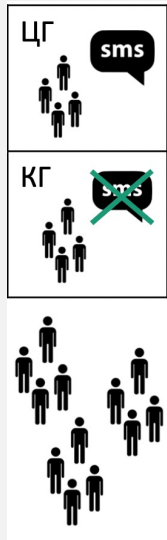
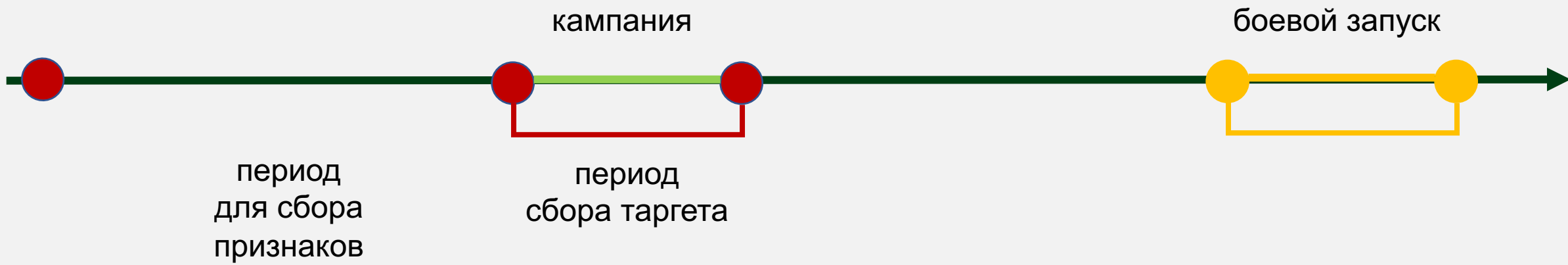


Uplift-конец и RecSys

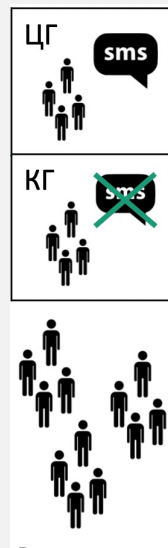
Детали обучения



T

1
0
1
0
0
1
1
0
1

Uplift model



Случайная аудитория

Отобранная моделью аудитория

Методы обучения?

Методы обучения

- Solo Learner, S-Learner

$$\textit{fit}\left(\begin{array}{cc} x_{11} & x_{1m} t_1 y_1 \\ x_{n1} & x_{nm} t_n y_n \end{array}\right)$$

Методы обучения

- Solo Learner, S-Learner

$$\text{fit}\left(\begin{array}{cc} x_{11} & x_{1m} t_1 y_1 \\ x_{n1} & x_{nm} t_n y_n \end{array}\right)$$

$$\text{predict}\left(\begin{array}{cc} x_{11} & x_{1m} 1 \\ & 1 \\ x_{n1} & x_{nm} 1 \end{array}\right) - \text{predict}\left(\begin{array}{cc} x_{11} & x_{1m} 0 \\ & 0 \\ x_{n1} & x_{nm} 0 \end{array}\right) = \text{uplift}$$

Методы обучения

- Two-models approach (T-Learner)

$$\begin{aligned} & model_{treatment} \\ &= fit\left(\begin{array}{cc} x_{11} & x_{1m}y_1 \\ x_{p1} & x_{pm}y_p \end{array} \right) \end{aligned}$$

$$\begin{aligned} & model_{control} \\ &= fit\left(\begin{array}{cc} x_{p1} & x_{pm}y_p \\ x_{n1} & x_{nm}y_n \end{array} \right) \end{aligned}$$

$$predict_{treatment} \begin{pmatrix} x_{11} & x_{nm} \\ x_{n1} & x_{nm} \end{pmatrix} - predict_{control} \begin{pmatrix} x_{11} & x_{nm} \\ x_{n1} & x_{nm} \end{pmatrix}$$

Методы обучения

- Transformed outcome
- Целевая переменная подвергается подобной трансформации

$$Y^{new} = Y_i \frac{(T_i - p)}{p(1 - p)}$$

- P – propensity score, вероятность отнесения к контрольной или целевой группе
- Обучаемся как на обычную регрессию
- [Тут](#) показывают, что матожидание трансформированного таргета и есть искомый causal effect, т.е.

$$\tau = uplift = E[Y^{new} | X]$$

Метрики качества

Uplift@k

- Отсекаем из всей тестовой выборки по порогу $k\%$
- Сортируем выборку по величине предсказанного uplift
- Смотрим на разницу в среднем по таргету: `response_rate` между целевой и контрольной группой

$$uplift@k = response\ rate@k_{(treatment)} - response\ rate@k_{(control)}$$

$$response\ rate@k = mean(Y@k)$$

$Y@k$ — таргет на $k\%$

Метрики качества

Uplift-curve

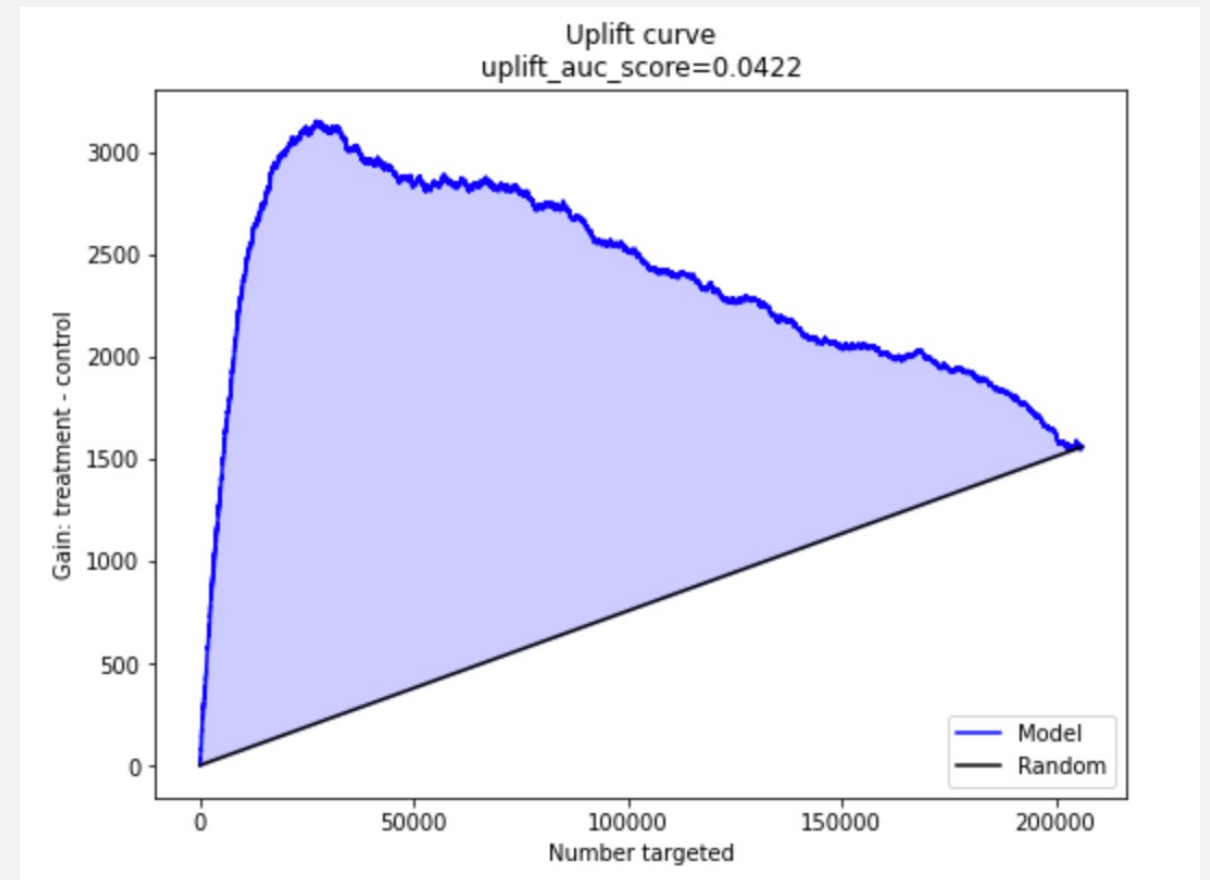
$$uplift\ curve(t) = \left(\frac{Y_t^T}{N_t^T} - \frac{Y_t^C}{N_t^C} \right) (N_t^T + N_t^C)$$

t — накопившееся количество объектов

Y_t^T, Y_t^C — таргет в *treatment* группе, таргет в *control* группе

N_t^T, N_t^C — размер *treatment* группы, размер *control* группы

- формы у кривой мб разные



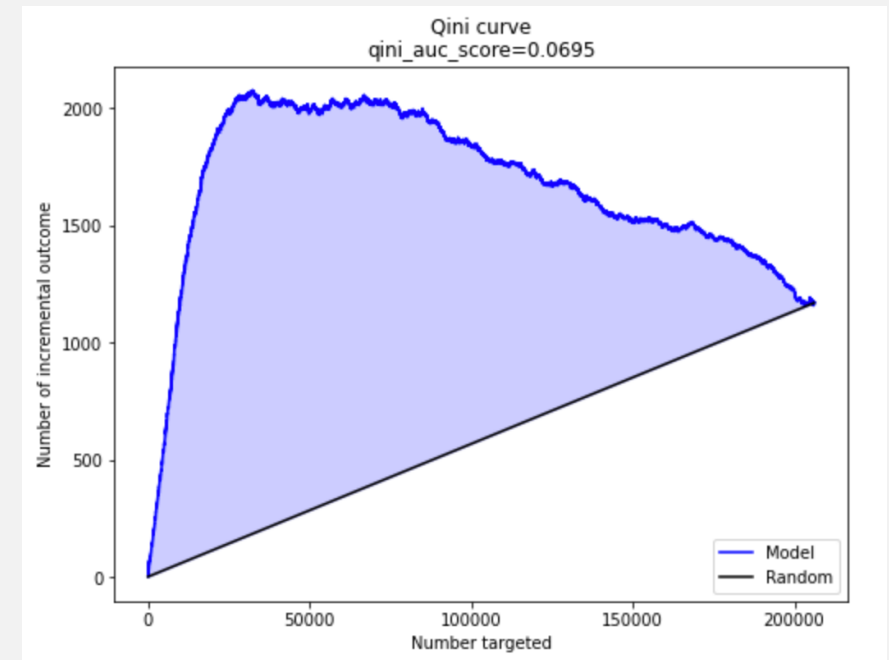
Метрики качества

- Qini curve

$$qini\ curve(t) = Y_t^T - \frac{Y_t^C N_t^T}{N_t^C}$$

Y_t^T, Y_t^C — таргет в *treatment* группе, таргет в *control* группе

N_t^T, N_t^C — размер *treatment* группы, размер *control* группы



Как оценить эффективность?

Короткая инструкция

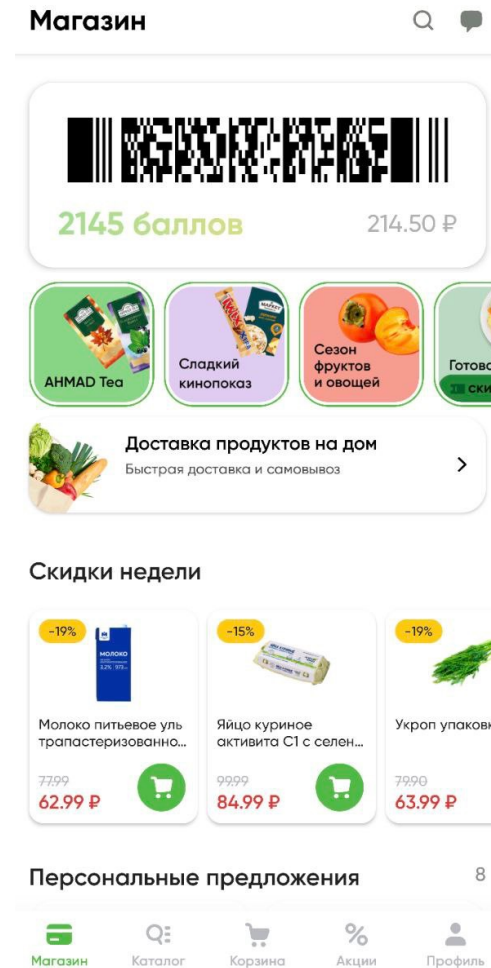
1. Определиться с таргетом (что мерим) и механикой (какой оффер)
2. Собираем обучающую выборку:
 1. Выделяем большой, репрезентативный, лишенный смещений, случайный сегмент базы
 2. Разбиваем его на контроль и тест
 3. Запускаем эксперимент
3. Дождавшись конца эксперимента, собираем таргет
4. Обучаем модель любым методом
5. Построили предсказания
6. Запустить эксперимент:
 1. Берем прогнозы модели, и на них запускаем АБ-тест
 2. Параллельно, мы берем случайную выборку, ее делим контроль-тест, и на ней тоже запускаем эксперимент

Каналы рекомендаций

- Мобильное приложение/Доставка

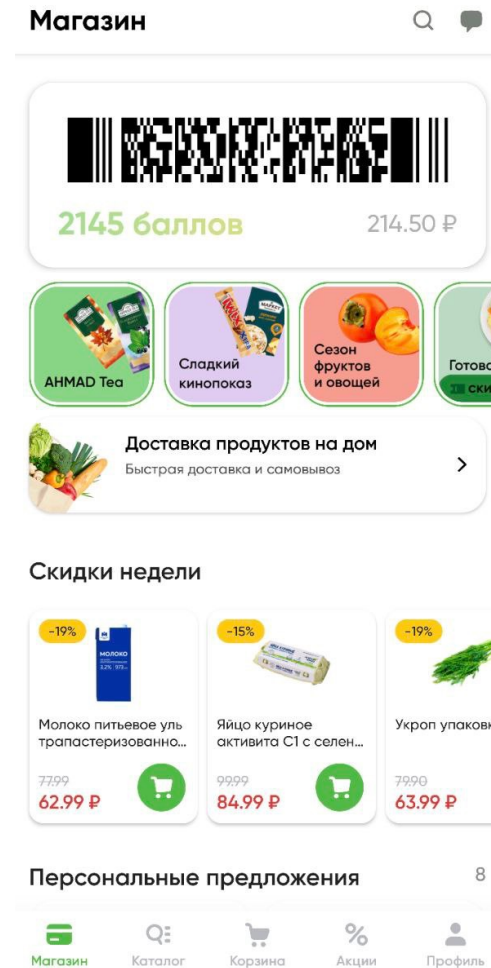
Где рекомендовать?

- Главная страница



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



Молоко Простоквашино пас...


75.99 Р






Молоко Простоквашино отб...


97.99 Р





Пирожное Kinder Молочный ломт...

49.99 Р



99.99 -15%

84.99 Р

[В корзину](#)

Магазин

Каталог

Корзина

Акции


Профиль

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


790 Р


К оплате: 562.83 Р
~ 60 мин


Оформить заказ

Магазин

Каталог

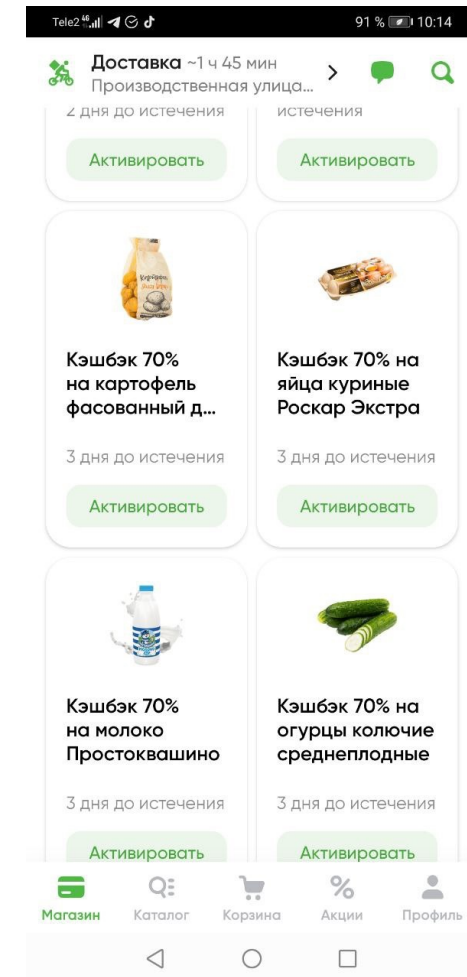
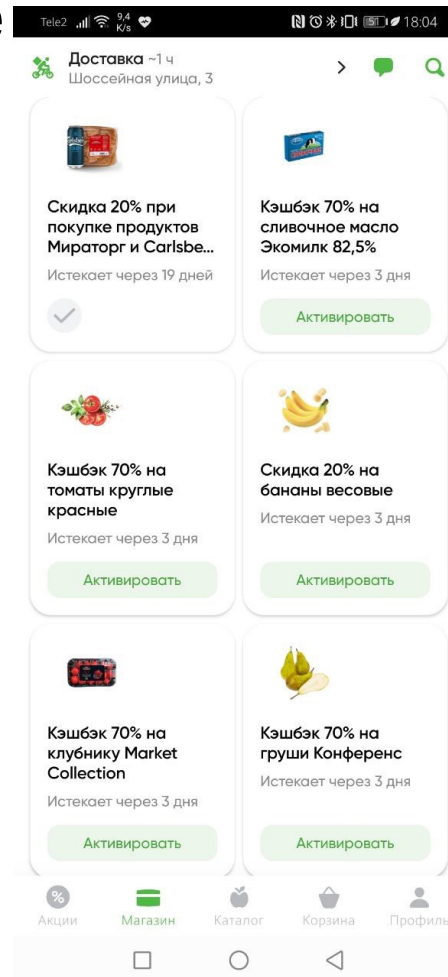
Корзина

Акции

Профиль

Где рекомендовать?

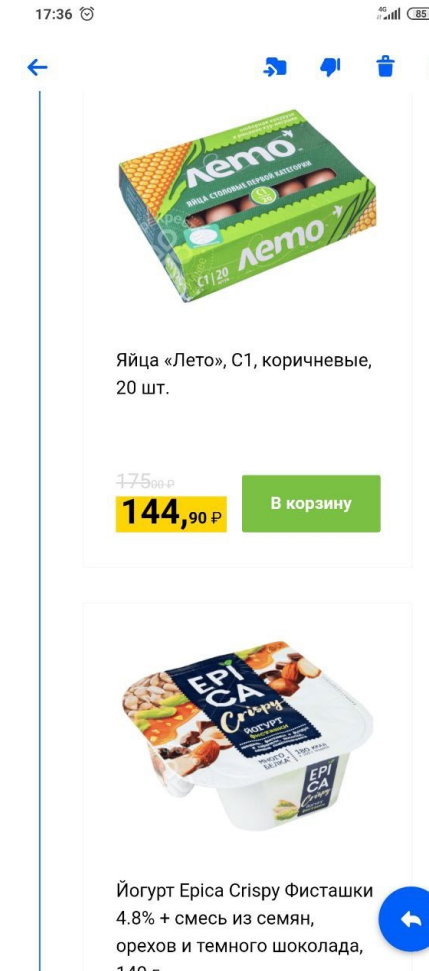
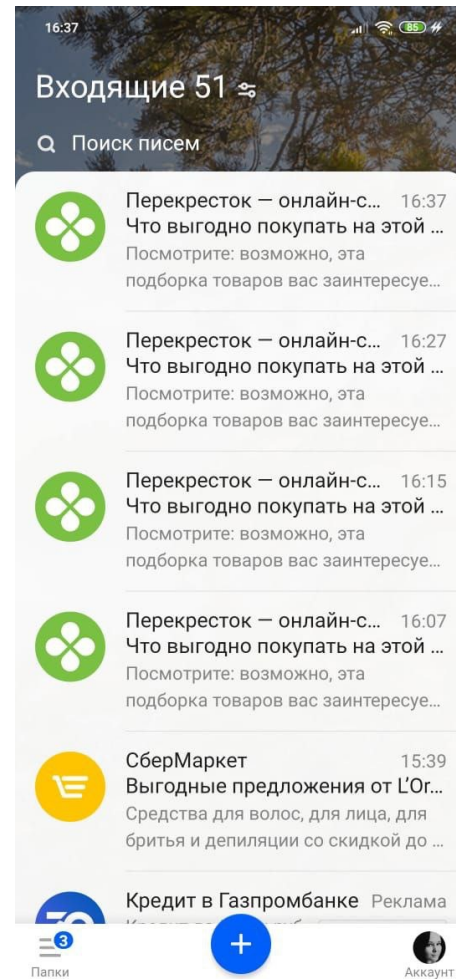
- Персонализированные скидки



Каналы рекомендаций

- Мобильное приложение/Доставка
- Электронная почта

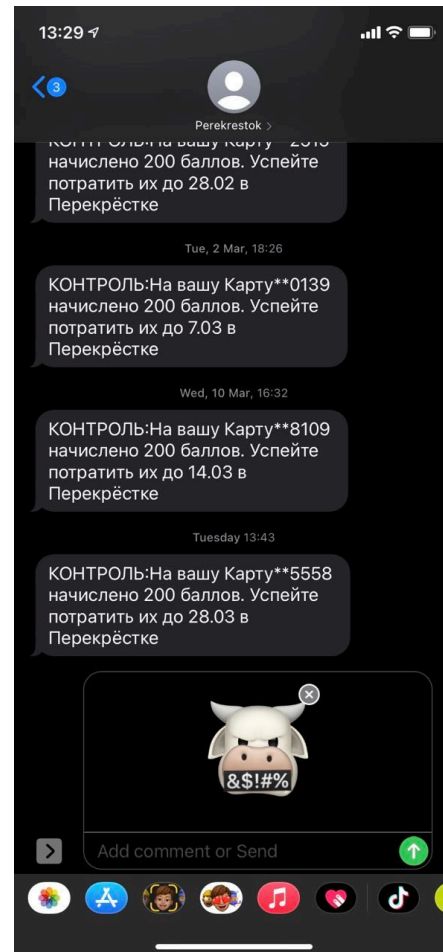
Реко в email



Каналы рекомендаций

- Мобильное приложение/Доставка
- Электронная почта
- SMS

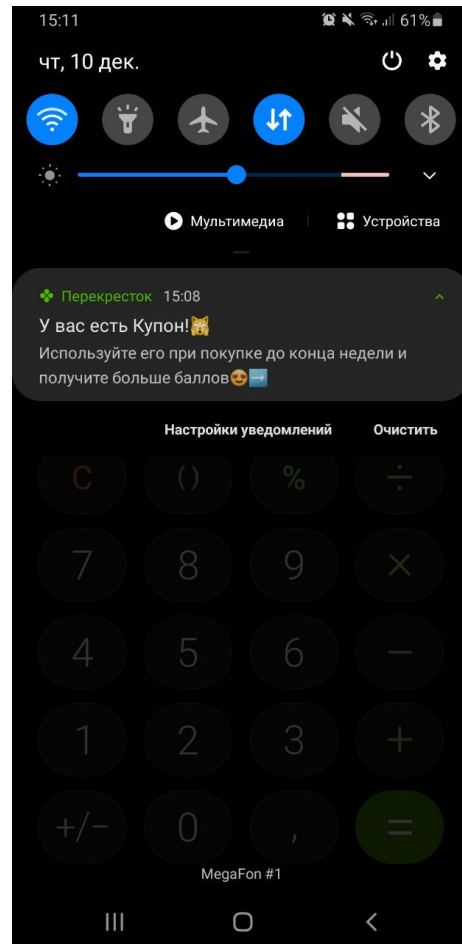
Реко в email



Каналы рекомендаций

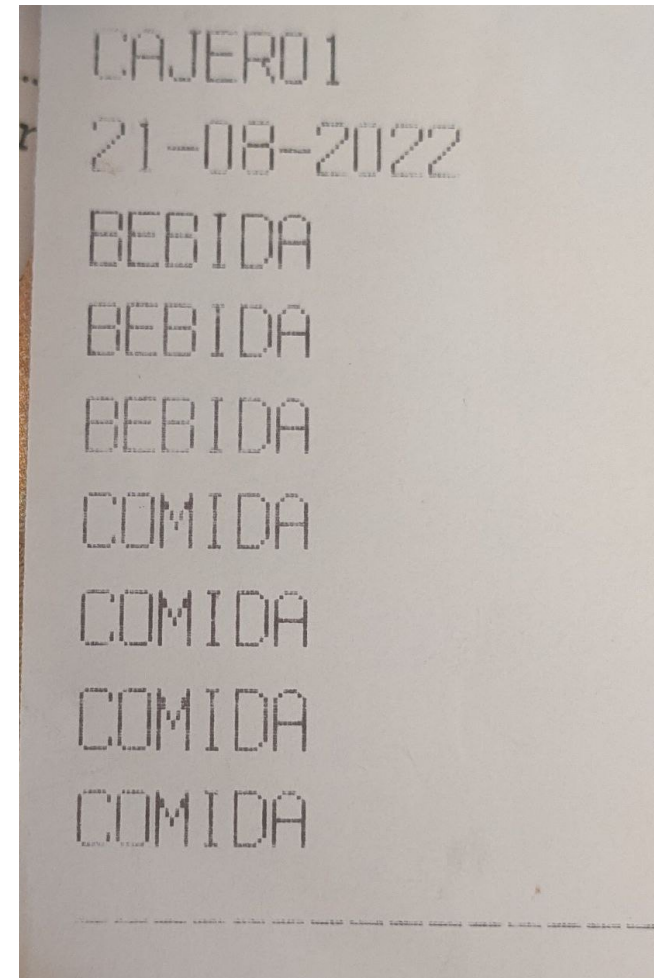
- Мобильное приложение/Доставка
- Электронная почта
- SMS
- push

Реко в email



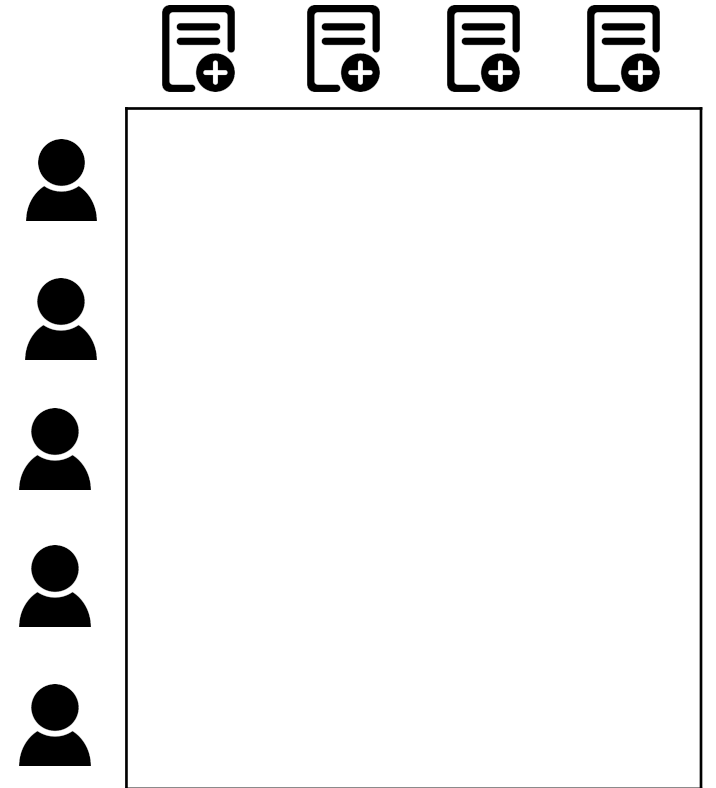
На чем обучать

- На чековых данных
- Короткий словарик:
 - Bebida – напиток
 - Comida – еда



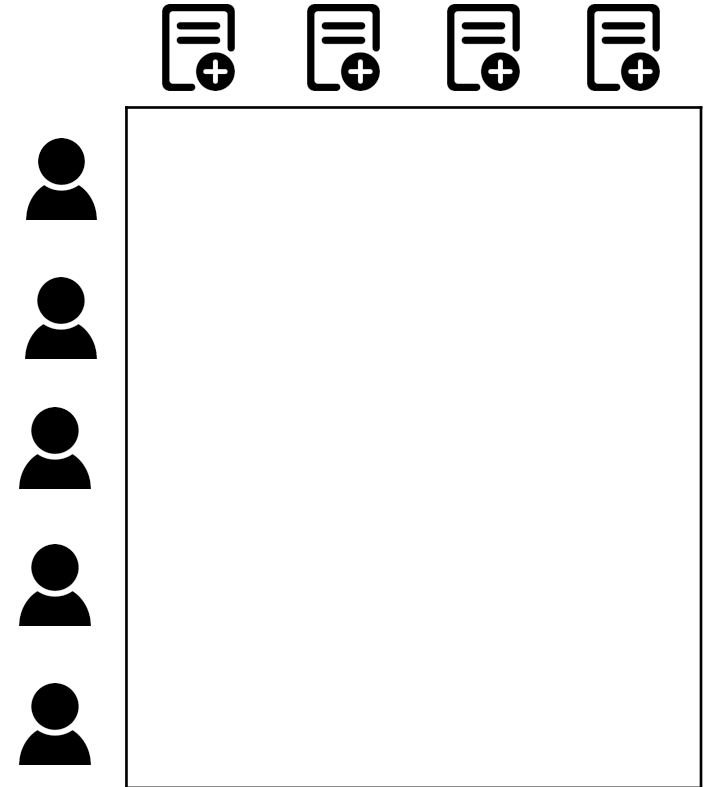
Проблемы с данными?

- 5 млн клиентов
- 20 тысяч товарных наименований



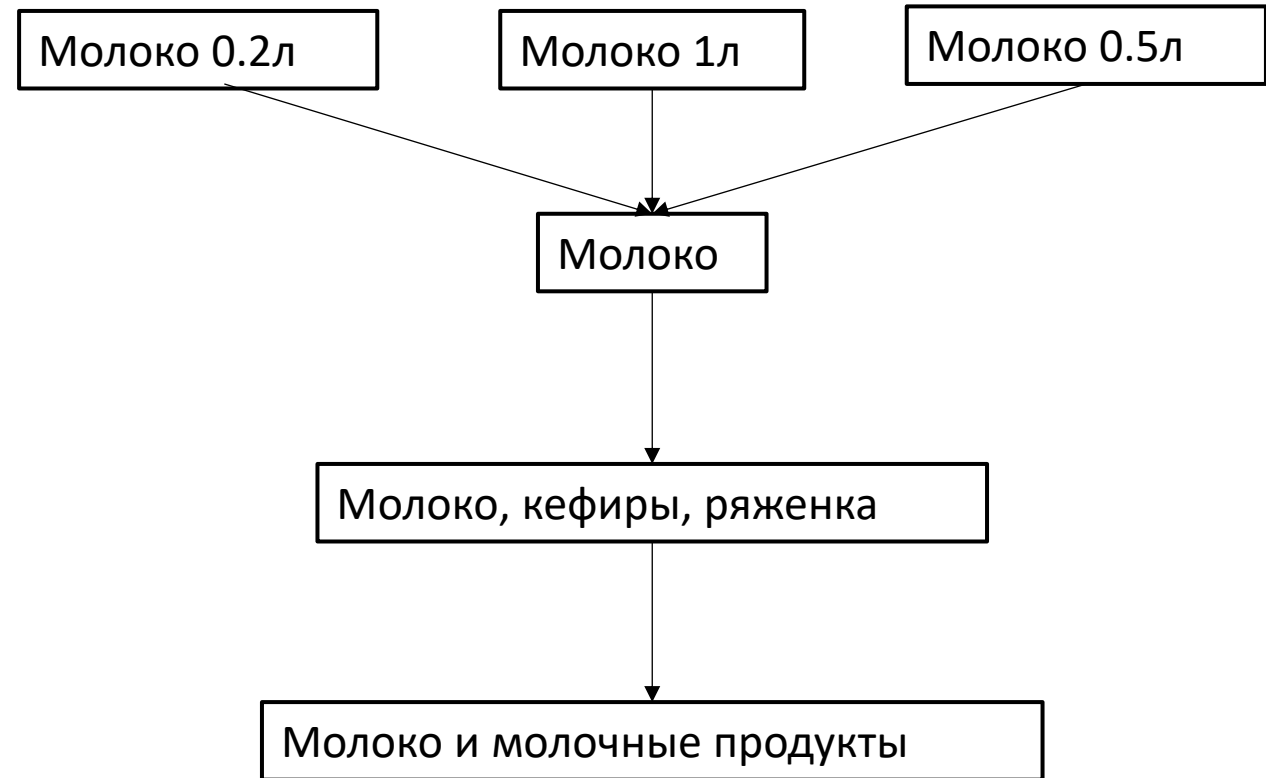
Разреженность данных

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



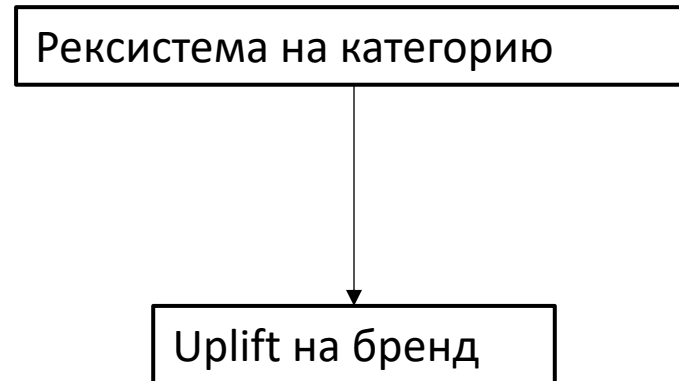
Решение проблемы [1]

- Существует товарная иерархия



Решение проблемы [2]

- Перейти к двухуровневым моделям



Подходы

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»

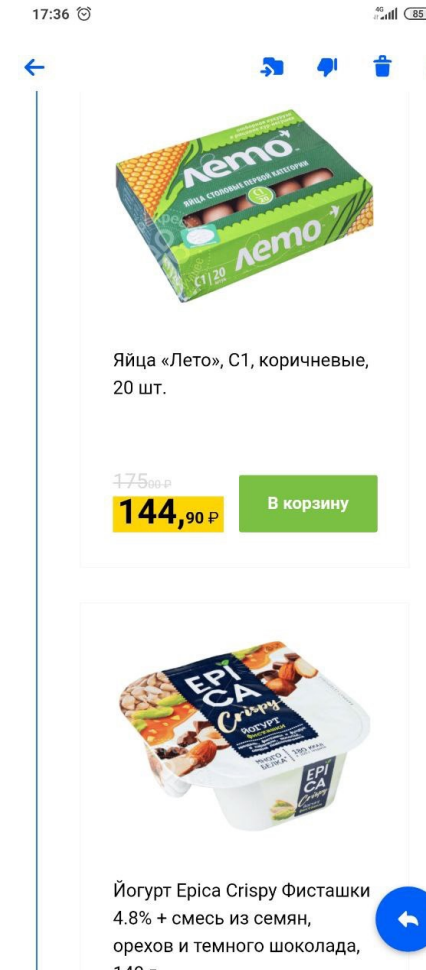
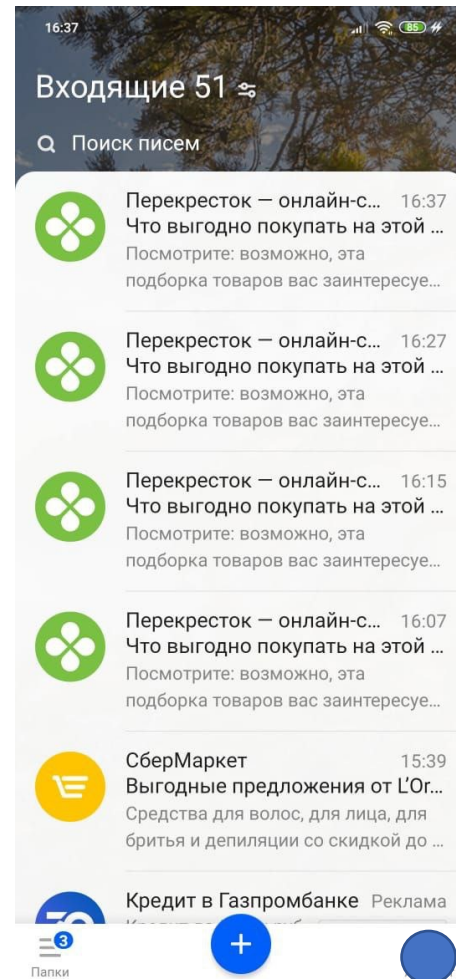


Подходы

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

Как использовать результаты ARL

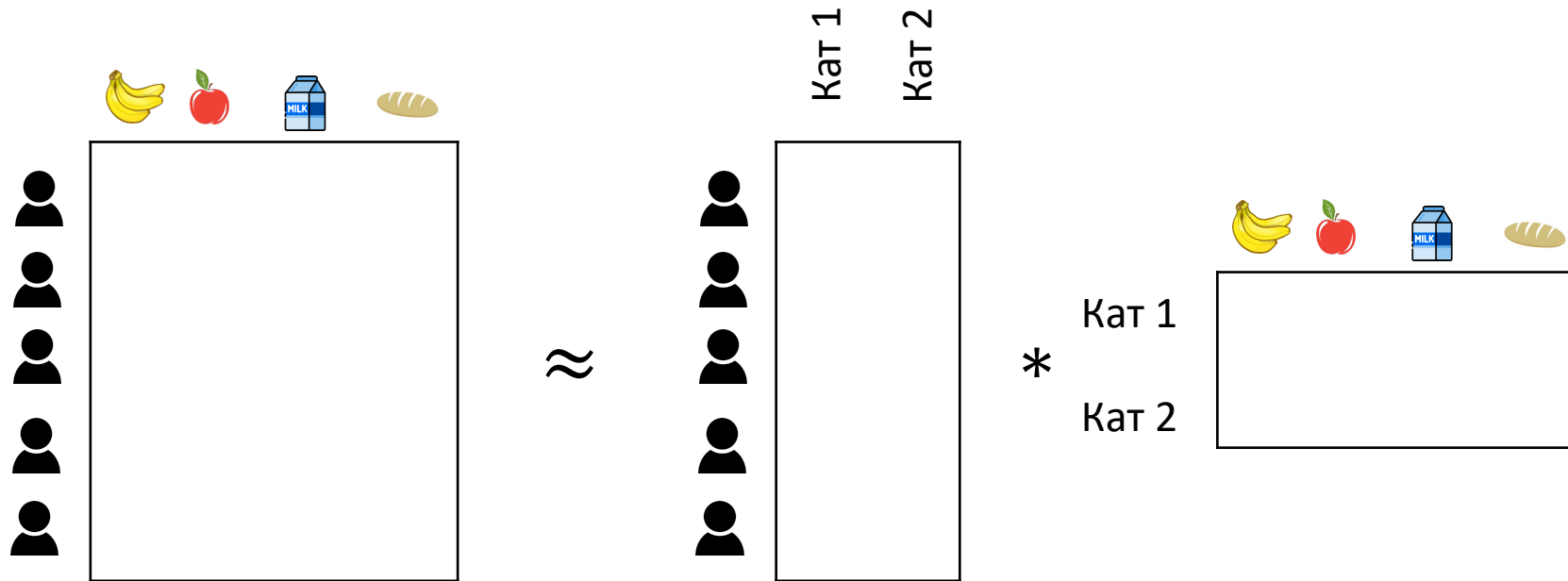


Подходы

Коллаборативная фильтрация

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

Подходы Факторизации



Подходы Факторизации. ALS

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

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

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

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

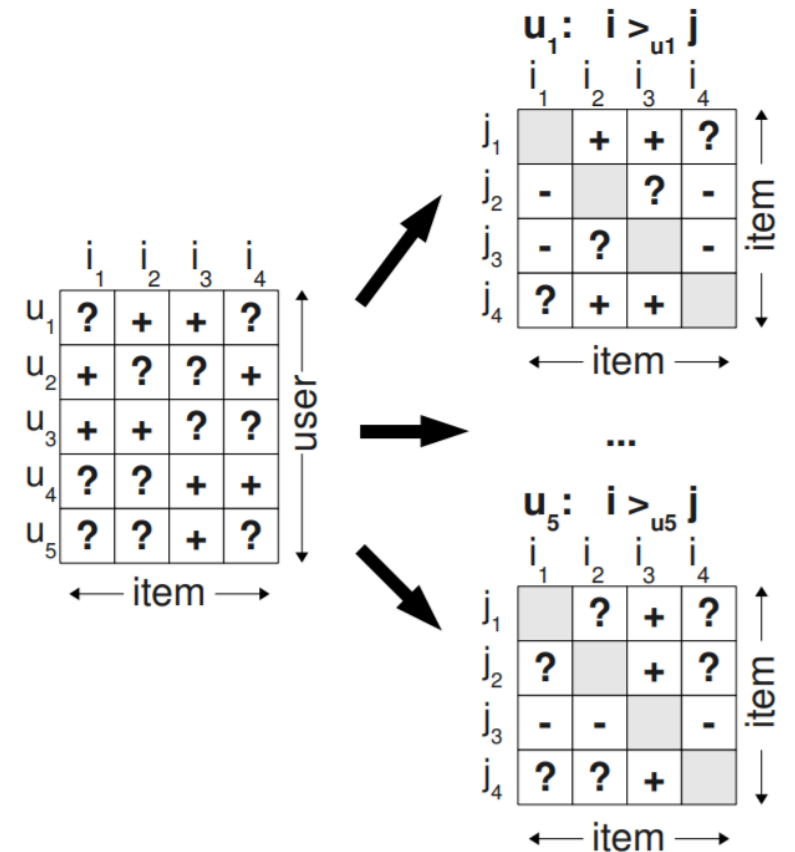
- После некоторых попыток оптимизации:

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

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

Подходы Факторизации. BPR

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



Эвристики вместо перс.реко?

Эвристики вместо перс.реко

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

Оценка качества?

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

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

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

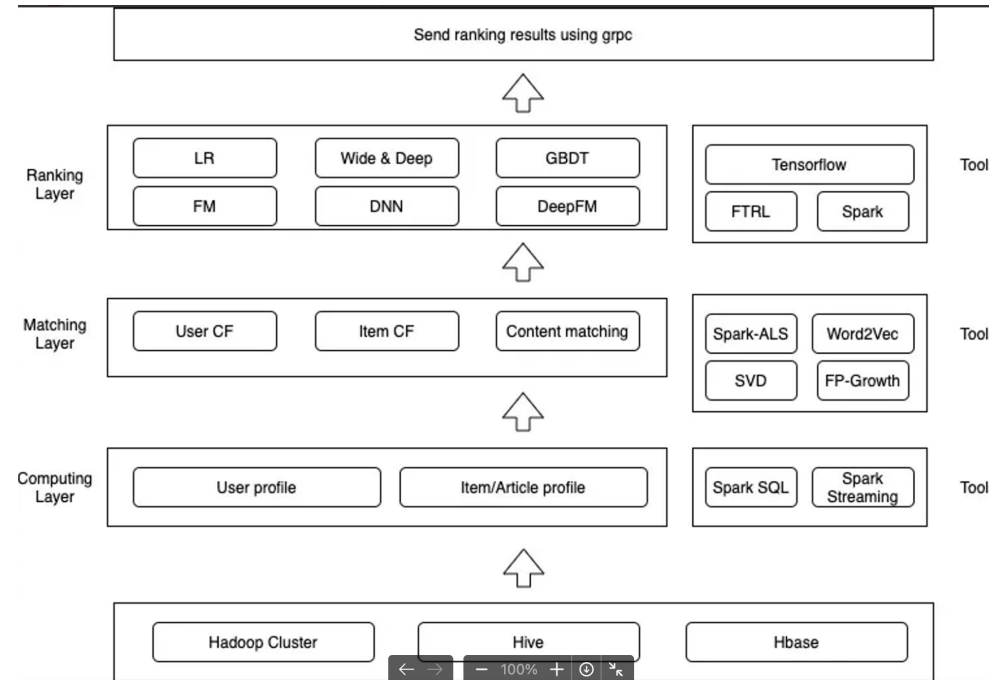
Оценка качества

- 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

Оценка качества

- 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

Схема систем



Ссылки по recsys

- 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/>

Ссылки по луковкам

- Целый диплом по One Class Classification
<https://homepage.tudelft.nl/n9d04/thesis.pdf>
- SVDD:
https://homepage.tudelft.nl/a9p19/papers/ML_SVDD_04.pdf
- Deep One Class Classification
<https://medium.com/analytics-vidhya/paper-summary-deep-one-class-classification-doc-adc4368af75c>
- Gaussian Mixture Models
<https://towardsdatascience.com/gaussian-mixture-models-explained-6986aaf5a95>

Ссылки по аплифтам

- **Подробная серия статей-объяснений:**
- https://habr.com/ru/company/ru_mts/blog/485980/
- **Курс по uplift-моделированию**
- <https://ods.ai/tracks/uplift-modelling-course/blocks/69a5fb73-458b-4fa0-9edf-201eb10b1538>
- **Библиотеки**
- <https://github.com/uber/causalml>
- <https://github.com/wayfair/pylift>
- <https://www.uplift-modeling.com/en/v0.3.0/index.html>