

**Разработка алгоритма  
предварительной группировки данных**

**Выполнила:**

Маркова Ангелина Викторовна, гр. 7383

**Руководитель:**

Жукова Наталия Александровна, д.т.н., доцент

**Консультанты:**

Заславский Марк Маркович, к.т.н., доцент

Семенов Виктор Павлович, д.э.н., профессор

Шевская Наталья Владимировна

Санкт-Петербург, 2021

# Цель и задачи

**Актуальность:** данные о товарах и их продажах в современных торговых системах имеют сложную структуру, которая без предварительной группировки сложно поддается анализу

## Цель:

разработать алгоритм предварительной группировки данных розничной торговли

## Задачи:

1. обзор предметной области и сравнительный анализ существующих алгоритмов группировки данных;
2. реализация алгоритма предварительной группировки данных;
3. оценка качества разработанного алгоритма.

# Обзор предметной области

## Группировка

представляет собой метод разбиения исследуемой совокупности данных на однородные по изучаемым признакам группы

### Группировка товаров для формирования ассортимента вендинговых автоматов

- осуществляется вручную и занимает слишком много времени;
- имеет низкую точность разбиения в связи с тем, что в них используется один параметр данных.

### С точки зрения маркетинговых задач применяется

- для выявления товаров со схожей структурой спроса;
- для разбиения потребителей на близкие по особенности поведения группы;
- для анализа спроса в зависимости от набора входных показателей.

# Определение критериев для сравнительного анализа

## Регулирование количества получаемых групп

разбиение набора данных на фиксированное число групп позволит задавать ассортимент товаров для конкретного места, на котором находятся вендинговые автоматы

## Возможность отнести элемент к более чем одной группе

необходима, так как одни и те же товары могут продаваться в нескольких местах одновременно

## Возможность группировки по нескольким критериям

алгоритмы, которые поддерживают многокритериальность, способны учесть несколько факторов одновременно

## Оптимальное значение вычислительной сложности

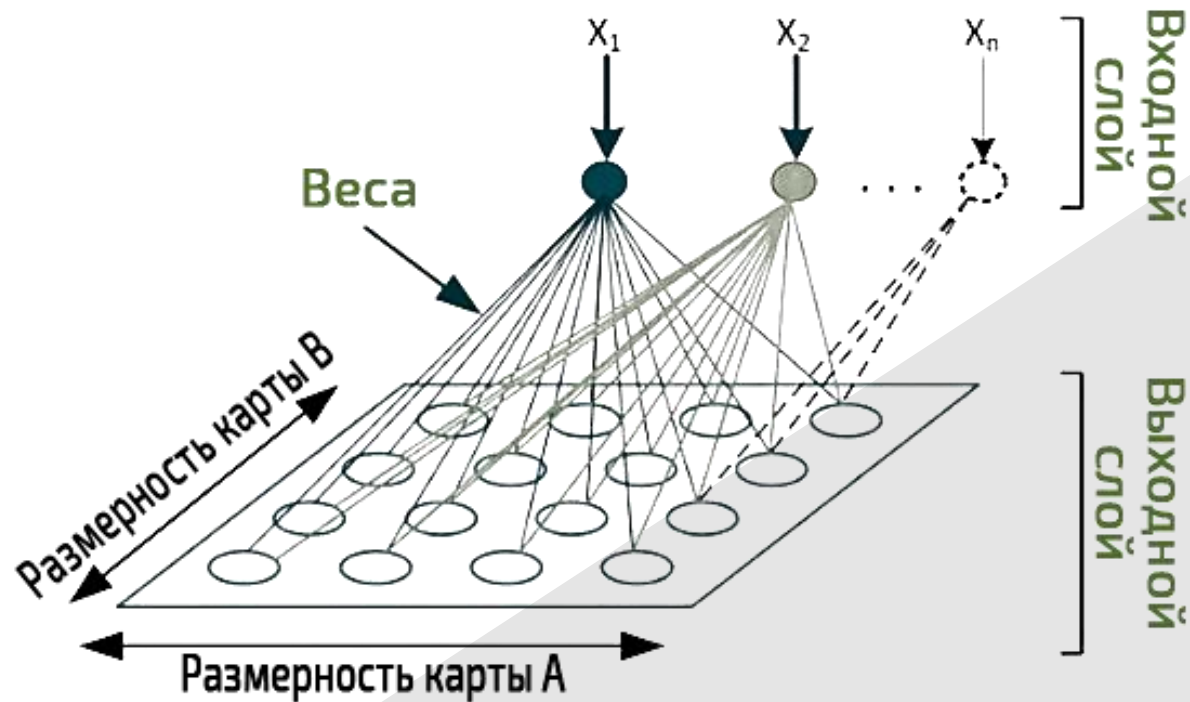
нужно, чтобы при увеличении размера входных данных алгоритм не расходовал дополнительные вычислительные ресурсы и время для анализа

# Сравнительный анализ алгоритмов по выбранным критериям

Название алгоритма группировки данных	Возможность регулирования количества групп	Возможность отнести элемент к более чем одной группе	Возможность группировки по нескольким критериям	Значение вычислительной сложности
Алгоритм <i>k</i> -средних	+	-	-	$O(n \cdot k \cdot l)$ , где $n$ – число объектов в наборе, $k$ – число групп, $l$ – число итераций
Алгоритм <i>c</i> -средних	+	+	-	
ЕМ-алгоритм	+	-	-	
Алгоритмы иерархической кластеризации	-	-	-	$O(n^2)$ , где $n$ – число объектов в наборе
Самоорганизующаяся карта Кохонена	-	+	+	$O(N \cdot M \cdot n)$ , где $N$ – число нейронов выходного слоя, $M$ – число векторов, $n$ – размерность

# Группировка на основе карты Кохонена

## Структура самоорганизующейся карты Кохонена



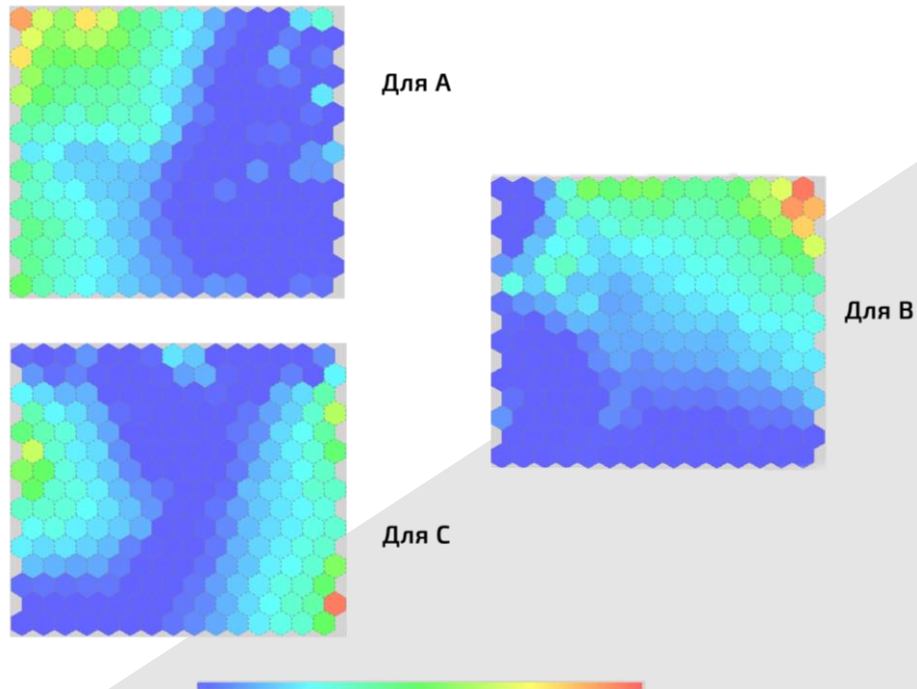
Алгоритм обучения разделяется на следующие шаги:

- инициализация весов каждого узла;
- выбор случайного вектора из набора входных данных;
- процесс конкуренции;
- процесс кооперации;
- процесс адаптации.

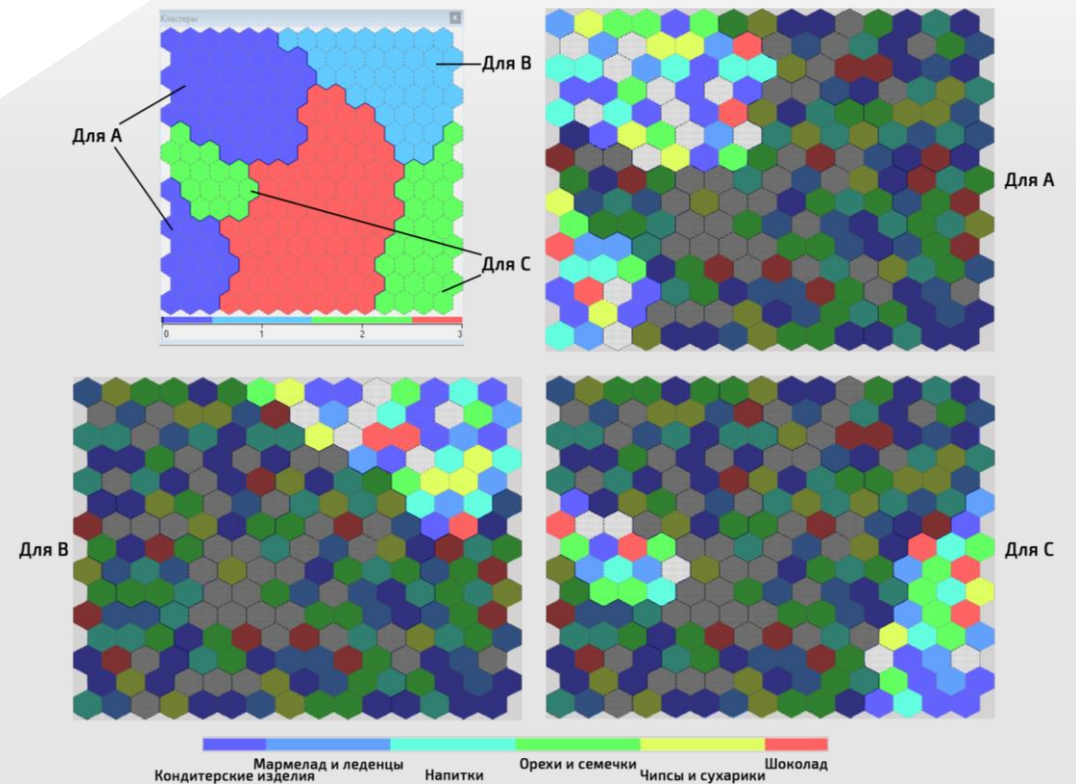
# Реализация алгоритма предварительной группировки данных

Шаг 1. Получение разбиения в Deductor Studio Academic 5.3

## Точки сгущения



## Полученные группы



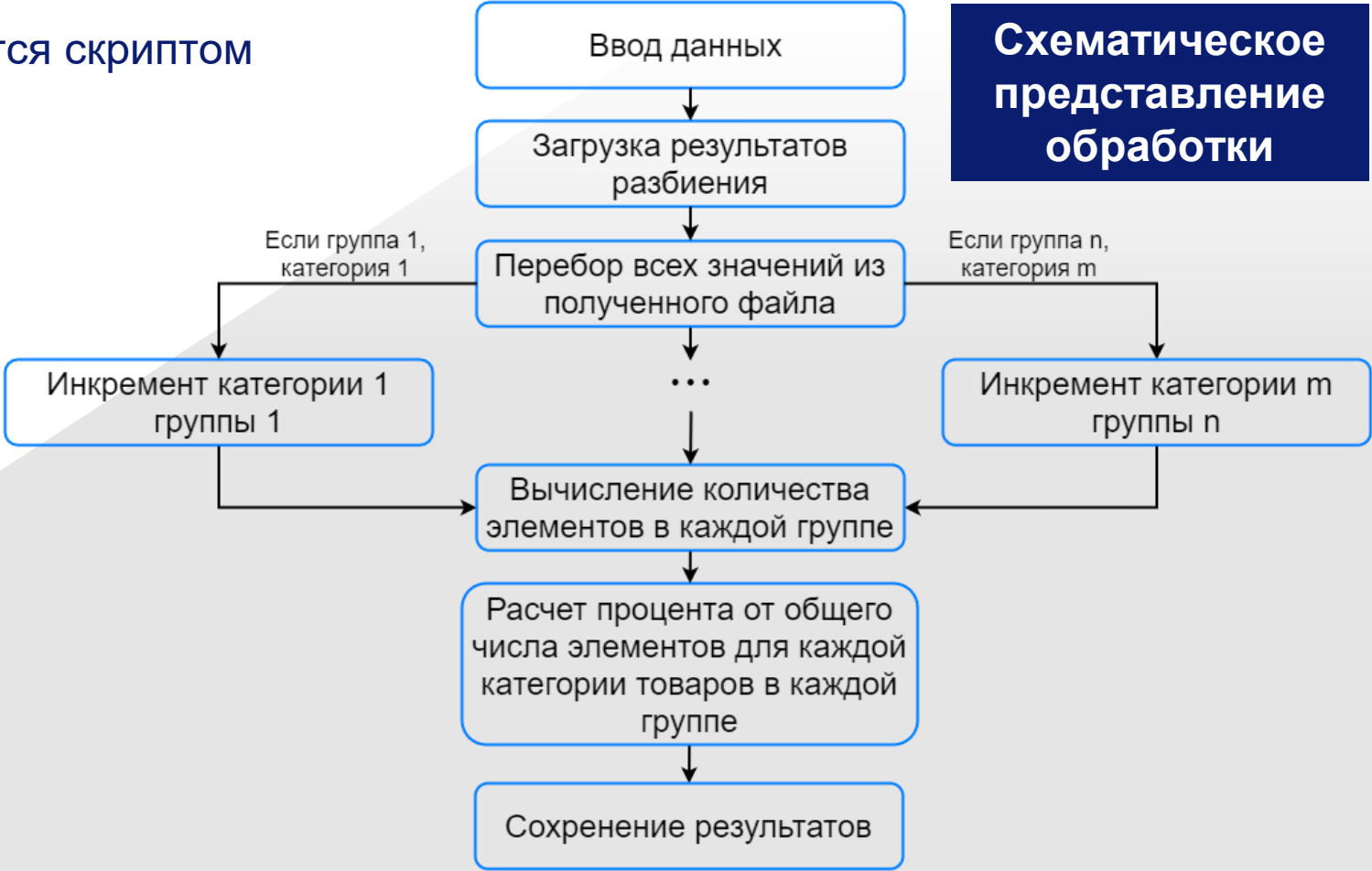
# Реализация алгоритма предварительной группировки данных (2)

Шаг 2. Построенные карты обрабатываются скриптом

Схематическое представление обработки

Экспортируемый файл

Наименование категории товаров	Номер группы
--------------------------------	--------------





# Результат алгоритма предварительной группировки данных

## Промежуточные вычисления

	Автомат А	Автомат В	Автомат С
Кондитерские изделия	22	11	16
Мармелад и леденцы	17	14	15
Чипсы и сухарики	25	11	11
Орехи и семечки	16	7	22
Напитки	20	9	18
Шоколад	24	8	17
Итого	124	60	99

## Результаты расчетов

Автомат А	Автомат В	Автомат С
17.74%	18.33%	16.16%
13.71%	23.33%	15.15%
20.16%	18.33%	11.11%
12.90%	11.67%	22.22%
16.13%	15.00%	18.18%
19.35%	13.33%	17.17%

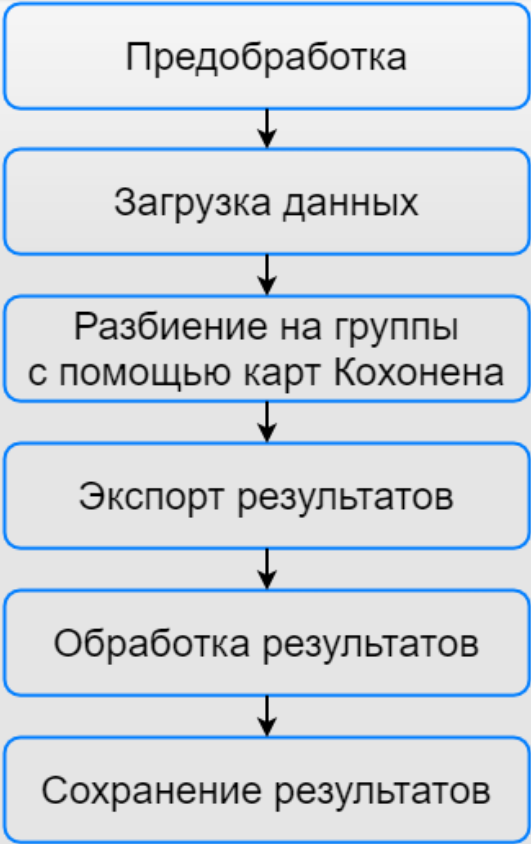
# Оценка качества разработанного алгоритма

Для определения корректности разбиения исходного множества на группы был использован ABC-анализ

## Результат ABC-анализа

	Чипсы и сухарики	Шоколад	Кондитерские изделия	Напитки	Мармелад и леденцы	Орехи и семечки
Автомат А	A	A	A	A	B	C
Автомат В	A	B	A	A	A	C
Автомат С	C	A	A	A	B	A

## Схематическое представление работы алгоритма



# Заключение

- Проведено сравнение существующих алгоритмов по назначенным критериям и выбран наиболее подходящий – группировка с помощью самоорганизующихся карт Кохонена.
- Разработан алгоритм предварительной группировки данных с использованием карт Кохонена.
- Для оценки точности работы алгоритма был проведен ABC-анализ, результаты которого подтвердили правильность работы разработанного алгоритма. Время работы алгоритма на тестовом наборе данных заняло 6-10 сек.

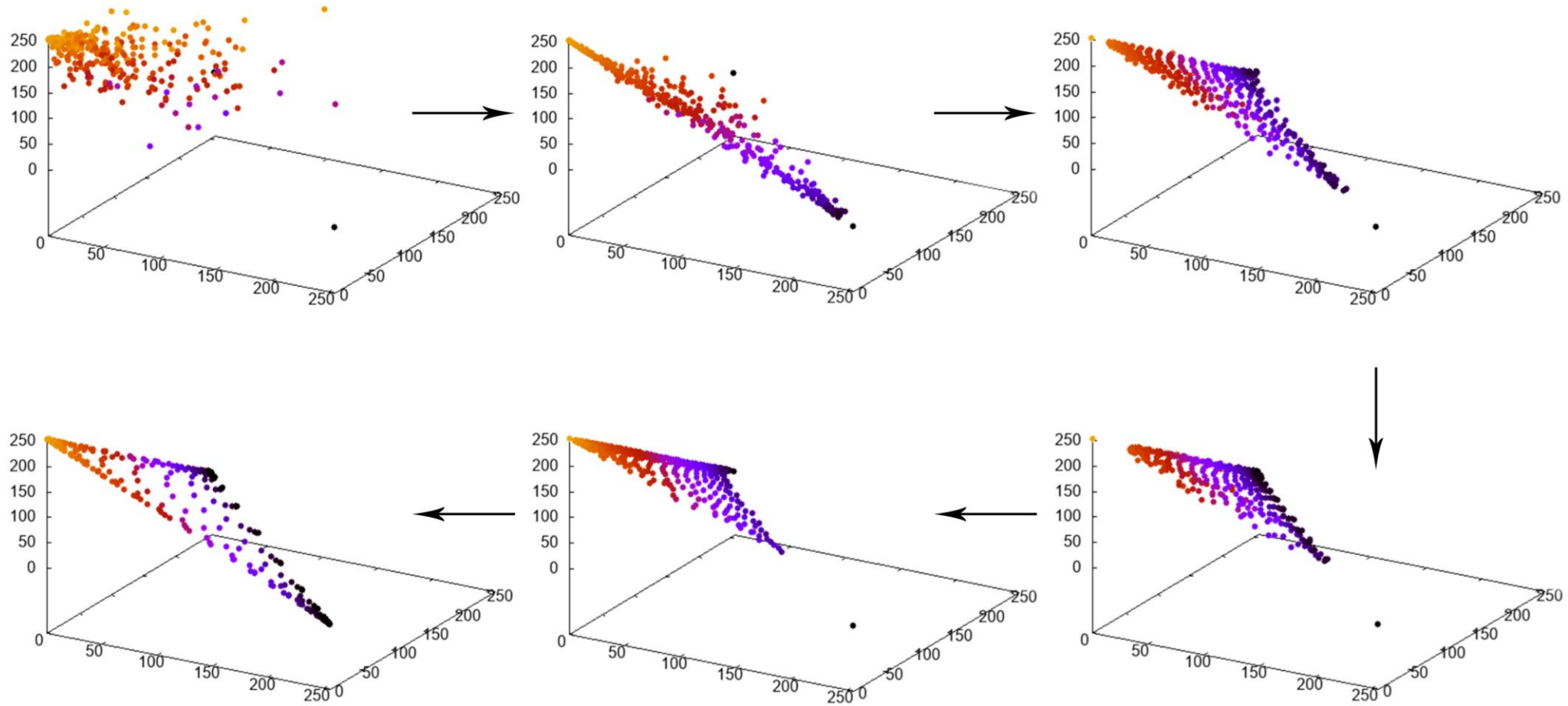
Дальнейшие направления исследований включают в себя использование разработанного алгоритма группировки данных розничной торговли в рекомендательных системах, а также создание графического интерфейса пользователя.

## Апробация работы

- Deryabina, P. S., Iolshina, V. M., Markova, A. V., Shevskaya, N. V., & Belov, V. (2021, January). Challenges and Perspectives of Recommender Systems in Vending Machines // 2021 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (ElConRus). – IEEE, 2021. – С. 284-289.
- Репозиторий проекта <https://github.com/AngelinaMarkova99/Processing-the-results-of-splitting-into-groups>.

# Запасные слайды

# Пример обучения карты Кохонена



# Исходный набор данных



Кондитерские изделия



Напитки



Шоколад



Чипсы и сухарики



Орехи и семечки



Мармелад и леденцы

A	B	C	Date	Product
10	6	2	01.01.2019	Кондитерские изделия
9	0	0	01.01.2019	Орехи и семечки
1	9	6	01.01.2019	Напитки
7	0	7	01.01.2019	Мармелад и леденцы
2	0	1	01.01.2019	Чипсы и сухарики
1	6	8	01.01.2019	Шоколад

...