**МИНОБРНАУКИ РОССИИ**

**Санкт-Петербургский государственный**

**электротехнический университет**

**«ЛЭТИ» им. В.И. Ульянова (Ленина)**

**Кафедра МОЭВМ**

отчет

**по лабораторной работе №4**

**по дисциплине «Машинное обучение»**

Тема: Ассоциативный анализ

|  |  |  |
| --- | --- | --- |
| Студент гр. 8304 |  | Холковский К.В |
| Преподаватель |  | Жангиров Т. Р. |

Санкт-Петербург

2021

**Цель работы**

Ознакомиться с методами ассоциативного анализа из библиотеки MLxtend

**Ход работы**

**Загрузка данных**

Были загружены данные:

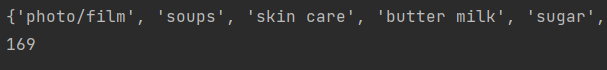


Рис 1 – Загруженные данные

**FPGrowth и FPMax**

Данные были приведены к удобному для анализа виду при помощи TransactionEncoder:

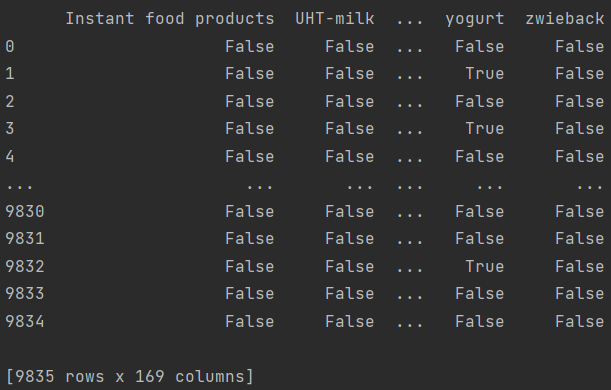


Рис 2 – Удобные для анализа данные

Был проведен ассоциативный анализ используя FPGrowth при уровне поддержки 0.03:



Рис 3 – Результат FPGrowth при minSup=0.03

Был проведен анализ полученных данных:

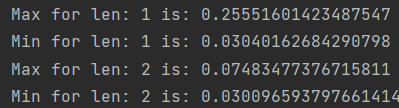
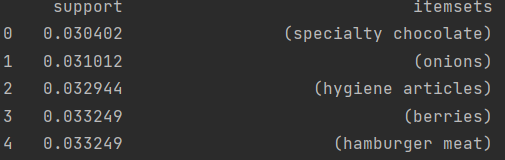


Рис 4 – Анализ полученных данных

Был проведен ассоциативный анализ используя FPMax при уровне поддержки 0.03:





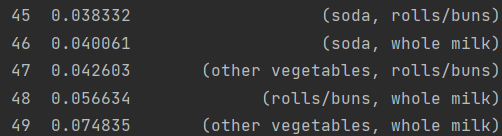


Рис 5 -Результаты FPMax

Был проведен анализ полученных данных:

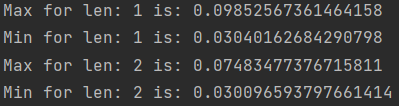


Рис 6 – Анализ полученных данных

Алгоритмы работают одинаково, но в FPMax входят только максимальные наборы элементов. Набор элементов называется максимальным, если он является частым и не существует частого супер-шаблона, содержащего его. Поэтому результаты для наборов длинны 2 не изменились.

Была построена гистограмма для товаров:

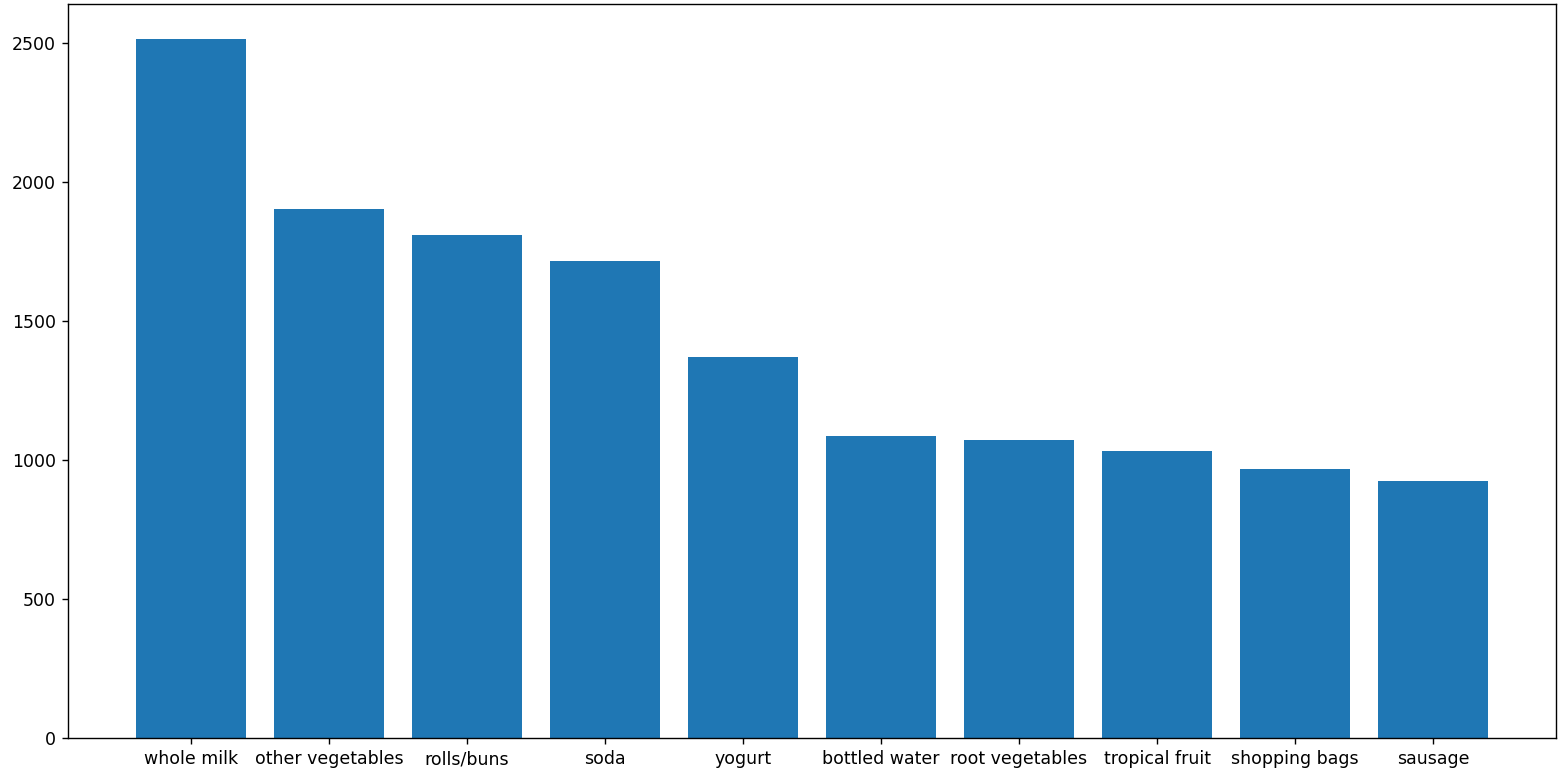
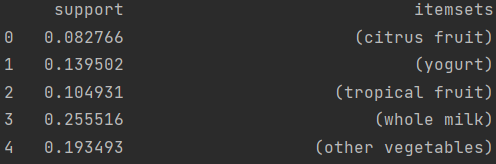


Рис 7 – Гистограмма для всех товаров

Был проведен анализ для нового набора данных:



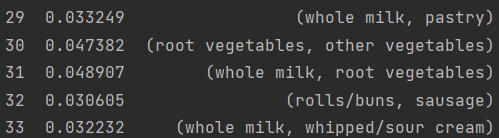


Рис 8 – FPGrowth для нового датасета

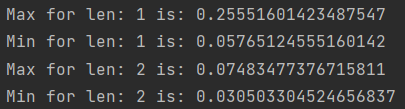


Рис 9 – Анализ полученных данных

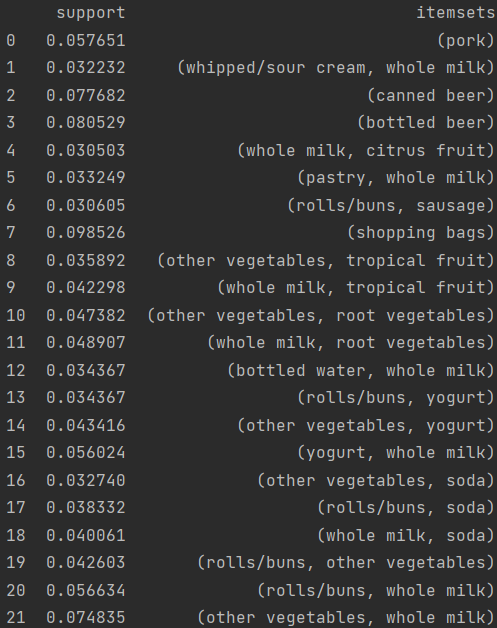


Рис 10 – FPMax для нового датасета

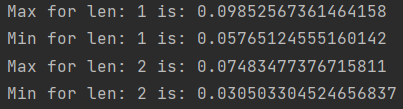


Рис 11 – Анализ полученных данных

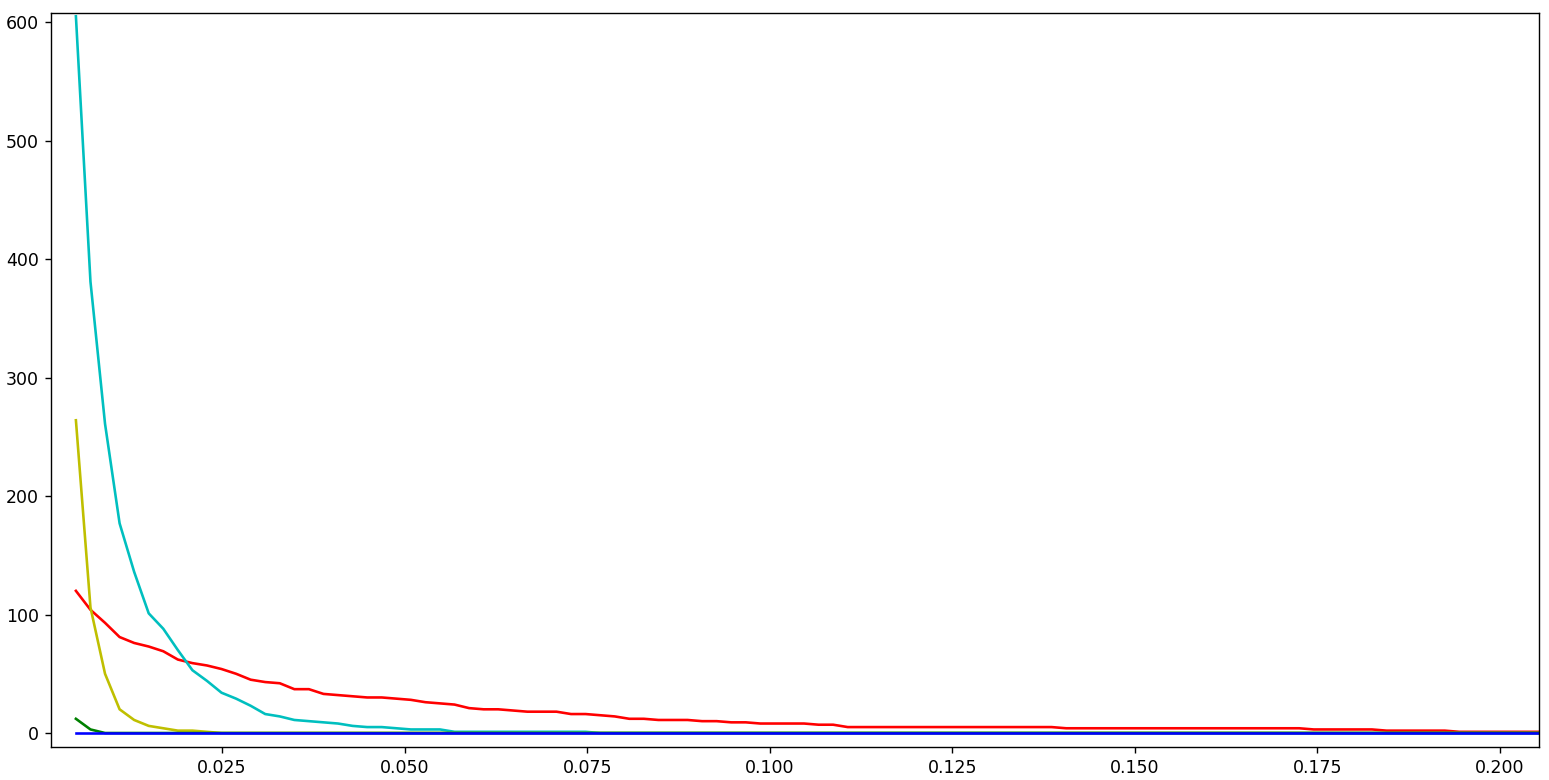


Рис 12 – График изменения количества правил от поддержки для разных длин наборов

1: красный, 2: циановый, 3: желтый, 4: зеленый, 5: blue.

**Ассоциативные правила**

Был выполнен анализ:

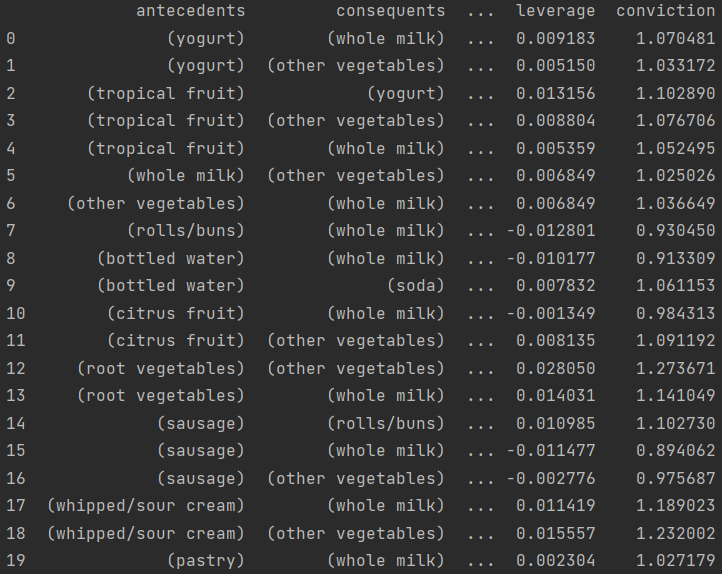


Рис 13 – Результаты анализа

Расчет проводился на основе метрики “confidence”

|  |  |
| --- | --- |
| Метрика | Описание |
| support | support(A→C)=support(A∪C) |
| confidence | confidence(A→C) = |
| lift | lift(A→C)= |
| leverage | levarage(A→C)=support(A→C)−support(A)×support(C) |
| conviction | conviction(A→C)= |

Были расчитаны среднее значение, медиана и СКО для каждой метрики:

|  |  |  |  |
| --- | --- | --- | --- |
| Метрика | Среднее | Медина | СКО |
| support | 0.07468 | 0.06695 | 0.02254 |
| confidence | 0.28957 | 0.26443 | 0.10368 |
| lift | 1.04299 | 1.05608 | 0.18326 |
| leverage | 0.01553 | 0.01359 | 0.00606 |
| conviction | 1.01719 | 1.02285 | 0.08399 |

Был построен граф:

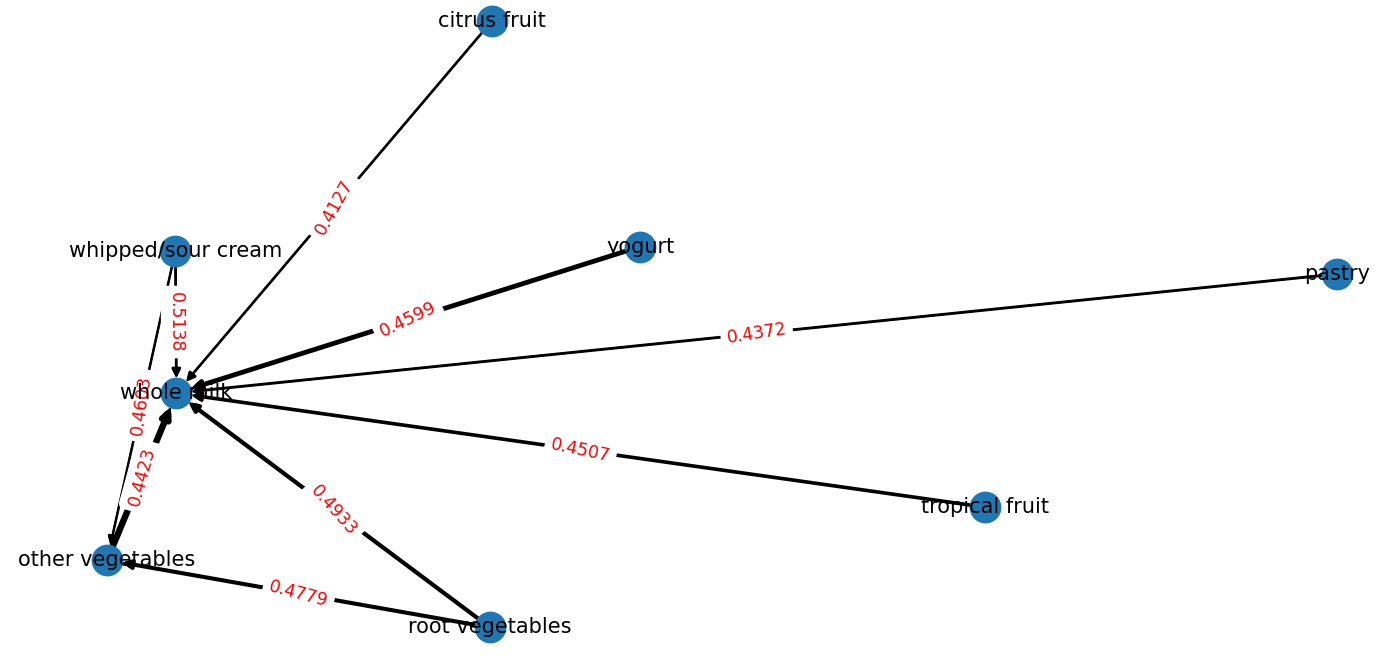


Рис 14 – Граф анализа

**Вывод**

Ознакомились с ассоциативного частотного анализа из библиотеки MLxtend.