

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное
учреждение высшего образования
**"Южно-Уральский государственный университет
(национальный исследовательский университет)"**
Высшая школа электроники и компьютерных наук
Кафедра системного программирования

ОТЧЕТ

о выполнении практической работы № 2

по дисциплине

«Технологии аналитической обработки информации»

Выполнил:
студент группы КЭ-403
Гольденберг Д.И.

Проверил:
Преподаватель кафедры СП
Гоглачев А.И.

Челябинск, 2025 г.

ОГЛАВЛЕНИЕ

1. ЗАДАНИЕ	2
2. РЕАЛИЗАЦИЯ АЛГОРИТМА.....	3
3. ЭКСПЕРИМЕНТЫ.....	4
3.1. Загрузка данных и тестовый запуск алгоритма.....	4
3.2. Диаграммы результатов.....	5

1. ЗАДАНИЕ

1. Доработайте программу из задания «Поиск частых наборов», чтобы она также выполняла поиск ассоциативных правил. Список результирующих правил должен выдаваться в удобочитаемом виде (антецедент→консеквент) с указанием поддержки и достоверности каждого правила. Дополнительные параметры программы: порог достоверности, способ упорядочивания результирующего списка наборов (по убыванию значения поддержки или лексикографическое).

2. Проведите эксперименты на наборах из задания 1. В экспериментах зафиксируйте значение пороговое значение поддержки (например, 10%), варьируйте пороговое значение достоверности (например, от 70% до 95% с шагом 5%).

3. Выполните визуализацию полученных результатов в виде следующих диаграмм:

- сравнение быстродействия поиска правил на фиксированном наборе данных при изменяемом пороге достоверности;
- общее количество найденных правил на фиксированном наборе данных при изменяемом пороге достоверности.

4. Подготовьте список правил, в которых антецедент и консеквент суммарно включают в себя не более семи объектов (разумное количество). Проанализируйте и изложите содержательный смысл результата.

5. Подготовьте отчет о выполнении задания и загрузите отчет в формате PDF в систему. Отчет должен представлять собой связный и структурированный документ со следующими разделами:

- формулировка задания;
- гиперссылка на каталог репозитория с исходными текстами, наборами данных и др. сопутствующими материалами;
- рисунки с результатами визуализации;
- пояснения, раскрывающие смысл полученных результатов.

2. РЕАЛИЗАЦИЯ АЛГОРИТМА

Реализация поиска частых наборов представлена в функции `run_apriori`, которая была представлена в предыдущем задании. Дополнительно реализована функция `find_association_rules`, которая:

1. Принимает частые наборы и извлекает ассоциативные правила по метрике `confidence`.
2. Фильтрует правила по заданному порогу достоверности.
3. Позволяет сортировать результаты по поддержке или в лексикографическом порядке.

Код метода `find_association_rules` представлен в листинге 1.

Листинг 1 – Код метода «`find_association_rules`»

```
def find_association_rules(itemsets, min_confidence, sort_by='support'):
    rules = association_rules(itemsets, metric="confidence", min_threshold=min_confidence)

    if sort_by == 'support':
        rules = rules.sort_values(by='support', ascending=False)
    elif sort_by == 'lexical':
        rules = rules.sort_values(by='antecedents')

    return rules[['antecedents', 'consequents', 'support', 'confidence']]
```

Код реализованной программы и всех проведенных экспериментов находится в репозитории по ссылке https://github.com/Goldria/analytics/blob/main/2_associative_rules_search/associative_rules_search.ipynb.

Был проведен анализ правил, в которых антецедент и консеквент суммарно включают не более 7 объектов. Выбранные правила могут быть полезны для выявления закономерностей в покупательском поведении. Код представлен в листинге 2.

Листинг 2 – Код поиска правил

```
min_support = 0.002
min_confidence = 0.65

frequent_itemsets = run_apriori(df, min_support)
rules = find_association_rules(frequent_itemsets, min_confidence=min_confidence)

filtered_rules = rules[rules.apply(lambda row: len(row['antecedents']) + len(row['consequents']) <= 7, axis=1)]
display(filtered_rules)
```

3. ЭКСПЕРИМЕНТЫ

3.1. Загрузка данных и тестовый запуск алгоритма

Набор данных `baskets.csv` содержит список покупок в супермаркете. Для предобработки данных были выполнены следующие шаги:

- Загрузка данных с определением кодировки.
- Очистка от пропущенных значений.
- Формирование списка транзакций.

Тестовый запуск алгоритма проводился с порогом поддержки 0.002 и порогом достоверности 0.7, результат полученных наборов и правил которого представлен на рисунке 1 и 2 соответственно.

support	itemsets
0.238267	(минеральная вода)
0.188000	(макароны)
0.179733	(яйца)
0.170933	(картофель-фри)
0.163867	(шоколад)
...	...
0.002000	(корм для животных, минеральная вода)
0.002000	(креветки, красное вино)
0.002000	(креветки, говяжий фарш, шоколад, макароны)
0.002000	(говяжий фарш, низкокалорийный йогурт, минерал...
0.002000	(травы и перец, говяжий фарш, рис, минеральная...

Рисунок 1 – Полученные наборы

antecedents	consequents	support	confidence
(грибной соус, эскалоп)	(макароны)	0.004267	0.744186
(растительное масло, яйца, макароны)	(минеральная вода)	0.003067	0.718750
(молоко, замороженные овощи, суп)	(минеральная вода)	0.003067	0.766667
(шоколад, оливковое масло, замороженные овощи)	(минеральная вода)	0.002800	0.700000
(молоко, оливковое масло, яйца)	(минеральная вода)	0.002667	0.714286
(креветки, говяжий фарш, замороженные овощи)	(макароны)	0.002533	0.791667
(обезжиренное молоко, макароны)	(минеральная вода)	0.002533	0.730769
(блинчики, растительное масло, макароны)	(минеральная вода)	0.002267	0.739130
(блинчики, суп, макароны)	(минеральная вода)	0.002267	0.772727
(помидоры, оливковое масло, замороженные овощи)	(макароны)	0.002133	0.842105
(индейка, красное вино)	(минеральная вода)	0.002133	0.727273
(блинчики, рис, говяжий фарш)	(минеральная вода)	0.002133	0.842105
(молоко, помидоры, рис)	(макароны)	0.002133	0.800000
(индейка, замороженные овощи, молоко)	(минеральная вода)	0.002000	0.750000
(шоколад, молоко, говяжий фарш, замороженные о...	(минеральная вода)	0.002000	0.750000
(креветки, шоколад, замороженные овощи, макароны)	(минеральная вода)	0.002000	0.882353

Рисунок 2 – Полученные правила

3.2. Диаграммы результатов

Был проведен эксперимент с различными значениями порога достоверности (70%, 75%, 80%, 85%, 90%, 95%). Для каждого значения фиксировалось время выполнения алгоритма, график зависимости времени выполнения от порога достоверности представлен на рисунке 3.

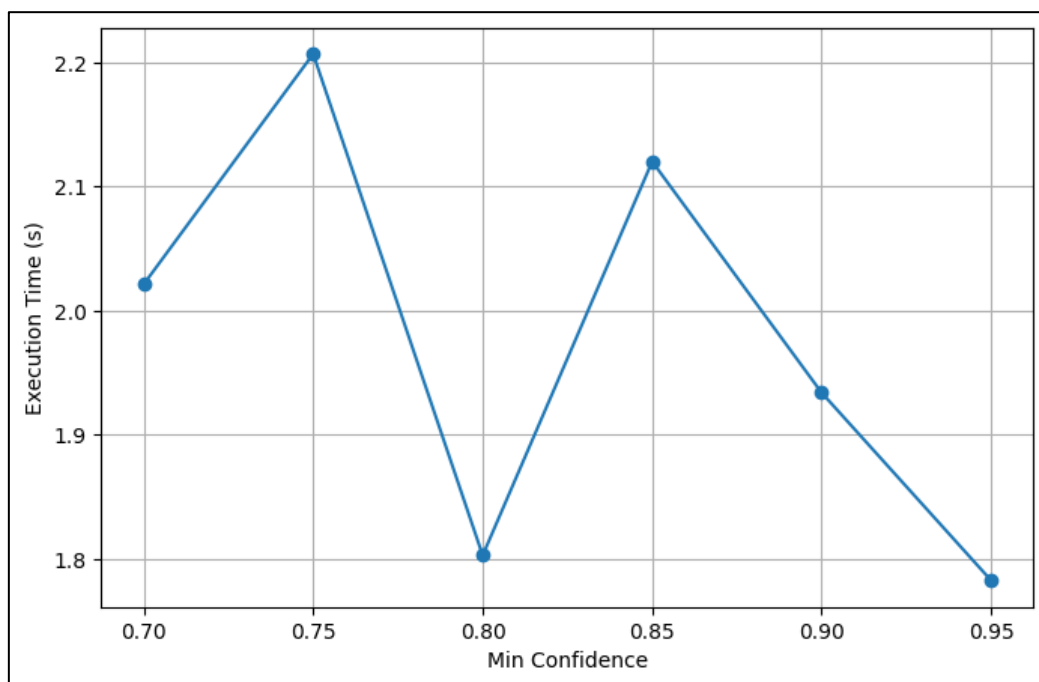


Рисунок 3 – Диаграмма с временем выполнения
По данным графика можно сделать следующие выводы.

1. При увеличении порога достоверности время выполнения алгоритма в среднем уменьшается, но скачкообразно.

2. Это связано с тем, что при высоких значениях confidence количество извлекаемых правил уменьшается.

Также был построен график зависимости числа найденных правил от порога достоверности, который представлен на рисунке 4.

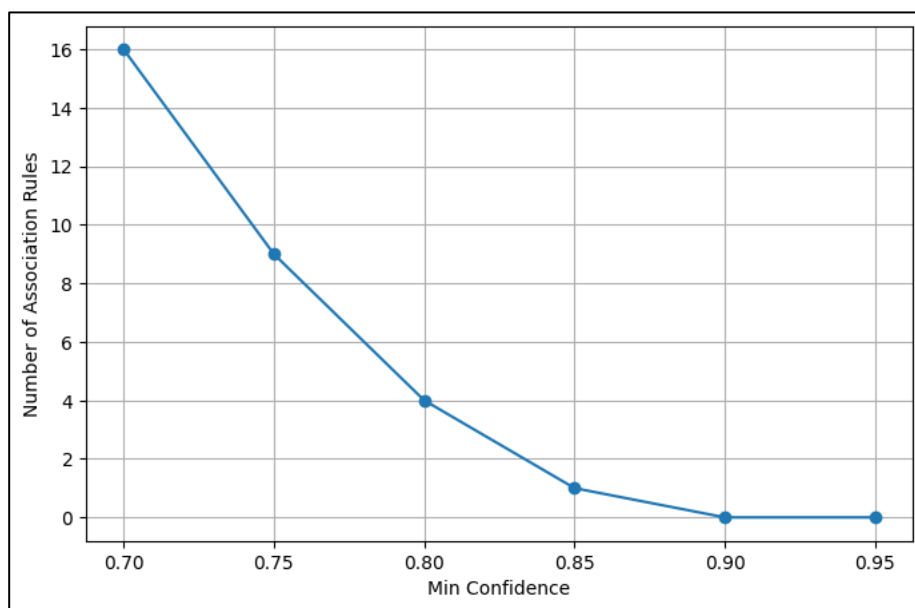


Рисунок 4 – Диаграмма числа правил

По графику можно сделать следующие выводы:

1. При увеличении порога достоверности количество найденных правил уменьшается.

2. При confidence = 0.9 число найденных правил падает до нуля, что говорит о том, что высокая достоверность встречается только у небольшого количества правил.

Также был проведен анализ правил, в которых антецедент и консеквент суммарно включают не более 7 объектов. По результатам большинство найденных правил имеют логическое объяснение: покупатели, приобретающие один товар, часто покупают связанные с ним товары, при ужесточении порога достоверности остаются только наиболее значимые связи.

В ходе работы был реализован алгоритм Apriori для поиска ассоциативных правил. Проведенные эксперименты показали:

- Время выполнения алгоритма снижается при увеличении порога достоверности.
- Число найденных правил уменьшается при увеличении порога достоверности.
- Высокие значения достоверности встречаются только у небольшого количества правил.