

# Индустриальный семинар

## Рекомендательные системы

Элен Теванян

Руководитель направления алгоритмического анализа, X5 Tech

x5retail.tech

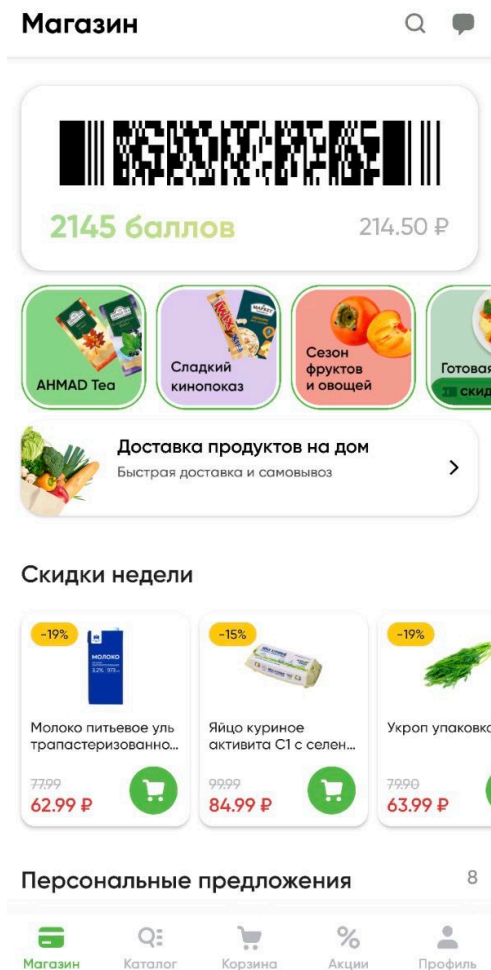


# Где рекомендовать?

- Мобильное приложение/Доставка

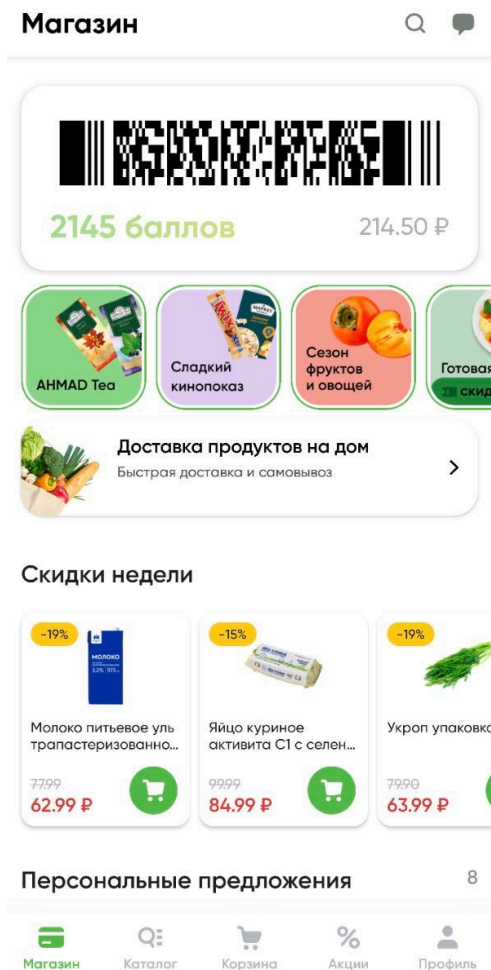
# Где рекомендовать?

## Главная страница: Скидки



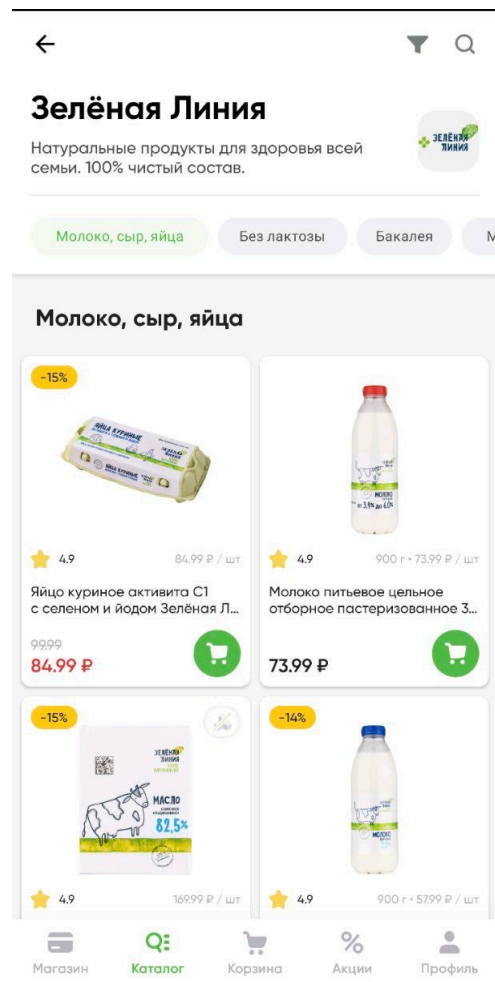
# Где рекомендовать?

## Главная страница: Ранжирование



# Где рекомендовать?

## Ранжирование в каталоге



# Где рекомендовать?

## Комплементарные товары

Яйцо Зелёная Линия куриное активита С1 с селеном и йодом, 10шт – натуральный продукт, полученный от птиц свободного содержания. Благодаря этому в нем содержится больше витаминов и микроэлементов. Яйца достаточно крупные, с плотным желтком ярко-желтого цвета. Белок хорошо держит структуру в отварном виде. Продукт подходит для приготовления салатов, блинов и оладий, сэндвичей с яйцом-пашот и авокадо. Им можно смазывать выпечку, чтобы она приобрела аппетитную румяную корочку.

### Состав

Яйцо куриное пищевое столовое.







### Характеристики товара

Кол-во в упаковке 10 шт

[Все характеристики](#)



### С этим товаром покупают

 <p>Молоко Простоквашино пас...</p> <p>75.99 Р</p> 	 <p>Молоко Простоквашино отб...</p> <p>97.99 Р</p> 	 <p>Пирожное Kinde Молочный ломт</p> <p>49.99 Р</p> 
--	--	---

99.99  
**84.99 Р**

-15%


В корзину

# Где рекомендовать?

## Расширение корзины

Корзина

Очистить



Сыр Аллгой Dorblu Classic с голубой плесенью 50%, 100г


299.99 Р

—

1 шт


+


Рекомендуемые товары



Крем-мыло жидкое Dove Красота и ухо...


399.99 Р






Сёмга стейк охлаждённый


1999.00 Р





Сёмга кусок охлаждённый

1979.00 Р



В вашем заказе

2 товара

455.93 Р

Доставка

99.00 Р

Пакет

7.90 Р

К оплате: 562.83 Р  
~ 60 мин

Оформить заказ

Магазин

Каталог

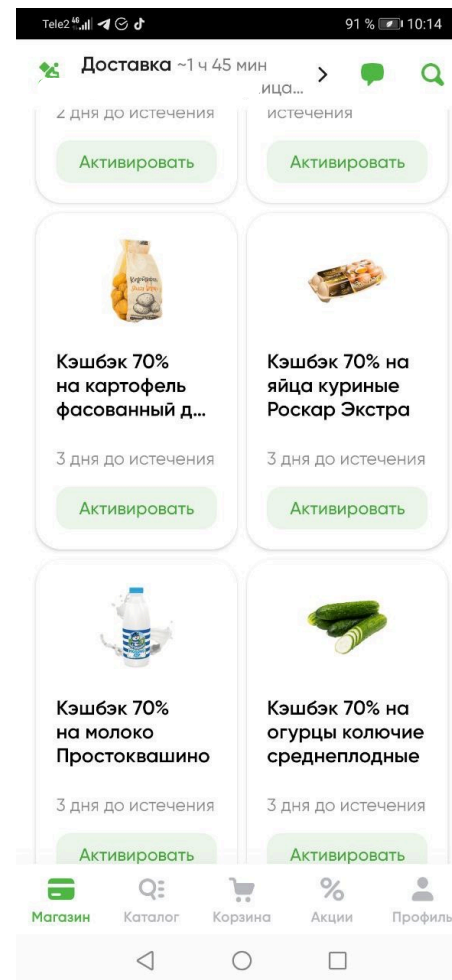
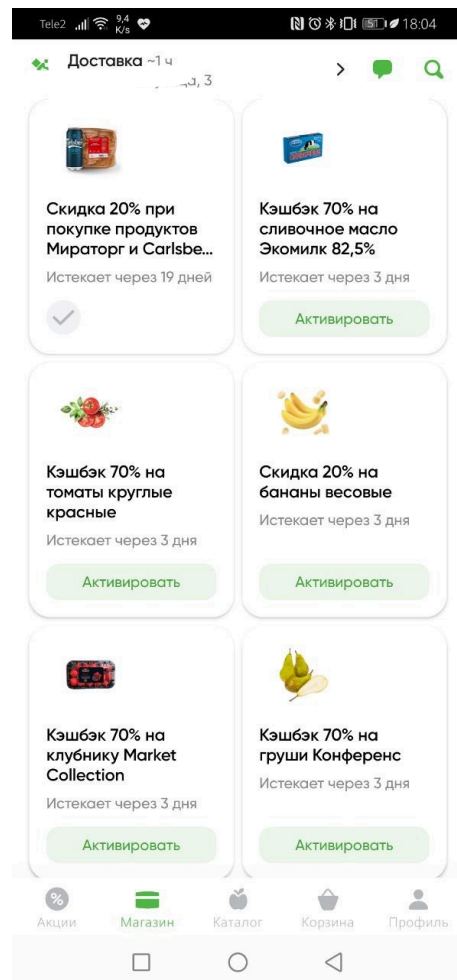
Корзина

Акции

Профиль

# Где рекомендовать?

## Персонализированные скидки



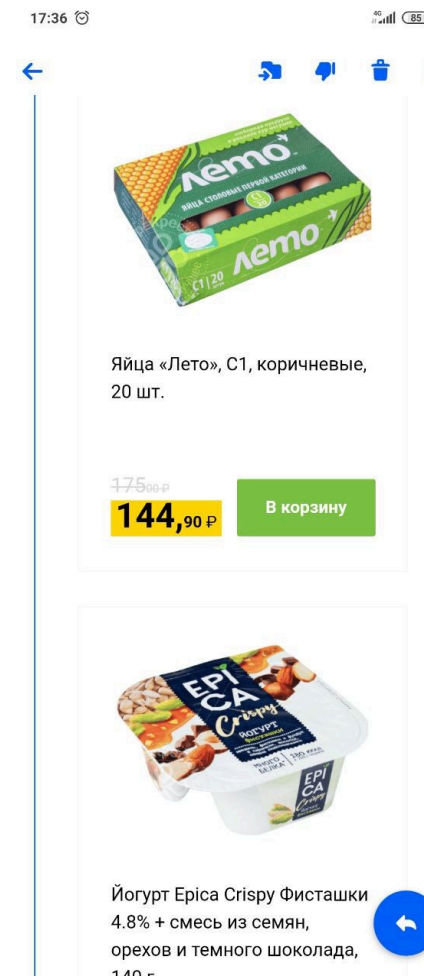
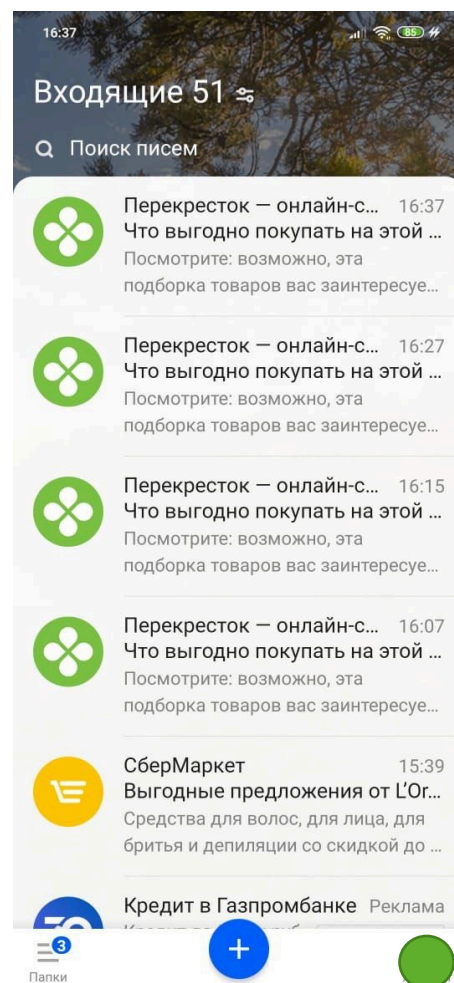


# Где рекомендовать?

- Мобильное приложение/Доставка
- email'ы

# Где рекомендовать?

## Email'ы

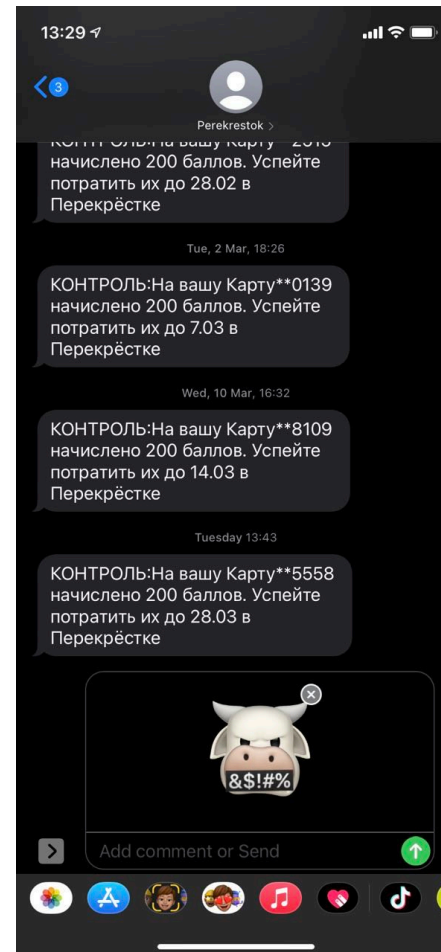


# Где рекомендовать?

- Мобильное приложение/Доставка
- email'ы
- Sms

# Где рекомендовать?

## SMS

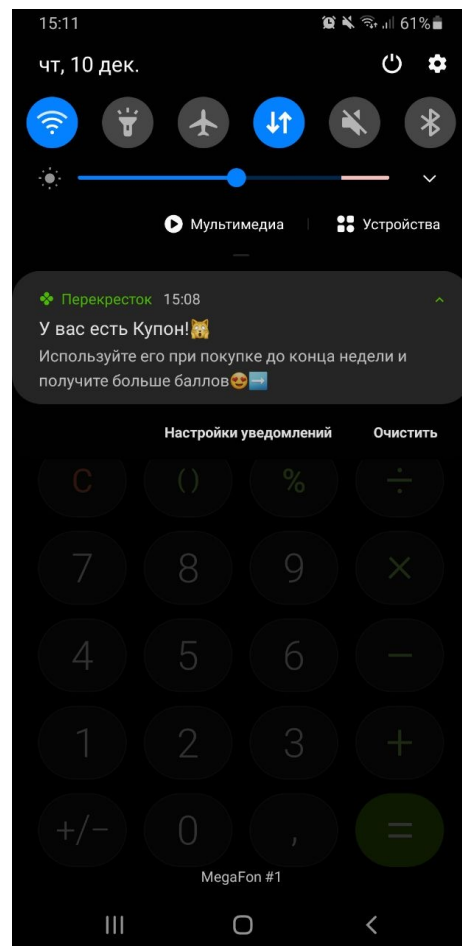


# Где рекомендовать?

- Мобильное приложение/Доставка
- email'ы
- Sms
- Push

# Где рекомендовать?

## PUSH



# Где рекомендовать?

- Мобильное приложение/Доставка
- email'ы
- Sms
- Push

# На чем обучать?

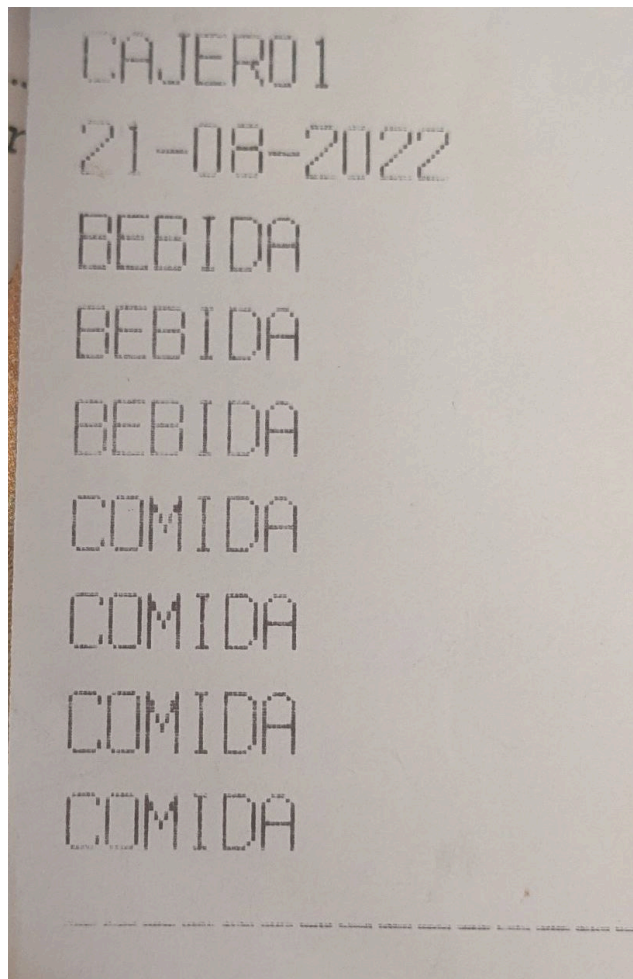


# На чем обучать?

## На чековых данных

Короткий словарь:

- Bebida – напиток
- Comida – еда

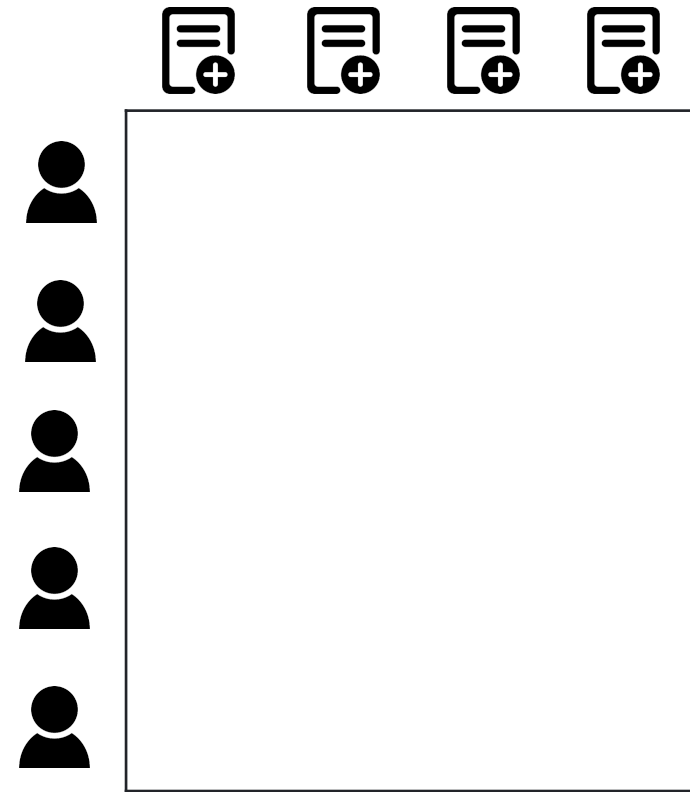


# Какие проблемы с данными?

# Какие проблемы с данными?

## Размер матрицы взаимодействий

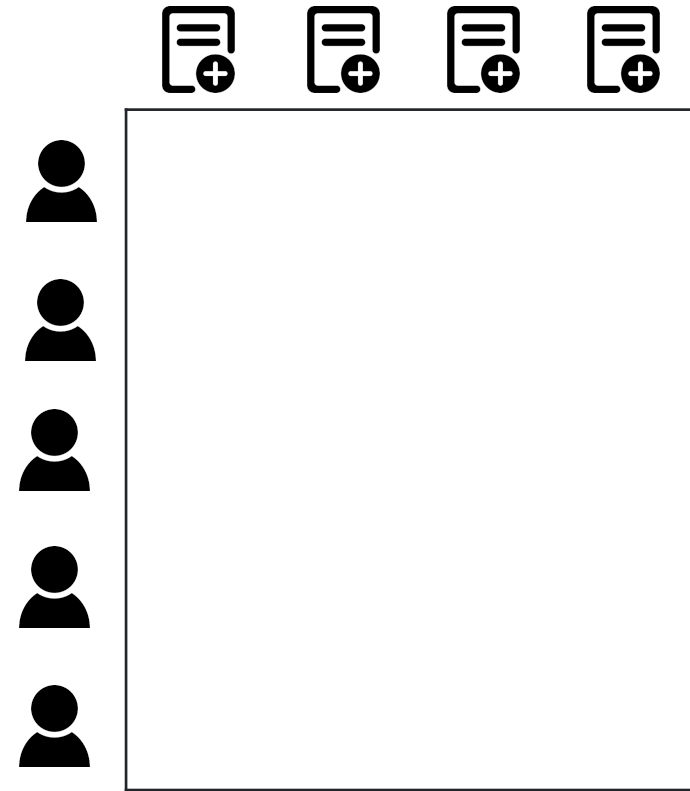
- 5 млн клиентов
- 20 тысяч товарных наименований



# Какие проблемы с данными?

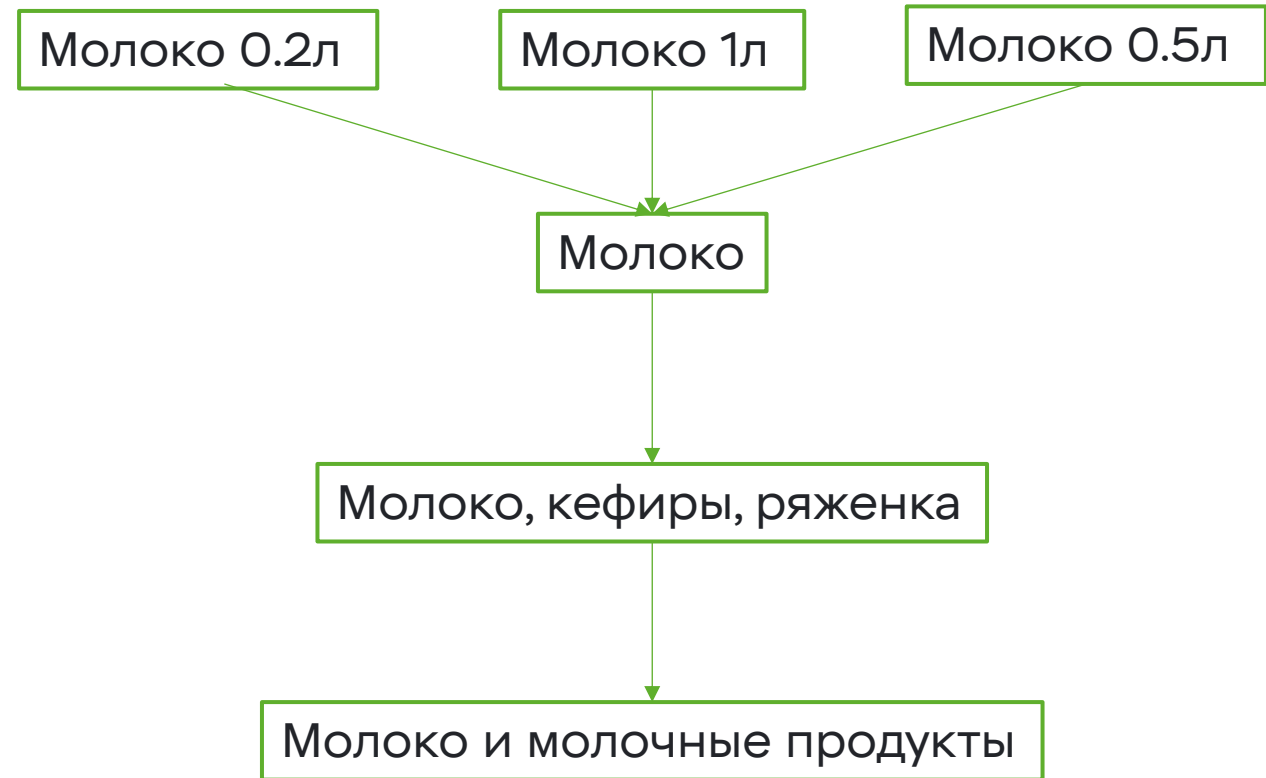
## Разреженность матрицы

- 5 млн клиентов
- 20 тысяч товарных наименований
- Гипотеза: человек покупает не более 2 тысяч товарных наименований



# Решение проблемы разреженности

- Вспомнить о товарной иерархии



# Решение проблемы разреженности

- Вспомнить о товарной иерархии
- Перейти к двухуровневым моделям

Рексистема на категорию



Uplift на бренд

# Что используется?

# Что используется?

## ARL – Associations Rules Learning

- В 1992 году группа по консалтингу в области ритейла компании Teradata под руководством Томаса Блишока провела исследование 1.2 миллиона транзакций в 25 магазинах для ритейлера Osco Drug (нет, там продавали не наркотики и даже не лекарства, точнее, не только лекарства. Drug Store – формат разнокалиберных магазинов у дома). После анализа всех этих транзакций самым сильным правилом получилось «Между 17:00 и 19:00 чаще всего пиво и подгузники покупают вместе»





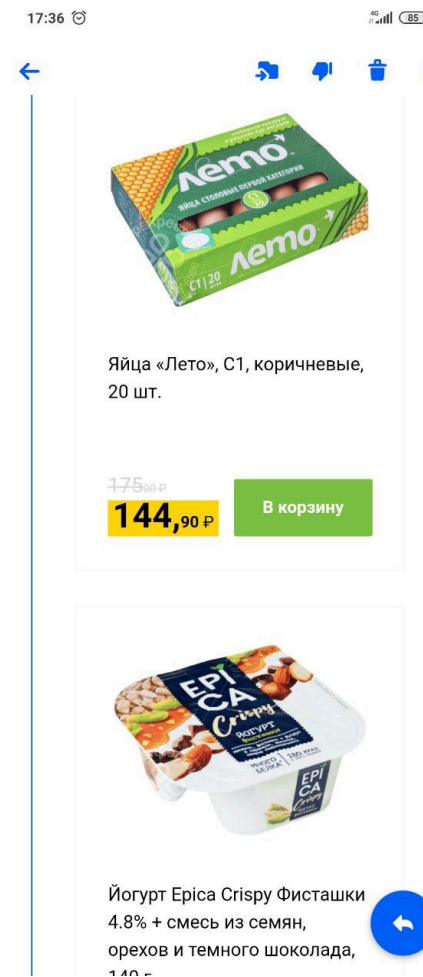
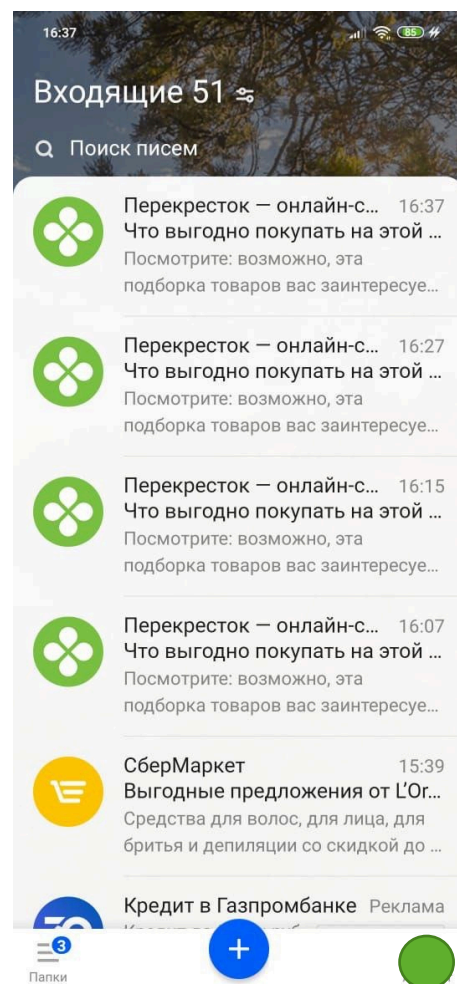
# Что используется?

## ARL – Associations Rules Learning

название	формула	Что значит	пример	трактовка
Поддержка $support(x)$	$\frac{\{t \in T; x \in t\}}{ T }$	частота встречаемости айтемсета в транзакциях	$\frac{\text{Транзакции с пивом и подгузниками}}{ \text{Все транзакции} }$	Доля транзакций с подгузником и пивом ото всех
Уверенность $confidence(x_1 \cup x_2)$	$\frac{supp(x_1 \cup x_2)}{supp(x_1)}$	как часто срабатывает правило в датасете	$\frac{supp(\text{Пиво} \cup \text{Подг})}{supp(\text{Пиво})}$	кто покупает пиво, тот покупает и подгузники
Лифт $lift(x_1 \cup x_2)$	$\frac{supp(x_1 \cup x_2)}{supp(x_1) * supp(x_2)}$	зависят ли айтемы друг от друга	$lift = 1.25$	правило, что пиво покупают с подгузниками, на 25% мощнее правила, что подгузники просто покупают
Убедительность $conviction(x_1 \cup x_2)$	$\frac{1 - supp(x_1)}{1 - conf(x_1 \cup x_2)}$	частотность ошибок, как часто один из айтемов покупали без другого	$conviction = 1.6$	правило «купил пиво и подгузники» было бы в 1.2 раза (на 60%) более верным, чем если бы совпадение этих items в одной транзакции было бы чисто случайным

# Куда отправить результаты?

## Email'ы



Что используется?

Что используется?

## Коллаборативная фильтрация

		Товары					
Пользователи		1	1	0		1	
		0	1	1			1
					1	1	0
			1	1		0	
			1				1

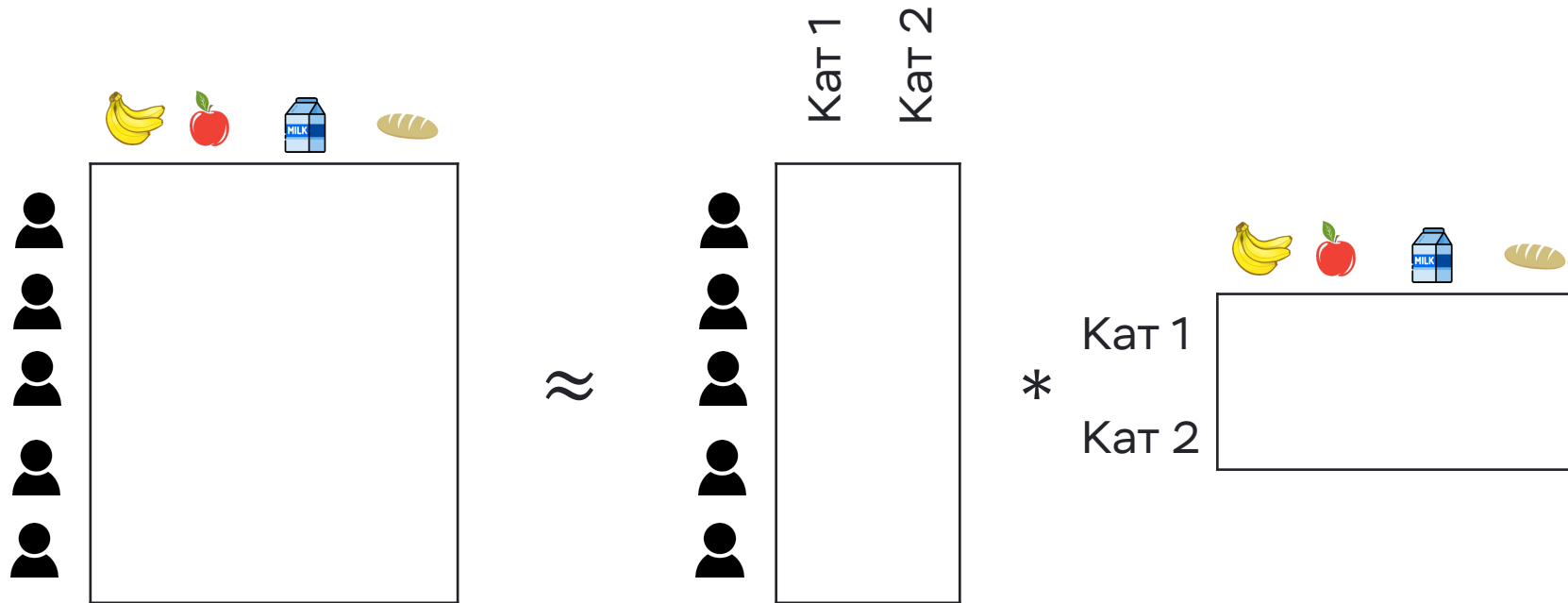
Что используется?

# Коллаборативная фильтрация

- User-User
- Item-Item

Что используется?

## КФ. Модели со скрытыми переменными



# Что используется?

## ALS

- $x_u$  — вектор пользователя  
 $y_i$  — вектор айтема
- $p_{ui}$  — предпочтение
- $c_{ui}$  — уверенность
- Оптимизационная задачаё

$$p_{ui} = \begin{cases} 1 & r_{ui} > 0 \\ 0 & r_{ui} = 0 \end{cases}$$

$$c_{ui} = 1 + \alpha r_{ui}$$

$$\min_{y_*, y_*} \sum_{u,i} c_{ui} (p_{ui} - x_u^T y_i)^2 + \lambda \left( \sum_u \|x_u\|^2 + \sum_i \|y_i\|^2 \right)$$

- После некоторых попыток оптимизации:

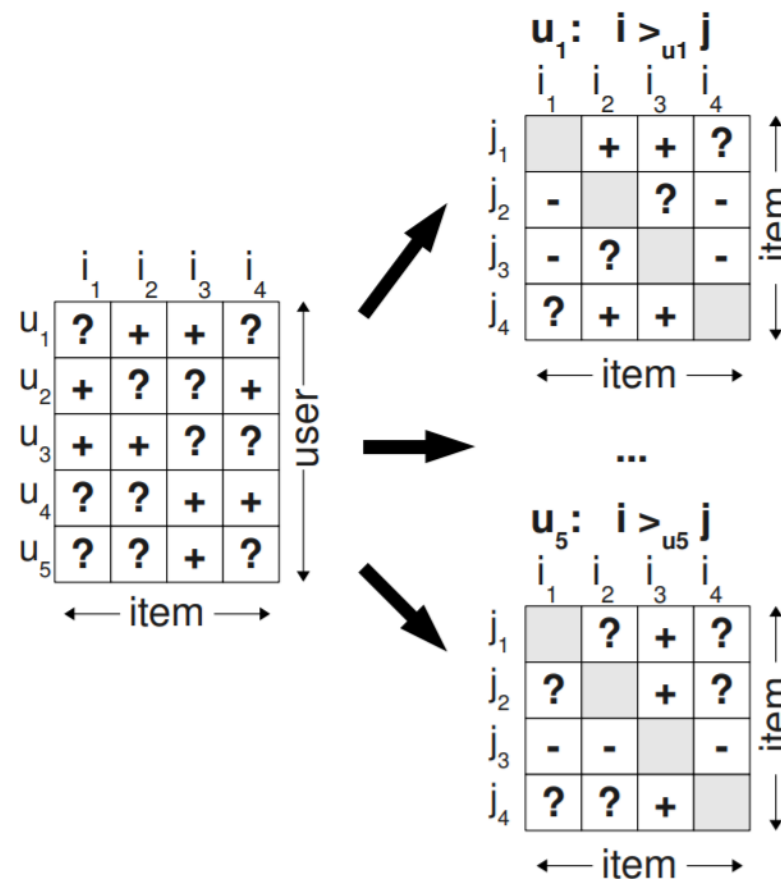
$$x_u = (Y^T Y + Y^T (C^u - I) Y + \lambda I)^{-1} Y^T C^u p(u)$$

$$y_i = (X^T X + X^T (C^i - I) X + \lambda I)^{-1} X^T C^i p(i)$$

# Что используется?

## Bayesian Personalized Ranking

- Имеем дело с тройкой: пользователей  $u$  и два айтема  $i, j$
- Работаем с предпочтением  $i$  над  $j$  пользователем  $u$

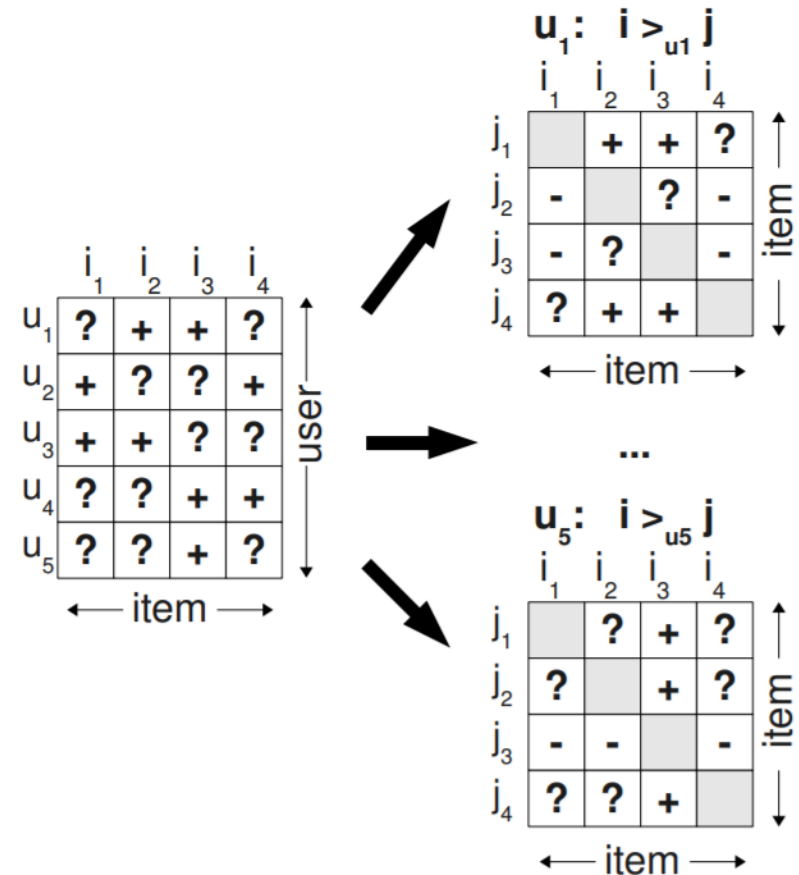




# Что используется?

## Bayesian Personalized Ranking

- Максимизируем:  $p(\Theta | >_u) \propto p(>_u | \Theta) p(\Theta)$



# Что используется?

## Bayesian Personalized Ranking

- Максимизируем:  $p(\Theta | >_u) \propto p(>_u | \Theta) p(\Theta)$

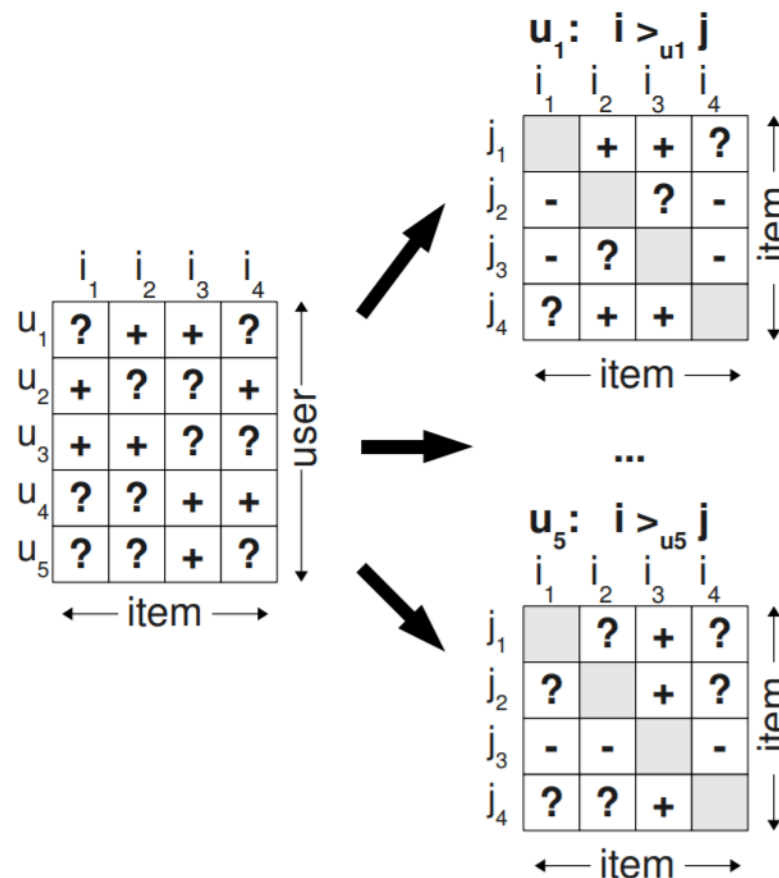
$$p(i >_u j | \Theta) := \sigma(\hat{x}_{uij}(\Theta))$$

where  $\sigma$  is the logistic sigmoid:

$$\sigma(x) := \frac{1}{1 + e^{-x}}$$

- $\hat{x}_{uij}(\Theta)$  получаем из матричных факторизаций
- Которые получаем отсюда:

$$\sum_{(u,i,j) \in D_S} \ln \sigma(\hat{x}_{uij}) - \lambda_{\Theta} \|\Theta\|^2$$



# Стеккинг моделей?

# Стеккинг моделей?

## Имеет место быть

- Стеккаем ALS и BPR

# Как оценивать?

# Как оценивать?

## Метрики стандартные

Позиция	Релевантность
1	1
2	0
3	1
4	1
5	0
6	0
7	1
...	...

# Как оценивать?

## Метрики стандартные

Позиция	Релевантность
1	1
2	0
3	1
4	1
5	0
6	0
7	1
...	...

$$\text{precision@5} = \frac{3}{5}$$

$$\text{recall@5} = \frac{3}{4} \text{ (если всего 4 релевантных айтема)}$$

# Как оценивать?

## P@k, R@k

Позиция	Релевантность
1	1
2	0
3	1
4	1
5	0
6	0
7	1
...	...

$$\text{precision@5} = \frac{3}{5}$$

$$\text{recall@5} = \frac{3}{4} \text{ (если всего 4 релевантных айтема)}$$



# Как оценивать?

## P@k, R@k

Позиция	Релевантность
1	1
2	0
3	1
4	1
5	0
6	0
7	1
...	...

$$\text{precision@5} = \frac{3}{5}$$

$$\text{recall@5} = \frac{3}{4} \text{ (если всего 4 релевантных айтема)}$$

# Как оценивать?

## Другие метрики

- **Покрытие**
  - Какая доля товаров рекомендовалась хотя бы раз?
  - Какой доле пользователей хотя бы раз показаны рекомендации?
- **Новизна**
  - Как много рекомендованных товаров пользователь встречал раньше?
- **Прозорливость (serendipity)**
  - Способность предлагать товары, которые отличаются от купленных ранее

# Эвристики?

## Безусловно да

- Топ-10/20/30/40 товаров
- Новинки
- Схожие темы
- И т.п.

# Ссылки

- ARL-туториал

<https://habr.com/ru/company/ods/blog/353502/>

- ALS-туториал

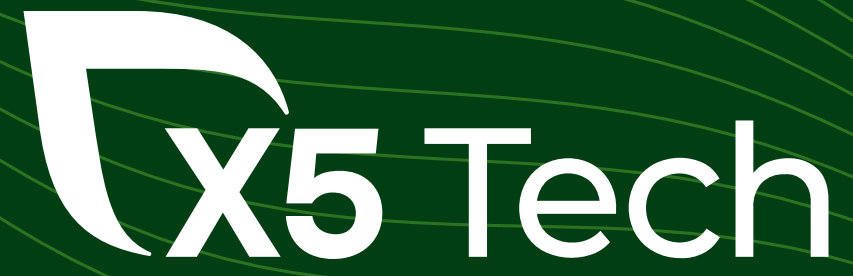
<https://medium.com/radon-dev/als-implicit-collaborative-filtering-5ed653ba39fe>

- ALS и BPR-туториал

<https://towardsdatascience.com/recommender-system-using-bayesian-personalized-ranking-d30e98bba0b9>

- Обзор методов:

<https://habr.com/ru/post/486802/>



20 октября 2022