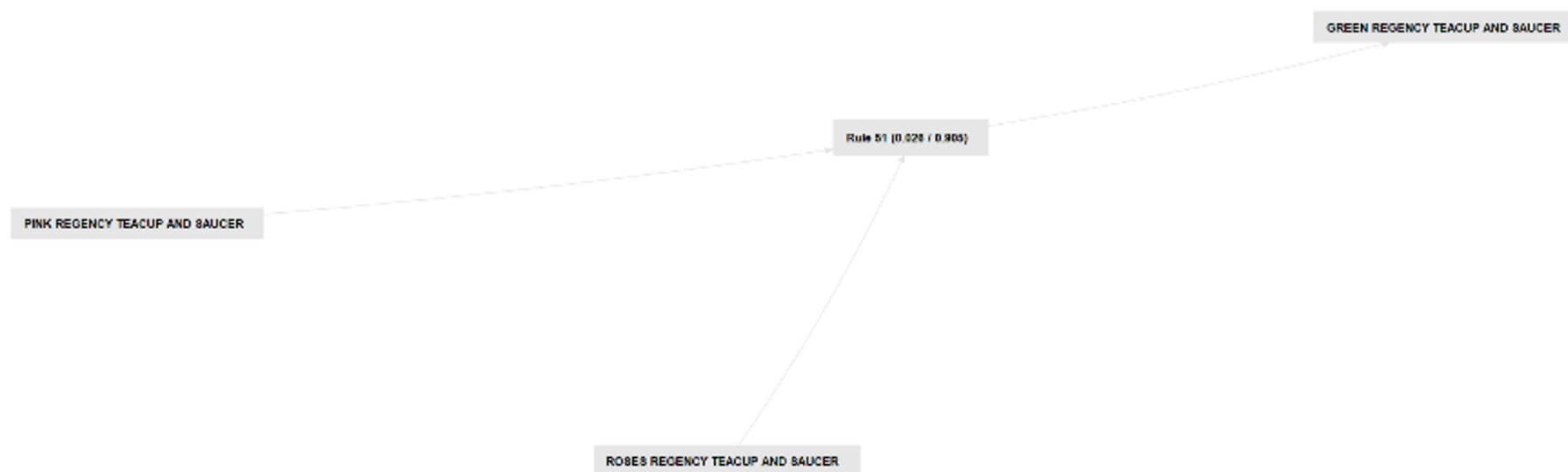
Визуализация ассоциаций в RapidMiner

Графическое представление правил помогает быстро выявить товары с большим количеством ассоциаций и выделить значимые сегменты покупателей с особыми предпочтениями.





Выводы из графического анализа



Центральные
товары имеют
много связей и
являются важными
для маркетинга и
рекомендаций.

Выделение сегментов покупателей

Изолированные группы товаров показывают специфические предпочтения отдельных сегментов покупателей, что полезно для персонализации предложений и акций.





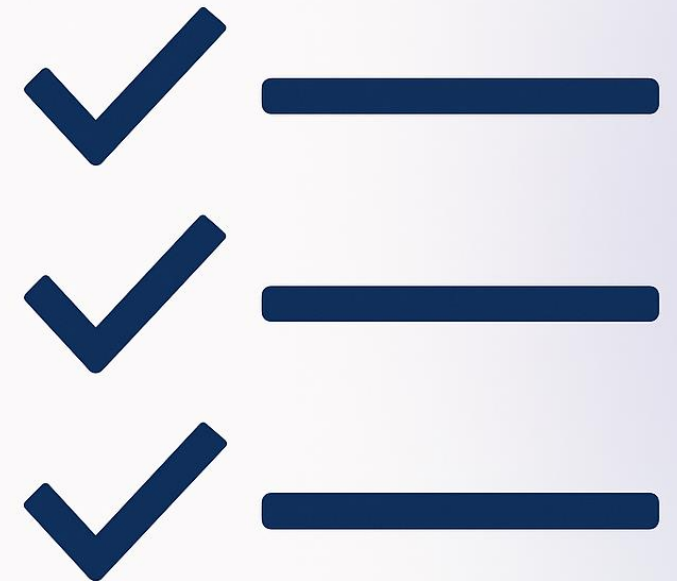
Типичные ошибки анализа

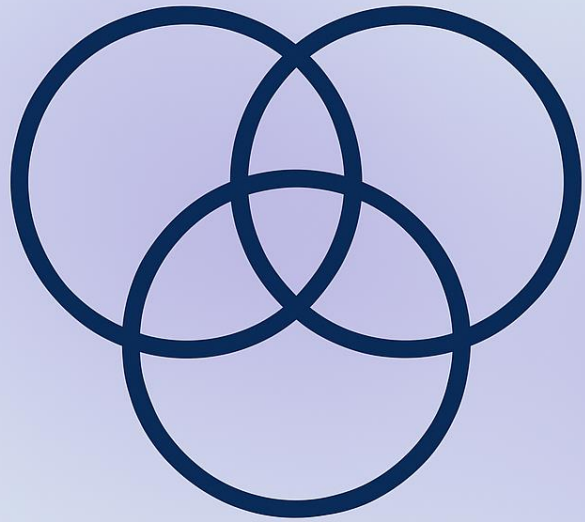


- Слишком низкое значение min support вызывает шум.
- Игнорирование возвратов искажает результаты.
- Высокий confidence не всегда означает практическую ценность правила.

Рекомендации по улучшению анализа

- Регулярная оптимизация параметров (min support, confidence).
- Тщательная предобработка данных.
- Применение бизнес-логики для оценки практической применимости правил.





Заключение

Ассоциативные правила в RapidMiner — эффективный инструмент анализа покупок.

Правила и их визуализация помогают улучшать ассортимент, формировать маркетинговые стратегии и оптимизировать бизнес-процессы.