

Лабораторная работа №7

Ассоциативные правила

Цель работы: изучить процесс поиска ассоциаций.

Задачи работы:

- изучить методику поиска ассоциативных правил в Deductor Studio;
- изучить приведенные в лабораторной работе примеры;
- выполнить контрольное задание.

1. Краткая теория

В Deductor Studio для решения задач ассоциации используется обработчик **Ассоциативные правила**. В нем реализован алгоритм a priori. Обработчик требует на входе два поля: идентификатор транзакции и элемент транзакции.

Например, идентификатор транзакции – это номер чека или код клиента. А элемент – это наименование товара в чеке или услуга, заказанная клиентом.

Оба поля (идентификатор и элемент транзакции) должны быть дискретного вида.

После работы обработчика по умолчанию предлагается визуализатор **Правила**. Вся остальная дополнительная информация, располагается в специализированных визуализаторах **Популярные наборы**, **Дерево правил**, **Что-если**.

Пример: анализ покупательских корзин для стимулирования спроса

Розничная сеть по продаже бытовой химии поставила задачу анализа покупательских корзин для оптимизации их размещения на витринах и проведения кросс-продаж. Отдел маркетинга предоставил 5 000 чеков, в которых отражены покупки сделанные предыдущими клиентами магазинов. Стоит следующая задача:

- предсказать то, какие товары покупатели могут выбрать в зависимости от того, что уже есть в их корзинах;
- выявить наиболее популярные товарные наборы, состоящие из более чем 1 предмета;
- предложить рекламные акции типа «каждому купившему А и В товар С в подарок».

Решение задачи

Шаг 1. В новом проекте в Deductor Studio импортируем данные из текстового файла Чеки.txt. В наборе данных два столбца:

| N | Поле | Тип поля | Назначение |
|---|------|-----------|------------|
| 1 | ID | строковый | Код чека |
| 2 | ITEM | строковый | Товар |

Шаг 2. К узлу импорта добавим обработчик Ассоциативные правила. Столбец ID сделаем идентификатором транзакции, а ITEM – ее элементом (рисунок 1).

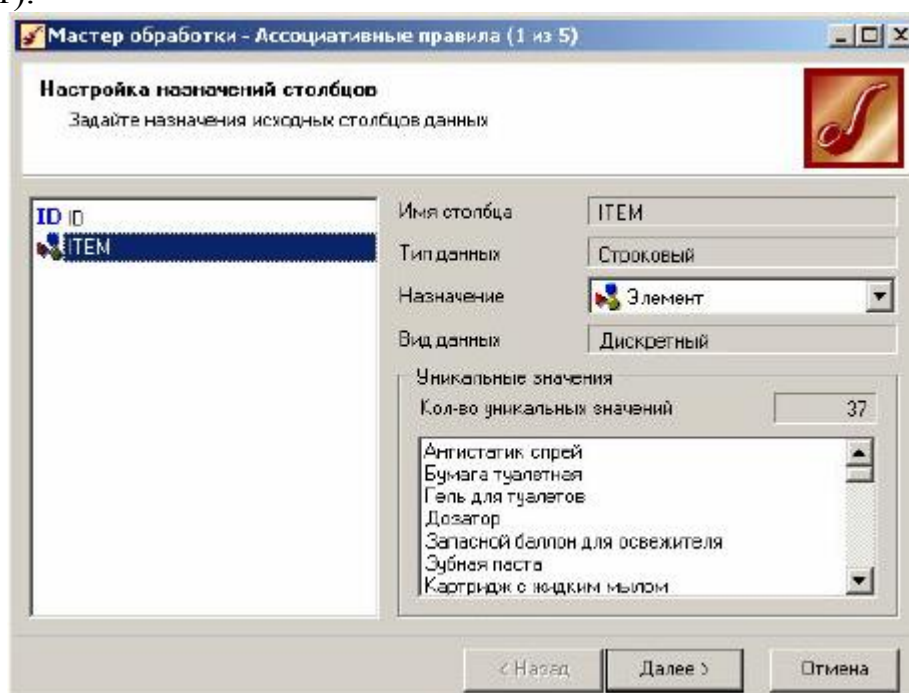


Рисунок 1 - Настройка входов для обработчика

На следующем шаге мастера настроим параметры построения ассоциативных правил (рисунок 2).

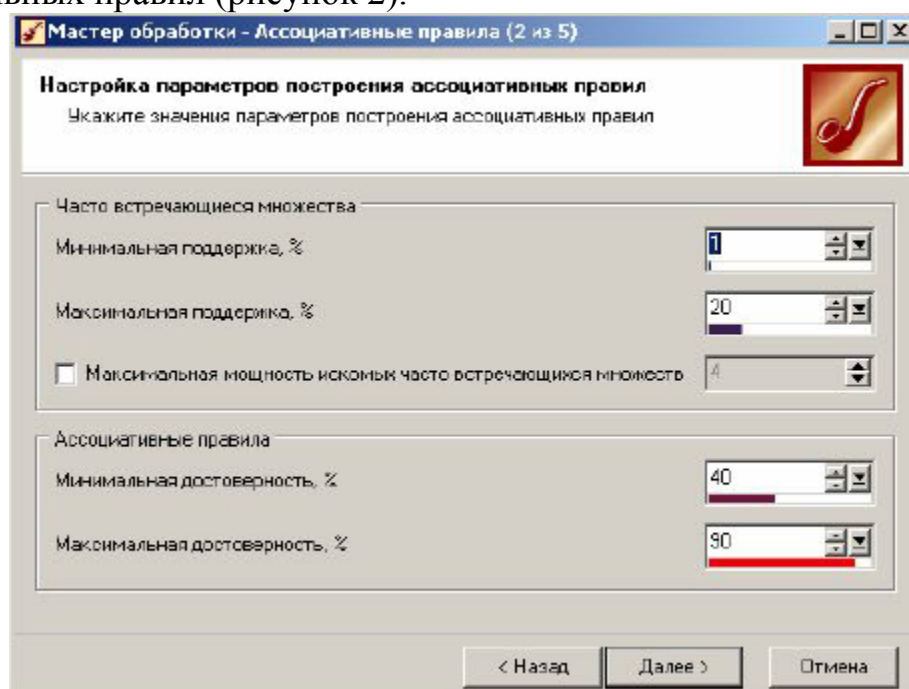


Рисунок 2 - Настройка параметров алгоритма a priori

Здесь для изменения доступны следующие параметры.

Минимальная и максимальная поддержка в % – ограничивают пространство поиска часто встречающихся предметных наборов. Эти границы определяют множество популярных наборов, из которых и будут создаваться ассоциативные правила.

Минимальная и максимальная достоверность в % – в результирующий набор попадут только те ассоциативные правила, которые удовлетворяют условиям минимальной и максимальной достоверности.

Максимальная мощность искомых часто встречающихся множеств – параметр ограничивает длину k-предметного набора. Например, при установке значения 4 шаг генерации популярных наборов будет остановлен после получения множества 4-предметных наборов. В конечном итоге это позволяет избежать появления длинных ассоциативных правил, которые трудно интерпретируются.

Оставим все настройки по умолчанию. Нажатие на кнопку Пуск приведет к работе алгоритма поиска ассоциативных правил. По окончании его работы справа в полях появится следующая информация (рисунок 3):

1. **Количество множеств** – число популярных наборов, удовлетворяющих заданным условиям минимальной поддержки и достоверности.
2. **Количество правил** – число сгенерированных ассоциативных правил.

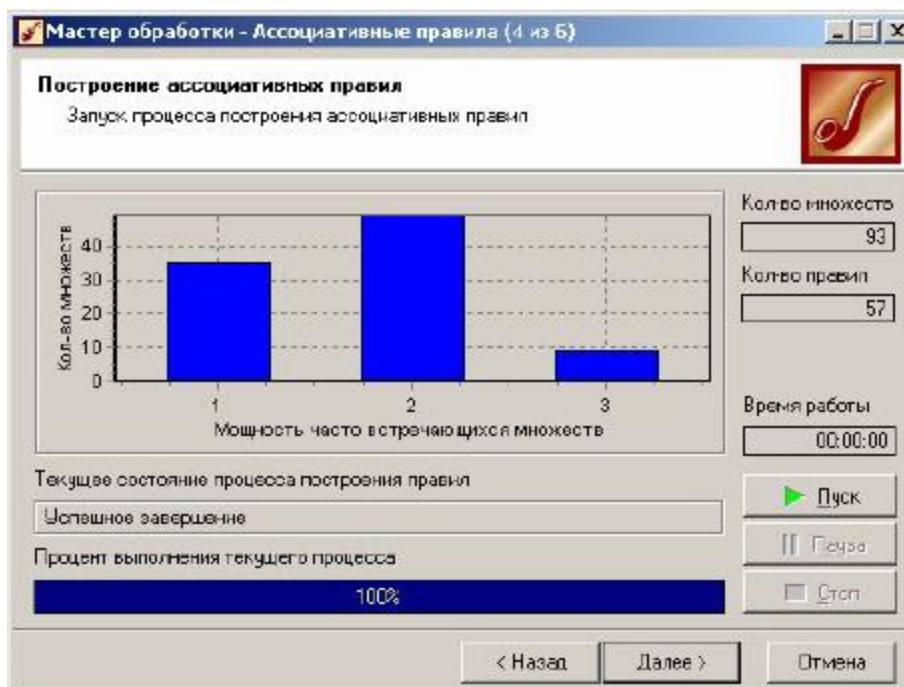


Рисунок 3 - Процесс поиска ассоциативных правил

Далее выбираем все доступные специализированные визуализаторы и визуализатор Таблица (рисунок 4).

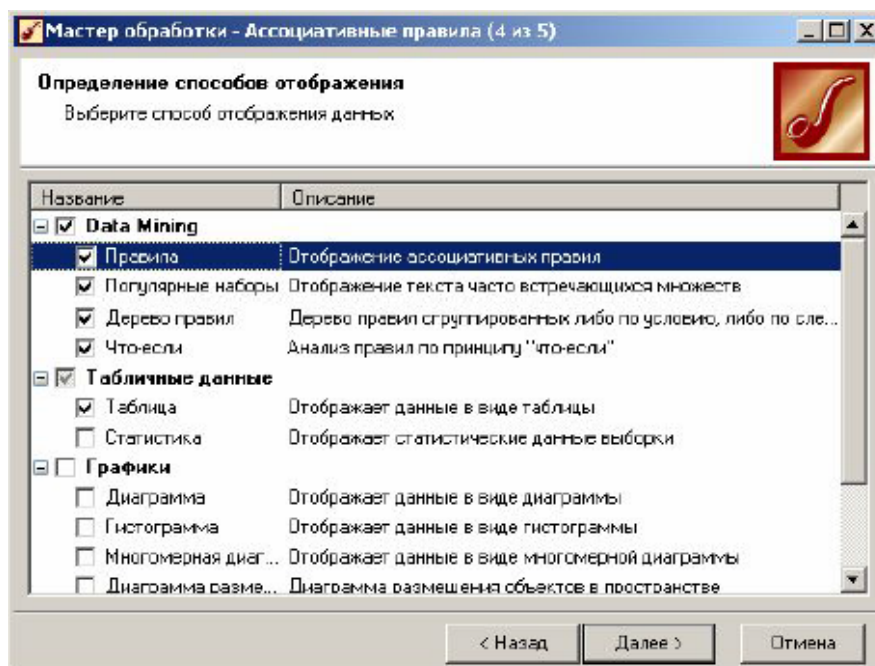




Рисунок 4 - Доступные визуализаторы после узла «Ассоциативные правила»

Все эти визуализаторы, кроме **Что-если**, отображают результаты работы алгоритма в различных формах.

На вкладке **Популярные наборы** отображается множество найденных популярных предметных наборов в виде списка. Кнопка  предлагает на выбор несколько вариантов сортировки списка, а кнопка  вызывает окно настройки фильтра множеств.

Например, задав в фильтре минимальное значение поддержки 6% и отсортировав их по убыванию поддержки, получим следующие 16 популярных наборов (рисунок 5).

На вкладке **Дерево правил** предлагается еще один удобный способ отображения множества ассоциативных правил, которое строится либо по условию, либо по следствию. При построении дерева правил по условию, на первом (верхнем) уровне находятся узлы с условиями, а на втором уровне – узлы со следствием. В дереве, построенном по следствию, наоборот, на первом уровне располагаются узлы со следствием.

| Популярные наборы | | | | | |
|--------------------|-----------------|--------------------------------------|-----------|-------|----------|
| Множеств: 16 из 91 | | Фильтр: Минимальная поддержка = 6,01 | | | |
| № | Номер множества | аб. Элементы | Поддержка | | Мощность |
| | | | Кол-во | % | |
| 1 | 10 | Мыло кусковое | 392 | 19,17 | 1 |
| 2 | 35 | Чистящий порошок универсальный | 292 | 14,26 | 1 |
| 3 | 6 | Зубная паста | 266 | 14,06 | 1 |
| 4 | 3 | Гель для туалетов | 221 | 10,81 | 1 |
| 5 | 11 | Освежитель воздуха | 208 | 10,17 | 1 |
| 6 | 9 | Мыло жидкое | 200 | 9,78 | 1 |
| 7 | 12 | Отбеливатель | 199 | 9,73 | 1 |
| 8 | 32 | Стиральный порошок ручной | 195 | 9,54 | 1 |
| 9 | 59 | Мыло жидкое | 167 | 8,17 | 2 |
| | | Мыло кусковое | | | |
| 10 | 26 | Средство для ухода за мебелью | 154 | 7,53 | 1 |
| 11 | 24 | Средство для мытья посуды | 152 | 7,43 | 1 |
| 12 | 19 | Пятновыводитель | 146 | 7,14 | 1 |
| 13 | 34 | Стиральный порошок-автомат | 144 | 7,04 | 1 |
| 14 | 21 | Сода кальцинированная | 133 | 6,50 | 1 |
| 15 | 67 | Мыло кусковое | 131 | 6,41 | 2 |
| | | Средство для мытья посуды | | | |
| 16 | 29 | Средство для чистки плит | 129 | 6,31 | 1 |

Рисунок 5 - Отфильтрованные популярные наборы

Справа от дерева расположен список правил, построенный по выбранному узлу дерева (рисунок 6).

| Дерево правил | | | | | |
|---------------------------------------|-----------|------|------------------|--|-------|
| Правило №16; Следствие: Мыло кусковое | | | | | |
| Условие | Поддержка | | Достоверность, % | | Лифт |
| | Кол-во | % | | | |
| Мыло жидкое И ... | 29 | 1,42 | 76,30 | | 3,981 |
| Мыло жидкое | 167 | 8,17 | 83,50 | | 4,356 |
| Стиральный пор... | 29 | 1,42 | 14,90 | | 0,776 |

Рисунок 6 - Дерево ассоциативных правил

Для каждого правила отображаются поддержка и достоверность и лифт. Если дерево построено по условию, то вверху списка отображается условие правила, а список состоит из его следствий. Тогда правила отвечают на вопрос, что будет при таком условии. Если же дерево построено по следствию, то вверху списка отображается следствие правила, а список состоит из его условий.

Эти правила отвечают на вопросы, что нужно, чтобы было заданное следствие или какие товары нужно продать для того, чтобы продать товар из следствия. На вкладке Правила помимо самих ассоциативных правил приводятся их основные расчетные характеристики: поддержка, достоверность и лифт (рисунок 7).

| Правила | | | | | | | | |
|------------------|---------------|------------------------|------------------------|-----------|------|---------------|--------|--|
| Правил: 17 из 17 | | | Фильтр: Без фильтрации | | | | | |
| № | Номер правила | Условие | Следствие | Поддержка | | Достоверность | Лифт | |
| | | | | Кол-во | % | | | |
| 1 | 4 | Стиральный порошок-а | Кондиционер для б | 79 | 3,86 | 54,86 | 12,064 | |
| 2 | 3 | Кондиционер для белья | Стиральный порош | 79 | 3,86 | 84,95 | 12,064 | |
| 3 | 2 | Запасной баллон для о | Освежитель возду | 53 | 2,59 | 59,55 | 5,855 | |
| 4 | 1 | Бумага туалетная | Освежитель возду | 53 | 2,59 | 58,89 | 5,790 | |
| 5 | 9 | Сода кальцинированна | Чистящий порошок | 96 | 4,69 | 72,18 | 5,055 | |
| 6 | 14 | Зубная паста | Чистящий порошок | 28 | 1,37 | 71,79 | 5,028 | |
| | | Сода кальцинированна | | | | | | |
| 7 | 13 | Гель для туалетов | Чистящий порошок | 34 | 1,66 | 70,83 | 4,961 | |
| | | Сода кальцинированна | | | | | | |
| 8 | 11 | Средство от накипи | Чистящий порошок | 76 | 3,72 | 69,72 | 4,883 | |
| 9 | 15 | Мыло кусковое | Мыло жидкое | 25 | 1,22 | 46,30 | 4,734 | |
| | | Отбеливатель | | | | | | |
| 10 | 8 | Салфетки бумажные | Освежитель возду | 50 | 2,44 | 47,17 | 4,638 | |
| 11 | 7 | Средство для мытья пс | Мыло кусковое | 131 | 6,41 | 86,18 | 4,496 | |
| 12 | 12 | Гель для туалетов | Мыло кусковое | 56 | 2,74 | 84,85 | 4,426 | |
| | | Мыло жидкое | | | | | | |
| 13 | 17 | Средство для мытья пс | Мыло кусковое | 37 | 1,81 | 84,09 | 4,387 | |
| | | Средство для чистки п. | | | | | | |
| 14 | 6 | Мыло кусковое | Мыло жидкое | 167 | 8,17 | 42,60 | 4,356 | |
| 15 | 5 | Мыло жидкое | Мыло кусковое | 167 | 8,17 | 83,50 | 4,356 | |

Рисунок 7 - Ассоциативные правила, отсортированные по убыванию лифта

Ассоциативные правила сами по себе, как результат работы некоторого алгоритма, еще не готовы к использованию. Их нужно проинтерпретировать, т.е. понять, какие из ассоциативных правил представляют интерес, действительно ли правила отражают закономерности или наоборот являются артефактом.

Все множество ассоциативных правил можно разделить на три вида:

Полезные правила – содержат действительную информацию, которая ранее была неизвестна, но имеет логичное объяснение. Такие правила могут быть использованы для принятия решений, приносящих выгоду.

Тривиальные правила – содержат действительную и легко объяснимую информацию, которая уже известна. Такие правила, хотя и объяснимы, но не могут принести какой-либо пользы, т.к. отражают или известные законы в исследуемой области, или результаты прошлой деятельности. При анализе рыночных корзин в правилах с самой высокой поддержкой и достоверностью окажутся товары-лидеры продаж. Практическая ценность таких правил крайне низка.

Непонятные правила – содержат информацию, которая не может быть объяснена. Такие правила могут быть получены или на основе аномальных значений, или глубоко скрытых знаний. Напрямую такие правила нельзя использовать для принятия решений, т.к. их необъяснимость может привести к непредсказуемым результатам. Для лучшего понимания требуется дополнительный анализ.

Варьируя верхним и нижним пределами поддержки и достоверности, можно избавиться от очевидных и неинтересных закономерностей. Как следствие, правила, генерируемые алгоритмом, принимают приближенный к реальности вид. Понятия «верхний» и «нижний» предел очень сильно зависят от предметной области, поэтому не существует четкого алгоритма их выбора. Но есть ряд общих рекомендаций.

1. Большая величина параметра Максимальная поддержка означает, что алгоритм будет находить хорошо известные правила, или они будут настолько очевидными, что в них нет никакого смысла. Поэтому ставить порог максимальная поддержка очень высоким (более 20%) не рекомендуется.

2. Большинство интересных правил находится именно при низком значении порога поддержки, хотя слишком низкое значение поддержки ведет к генерации статистически необоснованных правил. Поэтому правила, которые кажутся интересными, но имеют низкую поддержку, дополнительно анализируйте по лифту, а при необходимости рассчитывайте для них левередж и улучшение.

3. Ограничивайте мощность часто встречающихся множеств – правила с большим числом предметов в условии трудно интерпретируются и воспринимаются.

4. Уменьшение порога достоверности приводит к увеличению количества правил. Значение минимальной достоверности не должно быть слишком маленьким, так как ценность правила с достоверностью 5% чаще всего настолько мала, что это и правилом считать нельзя.

5. Правило с очень большой достоверностью (>85-90%) практической ценности в контексте решаемой задачи не имеет, т.к. товары, входящие в следствие, покупатель, скорее всего, уже купил.

Розничная сеть: интерпретируем ассоциативные правила

Продолжим рассматривать пример по стимулированию продаж розничной сети, занимающейся продажей бытовой химии. При настройках алгоритма a priori по умолчанию мы получили 17 правил (рисунок 8, сценарий Анализ покупательской корзины.ded).

Например, первое правило Бумага туалетная → Освежитель воздуха имеет $S = 2,59\%$; $C = 58,89\%$ и $L = 5,79$.

Это означает следующее:

- ожидаемая вероятность покупки набора Бумага туалетная + Освежитель воздуха равна 2,59%;
- если клиент положил в корзину товар Бумага туалетная, то с вероятностью 58,89% он купит и Освежитель воздуха;
- клиент, купивший Бумагу туалетную, в 5,8 раз чаще выберет Освежитель воздуха, нежели любой другой товар.

Анализ полученных правил позволяет прийти к выводу, что многие из них тривиальны:

– Мыло кусковое, Чистящий порошок, Зубная паста, Гель для туалетов часто встречаются в условиях и следствиях правил, это лидеры продаж магазина (см. популярные наборы), поэтому и правила с ними имеют высокую достоверность;

– группа правил Стиральный порошок-автомат → Кондиционер для белья и наоборот тривиальны сами по себе: люди часто покупают эти товары вместе.

– правила типа Запасной баллон для освежителя → Освежитель воздуха (и наоборот) тоже тривиальны, так как никому не нужен запасной баллон без освежителя.

– правило Салфетки бумажные → Освежитель воздуха не понятное: почему салфетки бумажные покупаются именно с освежителем воздуха.

| Правила | | | | | | | |
|------------------|----|------------------------------|------------------------------|-----------|------|-------|--------|
| Правил: 17 из 17 | | | Фильтр: Без фильтрации | | | | |
| № | No | Условие | Следствие | Поддержка | | Дост | Лифт |
| | | | | Кол-во | % | | |
| T 1 | 6 | Мыло кусковое | Мыло жидкое | 167 | 8,17 | 42,60 | 4,356 |
| T 2 | 15 | Мыло кусковое | Мыло жидкое | 25 | 1,22 | 46,30 | 4,734 |
| | | Отбеливатель | | | | | |
| ? 3 | 8 | Салфетки бумажные | Освежитель воздуха | 50 | 2,44 | 47,17 | 4,638 |
| T 4 | 10 | Средство для чистки кафеля | Чистящий порошок универсальн | 42 | 2,05 | 48,28 | 3,381 |
| T 5 | 4 | Стиральный порошок-автомат | Кондиционер для белья | 79 | 3,86 | 54,86 | 12,064 |
| T 6 | 1 | Бумага туалетная | Освежитель воздуха | 53 | 2,59 | 58,89 | 5,790 |
| T 7 | 2 | Запасной баллон для освежите | Освежитель воздуха | 53 | 2,59 | 59,55 | 5,855 |
| ? 8 | 11 | Средство от накипи | Чистящий порошок универсальн | 76 | 3,72 | 69,72 | 4,883 |
| T 9 | 13 | Гель для туалетов | Чистящий порошок универсальн | 34 | 1,66 | 70,83 | 4,961 |
| | | Сода кальцинированная | | | | | |
| ? 10 | 14 | Зубная паста | Чистящий порошок универсальн | 28 | 1,37 | 71,79 | 5,028 |
| | | Сода кальцинированная | | | | | |
| ? 11 | 9 | Сода кальцинированная | Чистящий порошок универсальн | 96 | 4,69 | 72,18 | 5,055 |
| T 12 | 16 | Мыло жидкое | Мыло кусковое | 29 | 1,42 | 76,32 | 3,981 |
| | | Стиральный порошок ручной | | | | | |
| T 13 | 5 | Мыло жидкое | Мыло кусковое | 167 | 8,17 | 83,50 | 4,356 |
| T 14 | 17 | Средство для мытья посуды | Мыло кусковое | 37 | 1,81 | 84,09 | 4,387 |
| | | Средство для чистки плит | | | | | |
| T 15 | 12 | Гель для туалетов | Мыло кусковое | 56 | 2,74 | 84,85 | 4,426 |
| | | Мыло жидкое | | | | | |
| T 16 | 3 | Кондиционер для белья | Стиральный порошок-автомат | 79 | 3,86 | 84,95 | 12,064 |
| T 17 | 7 | Средство для мытья посуды | Мыло кусковое | 131 | 6,41 | 86,18 | 4,496 |

Рисунок 8 - Первая попытка интерпретации правил (Т – тривиальное правило, ? – непонятное правило)

Тот факт, что при достоверности 42-43% встречаются тривиальные ассоциативные правила, говорит о том, что интересные правила содержатся при меньших значениях достоверности.

Необходимо сделать следующее:

- запустим алгоритм `apriori` с интервалом допустимой достоверности от 25% до 40%;

- не будем рассматривать правила, в следствиях и условиях которых содержатся мыло кусковое, чистящий порошок универсальный: это снова будут тривиальные правила (см. сценарий Анализ покупательской корзины.ded).

В итоге получим дополнительные правила (рисунок 9).

| Правила | | | | | | | |
|------------------|----|--------------------------------|--|-----------|------|--------|-------|
| Правил: 14 из 46 | | | Фильтр: Условие содержит [Антистатик спрей, Бумага туалетная, Дозатор, Запас | | | | |
| № | | Условие | Следствие | Поддержка | | Достов | Лифт |
| | | | | Кол-во | % | | |
| 1 | 14 | Пена/соль для ванн | Запасной баллон для освежителя | 28 | 1,37 | 31,82 | 7,311 |
| 2 | 13 | Запасной баллон для освежителя | Пена/соль для ванн | 28 | 1,37 | 31,46 | 7,311 |
| 3 | 30 | Средство для чистки плит | Салфетки бумажные | 38 | 1,86 | 29,46 | 5,683 |
| 4 | 29 | Салфетки бумажные | Средство для чистки плит | 38 | 1,86 | 35,85 | 5,683 |
| 5 | 43 | Мыло кусковое | Средство для мытья посуды | 21 | 1,03 | 38,89 | 5,232 |
| | | Отбеливатель | | | | | |
| 6 | 33 | Средство для чистки плит | Средство для мытья посуды | 44 | 2,15 | 34,11 | 4,589 |
| 7 | 32 | Средство для мытья посуды | Средство для чистки плит | 44 | 2,15 | 28,95 | 4,589 |
| 8 | 46 | Средство для чистки плит | Мыло кусковое | 37 | 1,81 | 28,68 | 4,477 |
| | | | Средство для мытья посуды | | | | |
| 9 | 45 | Мыло кусковое | Средство для чистки плит | 37 | 1,81 | 28,24 | 4,477 |
| | | Средство для мытья посуды | | | | | |
| 10 | 35 | Антистатик спрей | Мыло кусковое | 28 | 1,37 | 28,00 | 4,371 |
| | | | Средство для мытья посуды | | | | |
| 11 | 44 | Пена/соль для ванн | Мыло кусковое | 23 | 1,12 | 26,14 | 4,080 |
| | | | Средство для мытья посуды | | | | |
| 12 | 2 | Антистатик спрей | Средство для мытья посуды | 30 | 1,47 | 30,00 | 4,036 |
| 13 | 28 | Пена/соль для ванн | Средство для мытья посуды | 25 | 1,22 | 28,41 | 3,822 |
| 14 | 27 | Пятновыводитель | Отбеливатель | 41 | 2,00 | 28,08 | 2,886 |

Рисунок 9 - Полезные правила в примере имеют достоверность меньше 40%

Все полученные правила можно назвать полезными: они не очевидны, но понятны.

Например, возьмем правило Пятновыводитель→Отбеливатель. Проанализируем это правило с помощью лифта. Его величина равна 2,886, что больше чем 1, значит, с помощью правила предсказать покупку отбеливателя вероятнее, чем случайным угадыванием.

Как можно применить на практике это правило? Это зависит от конкретных целей. Приведем всевозможные варианты.

| Пятновыводитель → Отбеливатель |
|--|
| 1. Разместите их рядом на витрине |
| 2. Разместите их на большом расстоянии друг от друга |
| 3. Сформируйте подарочные наборы «пятновыводитель + отбеливатель» |
| 4. Сформируйте подарочные наборы «пятновыводитель + отбеливатель + плохо продаваемый товар» |
| 5. Поднимите цену на одно, снизьте на другое |
| 6. Не рекомендуем покупать к пятновыводителю отбеливатель |
| 7. Закажите комплекты пятновыводителей и отбеливателей одного бренда и серии |

«Что-если» в ассоциативных правилах

К узлу Ассоциативные правила доступен визуализатор Что-если. Он позволяет ответить на вопрос, что мы получим в качестве следствия, если выберем данные условия, например какие товары, приобретаются совместно с выбранными товарами (рисунок 10).

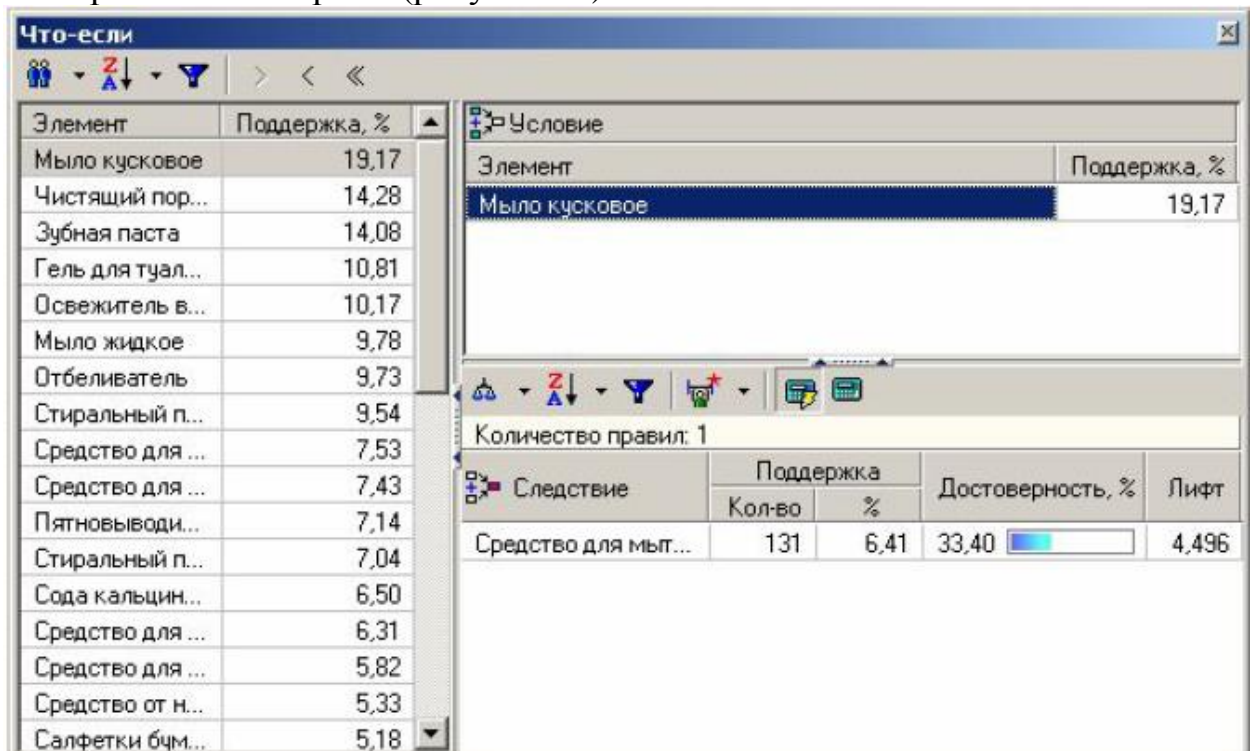






Рисунок 10 - Визуализатор «Что-если»

В окне слева расположен список всех элементов транзакций. Справа от каждого элемента указана поддержка: сколько раз данный элемент встречается в транзакциях.

В правом верхнем углу расположен список элементов, входящих в условие, выбирается он с помощью нажатия мышки или вспомогательных кнопок , , . Это, например, список товаров, которые приобрел покупатель. Для них можно найти следствие, нажав на кнопку  **Вычислить правила**. В условие могут входить несколько элементов, или товаров в данном случае. Тогда в следствие попадут все товары, условия которых удовлетворяют списку ассоциативных правил.

Например, клиент заказал Зубную пасту и Пятновыводитель. Что еще ему можно предложить? Поскольку у нас имеются два правила, а именно (1) Зубная паста → Чистящий порошок, (2) Пятновыводитель → Отбеливатель то в следствие попадут два элемента – Чистящий порошок и Отбеливатель.

Это проиллюстрировано на рисунке 11.

| Что-если | |
|--------------------------------|--------------|
| Элемент | Поддержка, % |
| Мыло кусковое | 19,17 |
| Чистящий порошок универса... | 14,28 |
| Зубная паста | 14,08 |
| Гель для туалетов | 10,81 |
| Освежитель воздуха | 10,17 |
| Мыло жидкое | 9,78 |
| Отбеливатель | 9,73 |
| Стиральный порошок ручной | 9,54 |
| Средство для ухода за мебел... | 7,53 |
| Средство для мытья посуды | 7,43 |
| Пятновыводитель | 7,14 |
| Стиральный порошок-автомат | 7,04 |
| Сода кальцинированная | 6,50 |
| Средство для чистки плит | 6,31 |
| Средство для мытья пола | 5,82 |
| Средство от накипи | 5,33 |
| Салфетки бумажные | 5,18 |
| Антистатик спрей | 4,89 |
| Перчатки резиновые | 4,74 |

| Условие | |
|-----------------|--------------|
| Элемент | Поддержка, % |
| Зубная паста | 14,08 |
| Пятновыводитель | 7,14 |


Количество правил: 2



| Следствие | Поддержка | | Достоверность, % | Лифт |
|---------------|-----------|------|------------------|-------|
| | Кол-во | % | | |
| Чистящий п... | 77 | 3,77 | 26,70 | 1,872 |
| Отбеливатель | 41 | 2,00 | 28,10 | 2,886 |


Рисунок 11 - Что еще предложить этому клиенту?


Информация о купленных товарах вносится продавцом в отчет и тут же формируется набор предложений, который озвучивается клиенту.

С помощью встроенных возможностей можно настроить интерактивный отчет возможных предложений-напоминаний клиенту. Для этого используются следующие кнопки:

 **Автоматически вычислить правила** – позволяет рассчитывать новый набор следствий после каждого добавления товара в набор условий;

 **Порядок сортировки**,  **Направление сортировки** – позволяют упорядочить «список предложений» по одному из выбранных параметров (поддержка, достоверность, лифт);

 **Фильтрация правил** – задаются ограничения на формируемый «список предложений»;

 **Тип определения лучшего правила** – при повторении наименований товаров в «списке предложений» позволяет отображать, один из них, в соответствии с лучшей характеристикой.

Использование узла Ассоциативные правила как модель

После выполнения обработчика Ассоциативные правила формируется два набора данных.

Один из них формируется на выходе узла по умолчанию и его можно просмотреть с помощью визуализатора Таблица, рисунок 12.

| ID | ITEM | Следствие | Номер правила | Поддержка, шт. | Поддержка, % | Достоверность, % | Лифт |
|---------|---------------------------|-------------------------------------|---------------|----------------|--------------|------------------|--------|
| S051184 | Гель для туалетов | <input type="checkbox"/> | | | | | |
| S051184 | Микроспрей | <input type="checkbox"/> | | | | | |
| S051184 | Сода кальцинированная | <input type="checkbox"/> | | | | | |
| S051184 | Чистящий порошок универс. | <input type="checkbox"/> | | | | | |
| S051184 | Мыло жидкое | <input checked="" type="checkbox"/> | 5 | 66 | 3,22738386 | 29,8642533937 | 3,0536 |
| S051184 | Мыло кусковое | <input checked="" type="checkbox"/> | 7 | 56 | 2,73838631 | 25,3393665158 | 1,3219 |
| S051184 | Стиральный порошок ручн. | <input checked="" type="checkbox"/> | 9 | 66 | 3,22738386 | 29,8642533937 | 3,1319 |
| S051184 | Зубная паста | <input checked="" type="checkbox"/> | 17 | 39 | 1,90709046 | 29,3233082707 | 2,0822 |
| S051184 | Зубная паста | <input checked="" type="checkbox"/> | 20 | 77 | 3,76528117 | 26,3698630137 | 1,8724 |
| S051184 | Средство от накипи | <input checked="" type="checkbox"/> | 34 | 76 | 3,71638142 | 26,0273972603 | 4,8831 |
| S051184 | Мыло жидкое | <input checked="" type="checkbox"/> | 37 | 56 | 2,73838631 | 25,3393665158 | 3,1029 |
| S051184 | Мыло кусковое | <input checked="" type="checkbox"/> | 37 | 56 | 2,73838631 | 25,3393665158 | 3,1029 |

Рисунок 12 - Визуализатор «Таблица» для узла «Ассоциативные правила»

В исходном наборе данных, изначально было всего два поля: ID и ITEM, после запуска обработчика добавились новые расчетные поля (таблица 1).

Таблица 1 – Новые поля выходного набора данных для узла «Ассоциативные правила»

| N | Имя столбца | Метка столбца | Тип поля | Описание |
|---|--------------|------------------|--------------|---|
| 1 | CONSEQUENT | Следствие | логический | Следствие ассоциативного правила, да/нет |
| 2 | RULENUMBER | Номер правила | целый | Номер ассоциативного правила |
| 3 | SUPPORTCOUNT | Поддержка, шт. | целый | Число транзакций, удовлетворяющих данному правилу |
| 4 | SUPPORT | Поддержка, % | вещественный | Поддержка ассоциативного правила в % |
| 5 | CONFIDENCE | Достоверность, % | вещественный | Достоверность ассоциативного правила в % |
| 6 | LIFT | Лифт | вещественный | Лифт ассоциативного правила |

Дополнительный набор данных можно получить, вызвав зависимый обработчик Правила (рисунок 13).

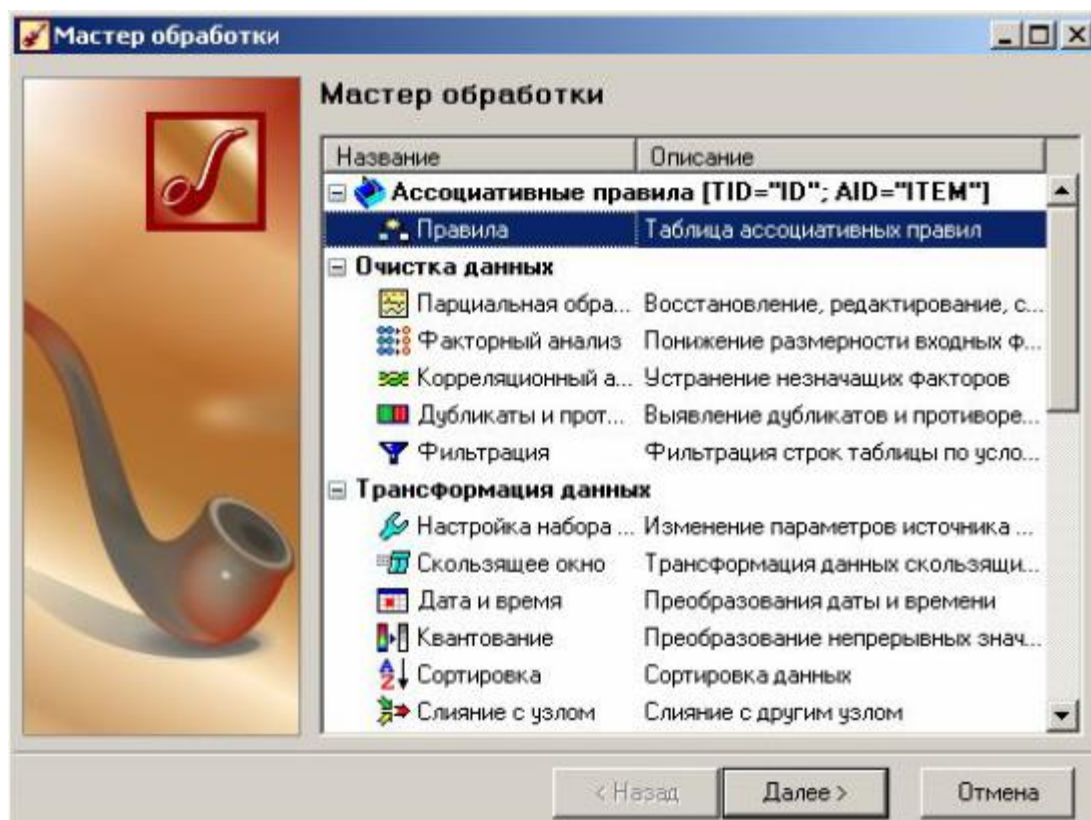


Рисунок 13 - Выбор зависимого обработчика «Правила»

Результирующий визуализатор Таблица для зависимого обработчика Правила имеет вид, показанный на рисунке 14. Он является более наглядным для анализа по сравнению с первым набором, поскольку сразу видно, где условие, а где следствие.

| Таблица | | | | | | | |
|---------------|----------------|-----------------------|--------------------|-----------|-------|---------------|----------|
| Номер правила | Номер элемента | Условие | Следствие | Поддержка | | Достоверность | Лифт |
| | | | | Кол-во | % | | |
| 36 | 2 | Мыло кусковое | | 56 | 2,738 | 33,532934132 | 3,10293 |
| 37 | 1 | Гель для туалетов | Мыло жидкое | 56 | 2,738 | 25,339366516 | 3,10293 |
| 37 | 2 | | Мыло кусковое | 56 | 2,738 | 25,339366516 | 3,10293 |
| 38 | 1 | Мыло жидкое | Гель для туалетов | 56 | 2,738 | 28,000000000 | 10,22500 |
| 38 | 2 | | Мыло кусковое | 56 | 2,738 | 28,000000000 | 10,22500 |
| 39 | 1 | Сода кальцинированная | Гель для туалетов | 34 | 1,663 | 35,416666667 | 3,27724 |
| 39 | 2 | Чистящий порошок уни | | 34 | 1,663 | 35,416666667 | 3,27724 |
| 40 | 1 | Сода кальцинированная | Гель для туалетов | 34 | 1,663 | 25,563909774 | 15,37594 |
| 40 | 2 | | Чистящий порошок | 34 | 1,663 | 25,563909774 | 15,37594 |
| 41 | 1 | Зубная паста | Сода кальцинирован | 28 | 1,369 | 36,363636364 | 5,59125 |
| 41 | 2 | Чистящий порошок уни | | 28 | 1,369 | 36,363636364 | 5,59125 |
| 42 | 1 | Сода кальцинированная | Зубная паста | 28 | 1,369 | 29,166666667 | 2,07104 |

Рисунок 14 - Визуализатор «Таблица» для зависимого обработчика «Правила»

Структура набора данных представлена в таблице 2.

Таблица 2 – Структура набора данных

| N | Имя столбца | Метка столбца | Тип поля | Описание |
|---|--------------|------------------|--------------|---|
| 1 | RULENUMBER | Номер правила | целый | Номер ассоциативного правила |
| 2 | ITEMNUMBER | Номер элемента | целый | Номер следствия входящего в предметный набор или количество условий |
| 3 | ANTECEDENT | Условие | строковый | Условие ассоциативного правила, заключено в "" |
| | CONSEQUENT | Следствие | строковый | Следствие ассоциативного правила, заключено в "" |
| | SUPPORTCOUNT | Поддержка Кол-во | целый | Число транзакций, удовлетворяющих данному правилу |
| 4 | SUPPORT | Поддержка % | вещественный | Поддержка ассоциативного правила в % |
| 5 | CONFIDENCE | Достоверность, % | вещественный | Достоверность ассоциативного правила в % |
| 6 | LIFT | Лифт | вещественный | Лифт ассоциативного правила |

Анализ исторических продаж – интересная и полезная задача, но довольно часто встает задача организации советующей системы, то есть быстрая автоматическая обработка новых чеков через уже сформированные правила.

Предположим такую ситуацию: есть организация, занимающая продажей товаров для предприятий. За каждым менеджером закреплено несколько десятков предприятий-покупателей.

Предприятия обычно заказывают продукцию 1 раз в месяц, а дозаказ осуществляется до обработки заявки, либо переносится на следующий месяц, в данном случае организация-продавец теряет прибыль. Менеджеры предприятия должны вовремя проверить заявку и предложить товар, который, возможно, забыли купить.

Наименований товарных групп очень много, и менеджеры продаж не в состоянии помнить их все, соответственно, упущенные выгоды возрастают с каждой ошибкой менеджера. Для этого можно в автоматическом режиме выявлять товары, которые следует предложить клиентам. Использование визуализатора «Что-Если» не совсем удобно, если наименований товара в заявке больше десяти, и объем заказов возрастает каждую минуту. Тогда лучшим решением будет обработка всех поступивших заявок на определение возможных недостающих товаров в корзине покупателя при помощи ассоциативных правил (рисунок 14).

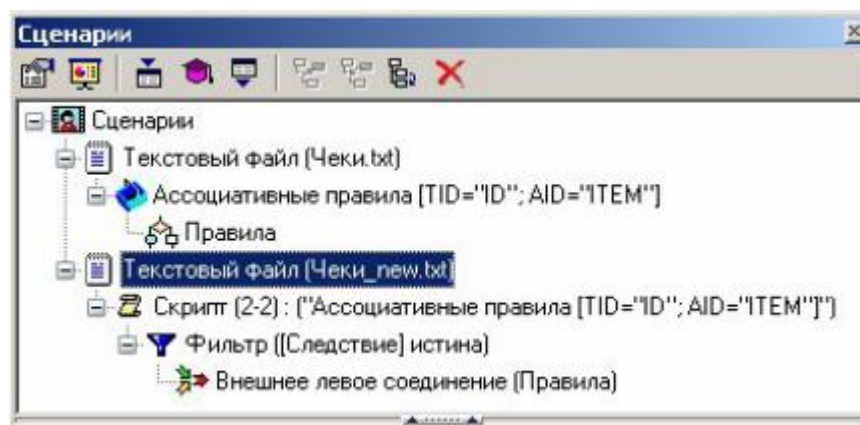


Рисунок 14 - Сценарий, иллюстрирующего использование узла «Ассоциативные правила» как модель

В первой ветке сценария импортируются исторические продажи, на основе которых выявляются основные ассоциативные правила. Во второй ветке происходит импорт новых заявок и их обработка с помощью узла Скрипт по выявленным до этого ассоциативным правилам. Пусть новая заявка расположена в файле Чеки_new.txt (рисунок 15). Результат работы обработчика

Скрипт через единственный узел Ассоциативные правила показан на рисунке 16.

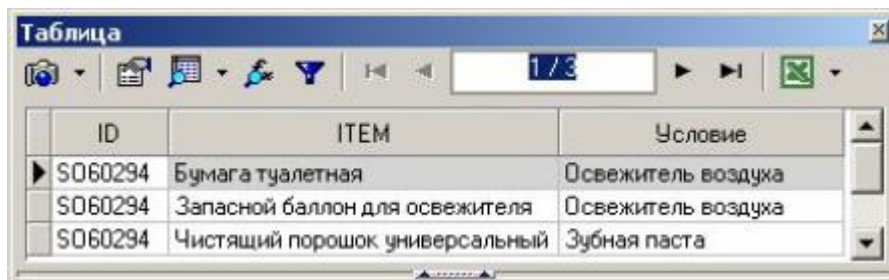
| ID | ITEM |
|---------|--------------------|
| SO60294 | Перчатки резиновые |
| SO60294 | Зубная паста |
| SO60294 | Освежитель воздуха |

Рисунок 15 - Новая заявка

| ID | ITEM | Следствие | Номер правила | Поддержка, шт. | Поддержка, % | Достоверность, % | Лифт |
|---------|------------------------------|-------------------------------------|---------------|----------------|--------------|------------------|-------|
| SO60294 | Зубная паста | <input type="checkbox"/> | | | | | |
| SO60294 | Освежитель воздуха | <input type="checkbox"/> | | | | | |
| SO60294 | Перчатки резиновые | <input type="checkbox"/> | | | | | |
| SO60294 | Бумага туалетная | <input checked="" type="checkbox"/> | 4 | 53 | 2,592 | 25,481 | 5,790 |
| SO60294 | Запасной баллон для освежите | <input checked="" type="checkbox"/> | 12 | 53 | 2,592 | 25,481 | 5,855 |
| SO60294 | Чистящий порошок универсаль | <input checked="" type="checkbox"/> | 19 | 77 | 3,765 | 26,736 | 1,872 |

Рисунок 16 - Результат работы обработчика «Скрипт»

Полученный набор данных содержит следствие и условие вперемешку. Для удобства работы конечного пользователя производят фильтрацию только по следствию. Наличие условий в данном наборе бессмысленно, так как не понятно по какому условию именно это следствие. Для правильной интерпретации и обоснования предлагаемой покупки используют зависимый обработчик Правила, который при слиянии с полученным набором данных дает объяснение выбора предлагаемого товара к уже зарезервированному. Результат работы сценария можно просмотреть на рисунке 17.



| ID | ITEM | Условие |
|---------|--------------------------------|--------------------|
| SO60294 | Бумага туалетная | Освежитель воздуха |
| SO60294 | Запасной баллон для освежителя | Освежитель воздуха |
| SO60294 | Чистящий порошок универсальный | Зубная паста |

Рисунок 17 - Обоснование выбора предлагаемого товара

Полученный набор можно экспортировать во внешний источник.

Задания для самостоятельной работы:

- 1 Изучите сценарий **Анализ покупательской корзины.ded**.
- 2 Поэкспериментируйте с настройкой фильтра в визуализаторах **Правила, Популярные наборы и Дерево правил**.
- 3 Добавьте новую ветвь Чеки_new.txt в сценарий **Анализ покупательской корзины.ded** и найти все имеющиеся ассоциативные правила.
- 4 Рассчитайте левередж для правила Пятновыводитель → Отбеливатель первой ветви сценария.
- 5 Определите, что приобретет клиент купивший: гель для туалетов; средство для мытья посуды и пятновыводитель с величиной лифта больше 3.
- 6 Сделайте интерактивный отчет, в котором покупателю на основе уже совершенных покупок предлагался бы товар со скидкой (мыло кусковое). Например, на все чистящие средства есть скидка, в соответствии с эти клиенту купившему мыло кусковое предлагается купить средство для мытья посуды со скидкой.