Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение высшего образования

ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

кафедра Автоматизированных систем

ОТЧЕТ

к лабораторной работе по дисциплине:

|  |
| --- |
| **Методы и средства проектирования информационных систем и технологий** |
| Применение методов и средств Data Mining |

наименование темы

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Выполнил студент группы: | ИСМб-19-1 |  |  |  | Вовиков Д.Е. |
|  |  | подпись |  | Фамилия И.О |
| Проверил: |  |  |  |  | Юрин А.Ю. |
|  | должность |  | подпись |  | Фамилия И.О |

Иркутск 2022 г.

Содержание

[Содержание 2](#_Toc115903609)

[1 Постановка задачи 3](#_Toc115903610)

[2 Неформальное описание предметной области 4](#_Toc115903611)

[3 Перечень полезных, тривиальных и неясных правил. 4](#_Toc115903612)

[4 Значения достоверности и поддержки для получения полезных правил. 5](#_Toc115903613)

[5 Визуализация результатов при помощи правил и дерева правил. 6](#_Toc115903614)

[6 Рекомендации для принятия решения. 6](#_Toc115903615)

1 Постановка задачи

Цель работы: выработка и закрепление навыков по применению специализированных алгоритмических и программных средств Data Mining.

Для выбранного варианта задания необходимо:

1. Ознакомится с методическим руководством по выполнению лабораторной работы.

2. Загрузить файл исходных данных в программное средство Deductor.

3. Определить ассоциативные правила, изменяя значения достоверности и поддержки.

4. Сравнить полученные результаты.

5. Указать полезные, тривиальные и неясные правила.

6. Определить наиболее походящие значения достоверности и поддержки для получения полезных правил.

7. Интерпретировать полученные правила и дать рекомендации для принятия решения.

8. Оформить отчет по результатам исследования.

**Вариант**

|  |  |
| --- | --- |
| **N** | **Предметная область/Задача** |
| 2. | Чек из супермаркета 2 |

2 Неформальное описание предметной области

В данной предметной области находится информация о содержимом чеков из супермаркета — это соусы, чай, макаронные изделия и др.

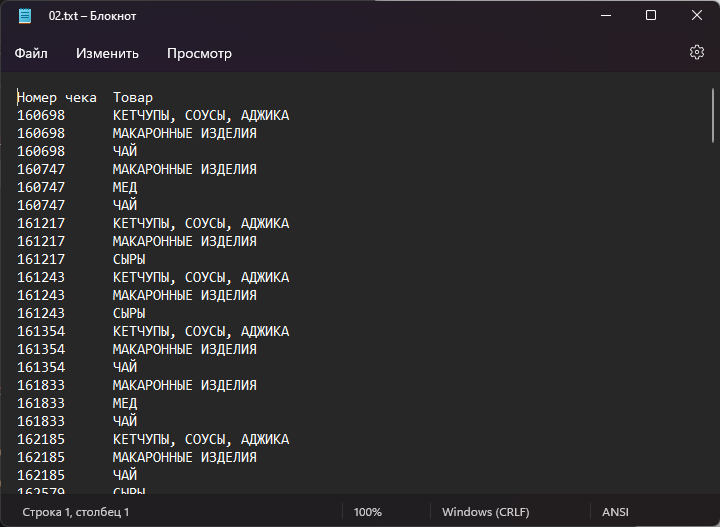


Рисунок 1 - Информация по предметной области

3 Перечень полезных, тривиальных и неясных правил.

Файл с данными согласно варианту задания был импортирован и обработан с помощью специального программного обеспечения Deductor. При импорте исходный данных для корректной обработки информации было использовано разделение с помощью символа табуляции.

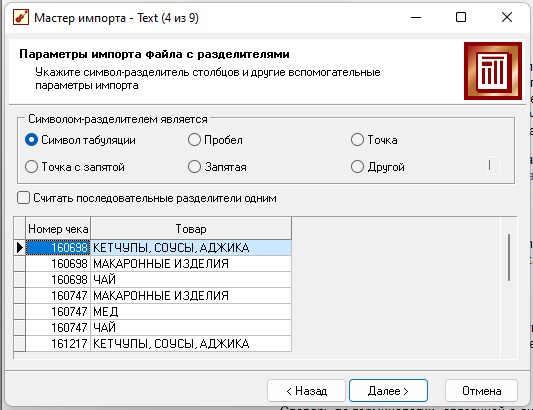


Рисунок 2 – Работа с «Мастером импорта»

В результате обработки информации получились следующие правила, представленные на рисунке 3.

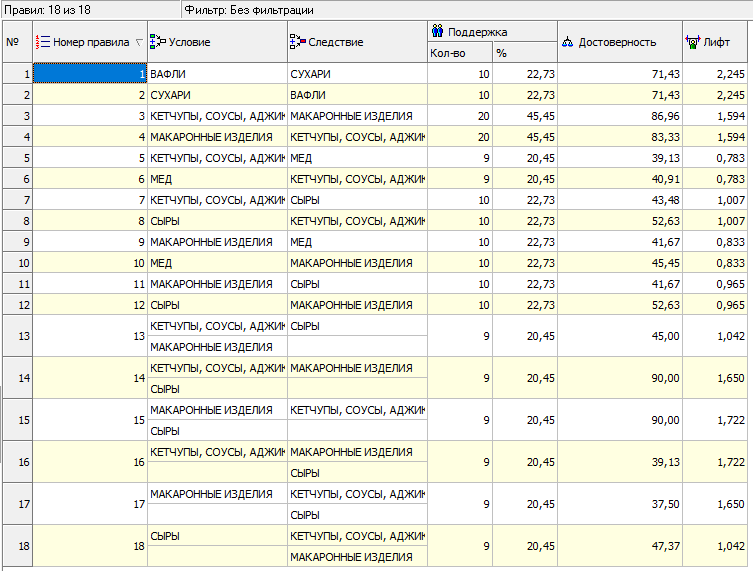


Рисунок 3 – Правила предметной области

Из данных правил можно выделить 3, 4, 7, 8, 10-18 правила полезными, т.к. указывают на возможные верные сочетания продуктов в чеке.

Остальные правила были отнесены к тривиальным, т.к. они отражают нестандартный набор продуктов в чеке.

4 Значения достоверности и поддержки для получения полезных правил.

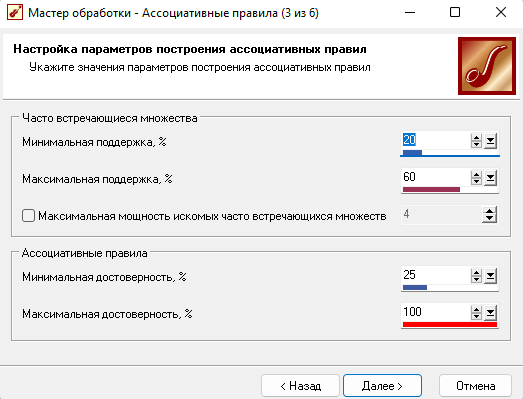


Рисунок 4 - Значения достоверности и поддержки

В качестве значения минимальной поддержки часто встречающихся множеств было выставлено значение равное 20%. В качестве максимального – 60%.

Параметр минимальной достоверности был выставлен на уровне 25%. Максимальной достоверности – 100%.

5 Визуализация результатов при помощи правил и дерева правил.

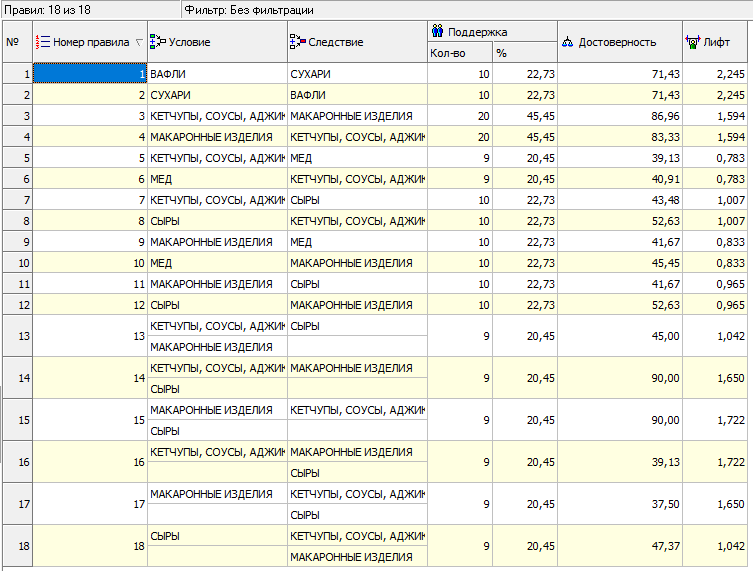


Рисунок 5 – Визуализация результатов при помощи правил

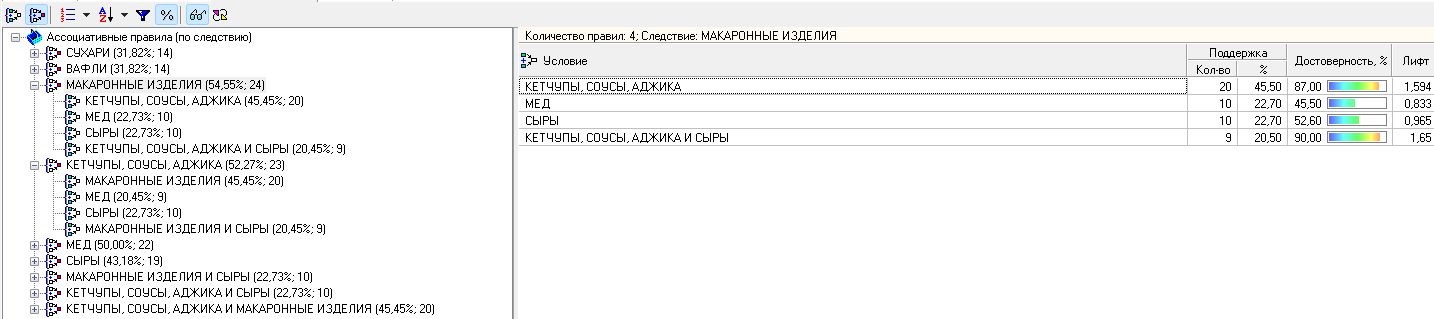


Рисунок 6 - Визуализация результатов при помощи дерева правил

6 Рекомендации для принятия решения.

По результатам обработки информации по данной предметной области можно дать рекомендацию для нахождения полезных правил для данной системы, а именно, это использовать диапазон поддержки от 20 до 60 % и диапазон достоверности от 25 до 100 %.