



РЕКОМЕНДАТЕЛЬНЫЕ СИСТЕМЫ НА ПРАКТИКЕ

СЕЛЕЗНЕВ АРТЕМ



ДЕНЬ 3



РАНЖИРОВАНИЕ

У меня в приложении есть каталог товаров,
мы могли бы сделать его индивидуальным для каждого?



Задачу рекомендаций можно перевести в
задачу ранжирования






Запомним, что ранжированный каталог, может не содержать элементов, которые понравятся пользователю, но надо стараться сделать каталог как можно лучше






РАНЖИРОВАНИЕ

самые романтические кафе


Места
Подборки

Romantic
Ресторан
Кутузовский просп., 2/1с1, этаж 35
★★★★★ 4,3 80 оценок
До открытия 15 мин




Unlock Cafe ✓
Кафе, банкетный зал
Лубянский пр., 19, стр. 1
★★★★★ 4,3 797 оценок
«Прикольная кафешка, не супер, но вполне прилично»
До открытия 15 мин






РАНЖИРОВАНИЕ

самые романтические кафе

Места Подборки

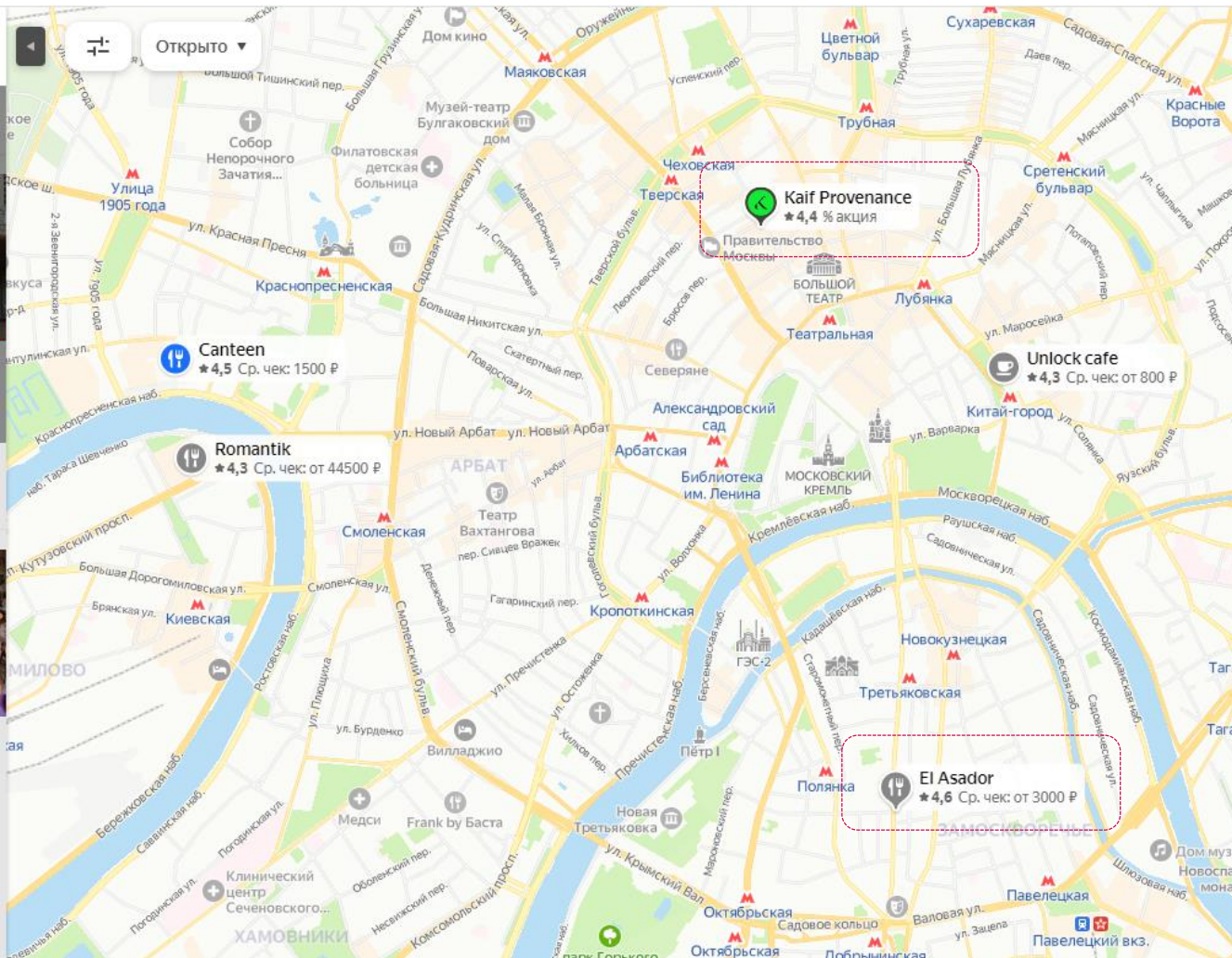


Romantic
Ресторан
Кутузовский просп., 2/1с1, этаж 35
★★★★★ 4,3 80 оценок
До открытия 15 мин



Unlock Cafe
Кафе, банкетный зал
Лубянский пр., 19, стр. 1
★★★★★ 4,3 797 оценок
«Прикольная кафешка, не супер, но вполне прилично»
До открытия 15 мин

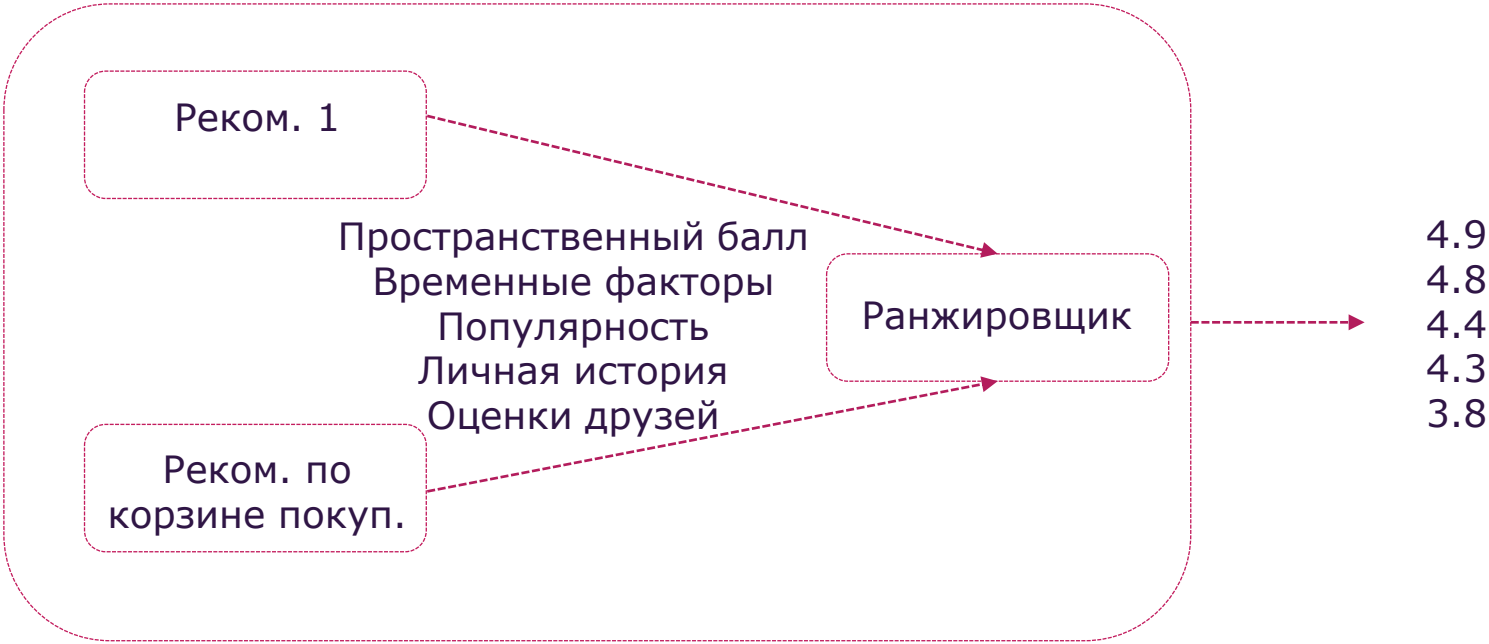
Открыто



РАНЖИРОВАНИЕ

Работают по принципу
гибридных
рекомендаций

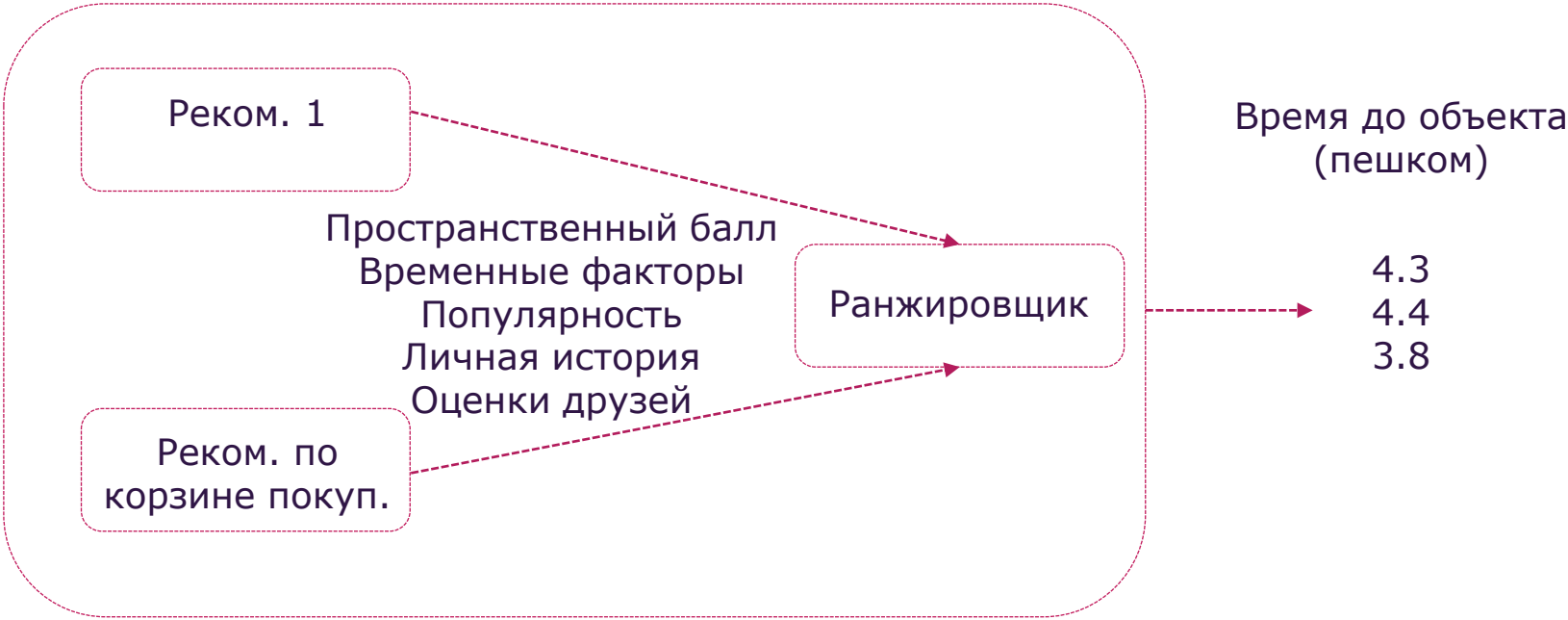
Но на основе
других факторов



РАНЖИРОВАНИЕ

Работают по принципу
гибридных
рекомендаций

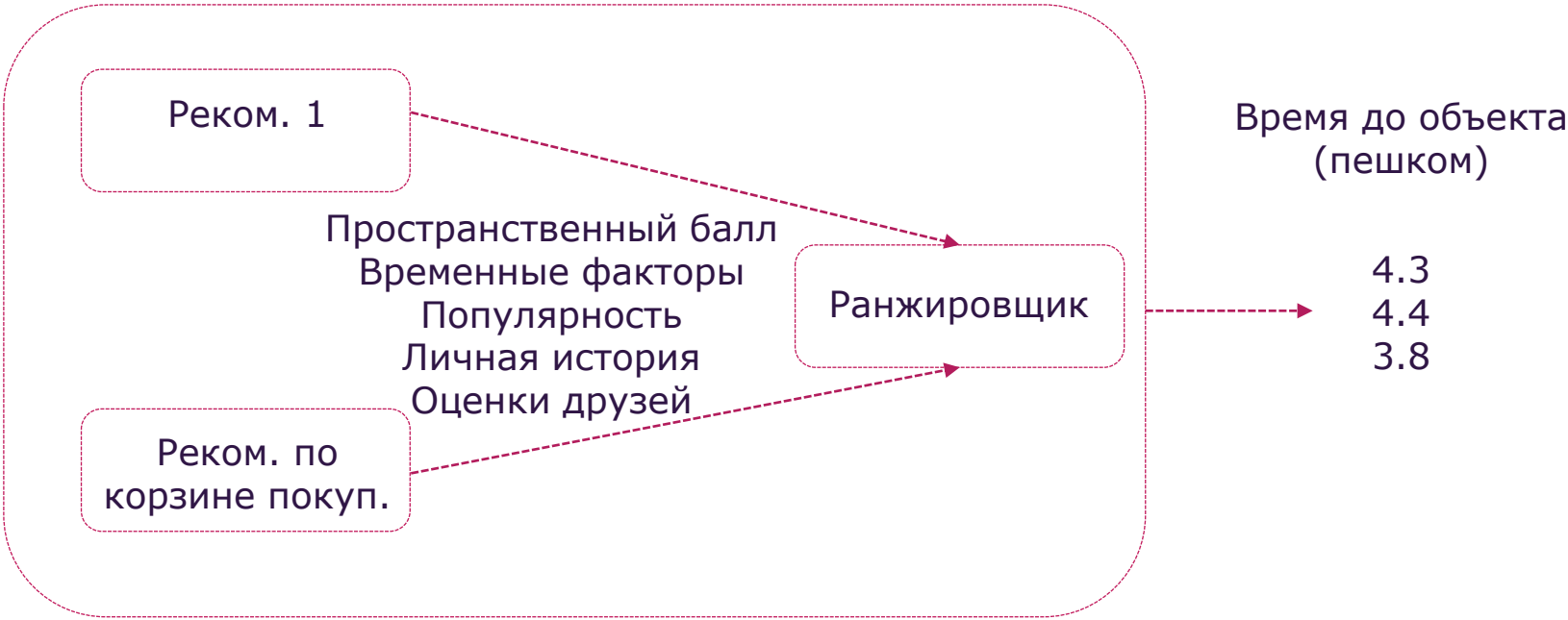
Но на основе
других факторов



РАНЖИРОВАНИЕ

Работают по принципу гибридных рекомендаций

Но на основе других факторов



РАНЖИРОВАНИЕ

Время							
Объект	Скоринг	пешком					
1	4.9	25	<div><div></div><div>$x - \min(x)$ $\max(x) - \min(x)$</div></div>	1	4.9	25	1.0
2	4.8	10		2	4.8	10	0.37
3	4.4	5		3	4.4	5	0.16
4	4.3	1		4	4.3	1	0
5	3.8	7		5	3.8	7	0.25

РАНЖИРОВАНИЕ

Объект	Время	
	Скоринг	пешком
1	4.9	25
2	4.8	10
3	4.4	5
4	4.3	1
5	3.8	7



Объект	Время		
	Скоринг	пешком	Оценка
4	4.3	1	0
3	4.4	5	0.16
5	3.8	7	0.25
2	4.8	10	0.37
1	4.9	25	1.0

РАНЖИРОВАНИЕ

Объект	Время	
	Скоринг	пешком
1	4.9	25
2	4.8	10
3	4.4	5
4	4.3	1
5	3.8	7



Объект	Время		
	Скоринг	пешком	Оценка
4	4.3	1	0
3	4.4	5	0.16
5	3.8	7	0.25
2	4.8	10	0.37
1	4.9	25	1.0



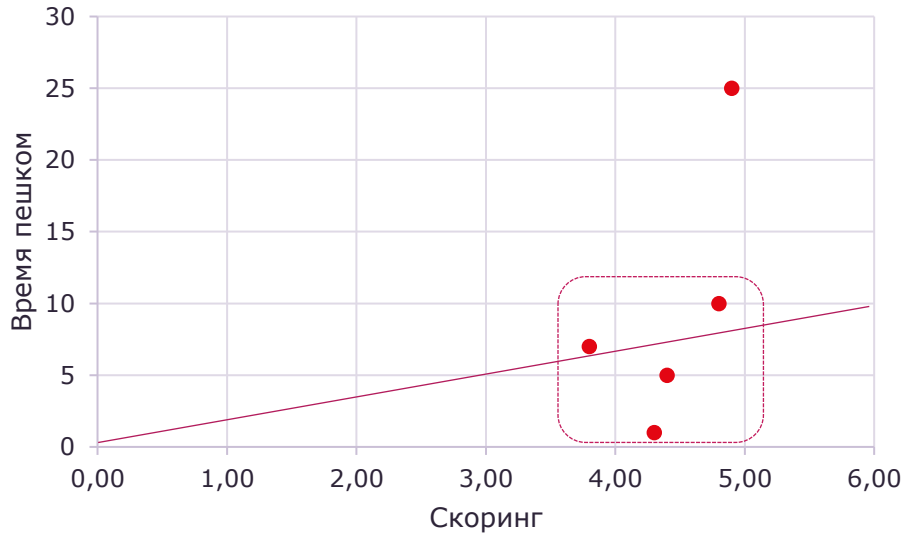
F (расстояние, скоринг)

РАНЖИРОВАНИЕ

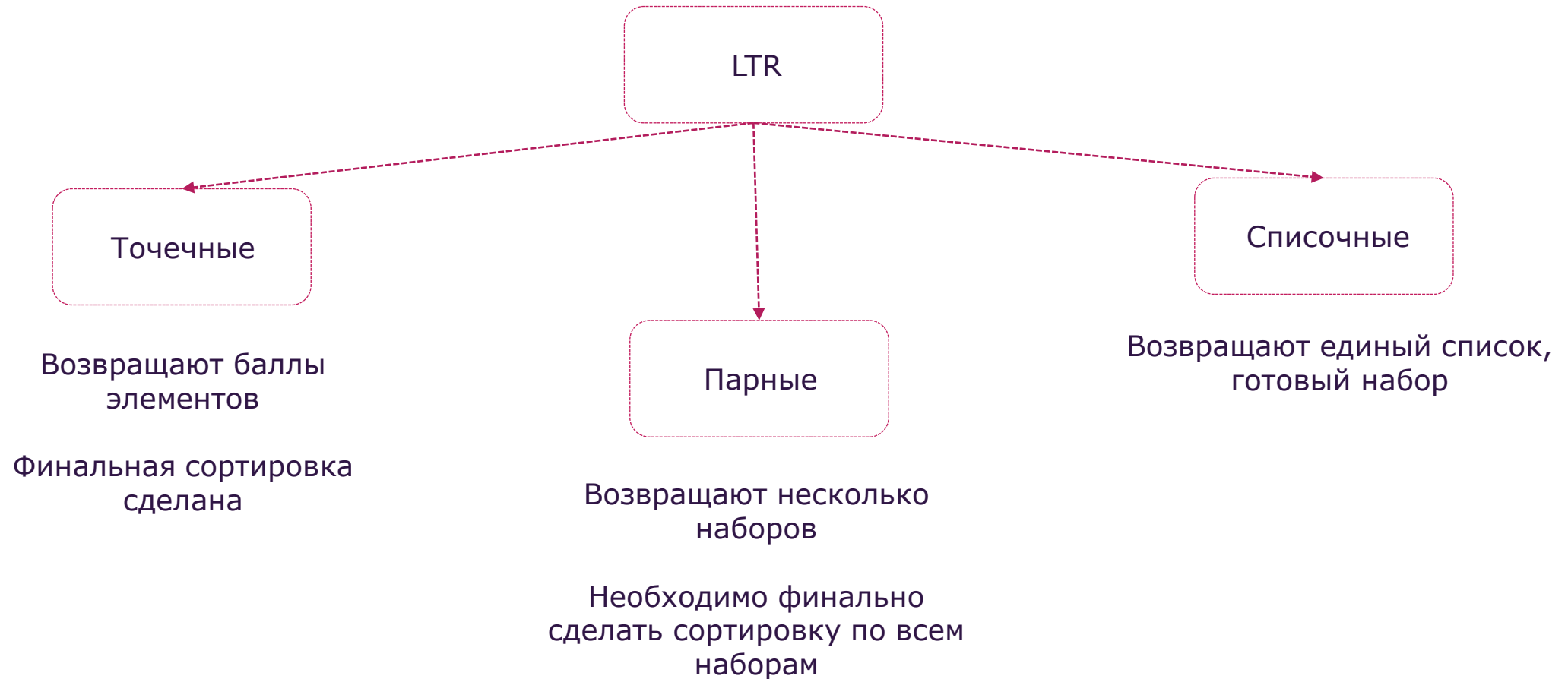
Объект	Скоринг	Время	
		пешком	Оценка
4	4.3	1	0
3	4.4	5	0.16
5	3.8	7	0.25
2	4.8	10	0.37
1	4.9	25	1.0



$F(\text{расстояние, скоринг}) = w_0 * \text{расстояние} + w_1 * \text{скоринг}$



РАНЖИРОВАНИЕ – LEARNING TO RANK



Рассмотрим на примере



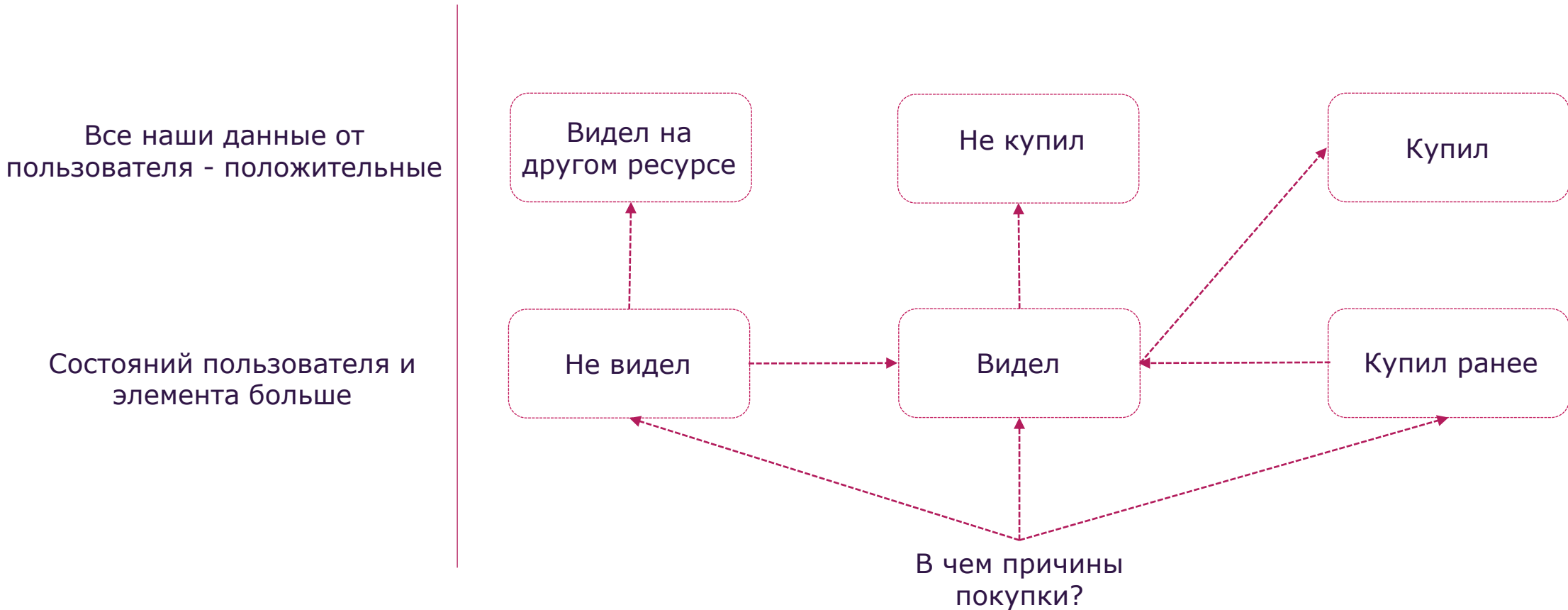
Все мои коллеги используют разные Байесовские подходы
А тут это возможно?



Есть BPR...



РАНЖИРОВАНИЕ – BAYESIAN PERSONALIZED RANKING





ДРУГИЕ РЕКОМЕНДАЦИИ

АНАЛИЗ КОНТЕНТА

Мы знаем больше о
элементах, которые
рекомендуем

Факты

Год

Состав

Качество

Любая неоспоримая информация

Теги

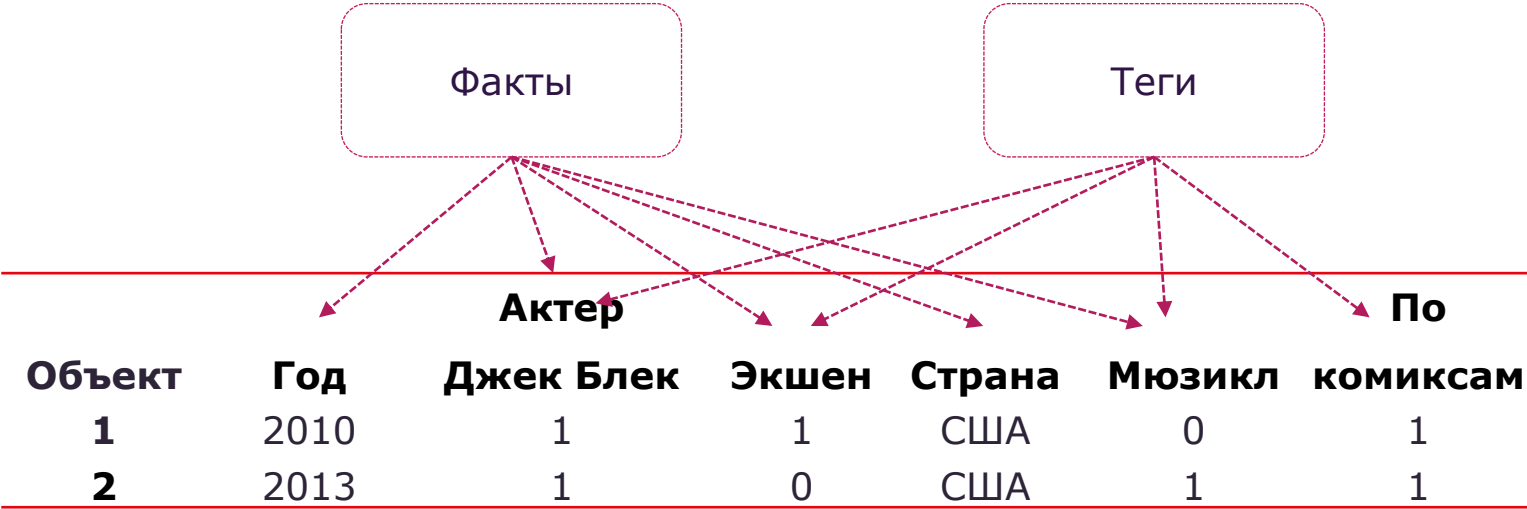
социальная метка

Общее мнение людей
в микро сочетании

Любая «социальная» информация

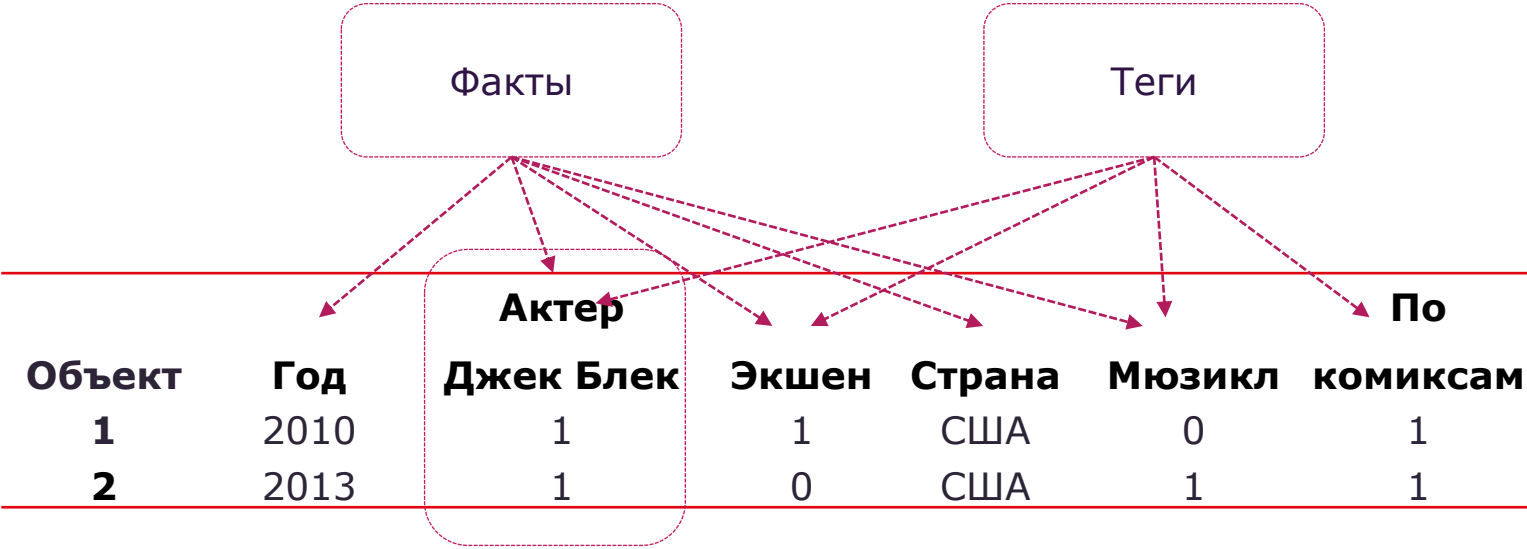
АНАЛИЗ КОНТЕНТА

Мы знаем больше о
элементах, которые
рекомендуем

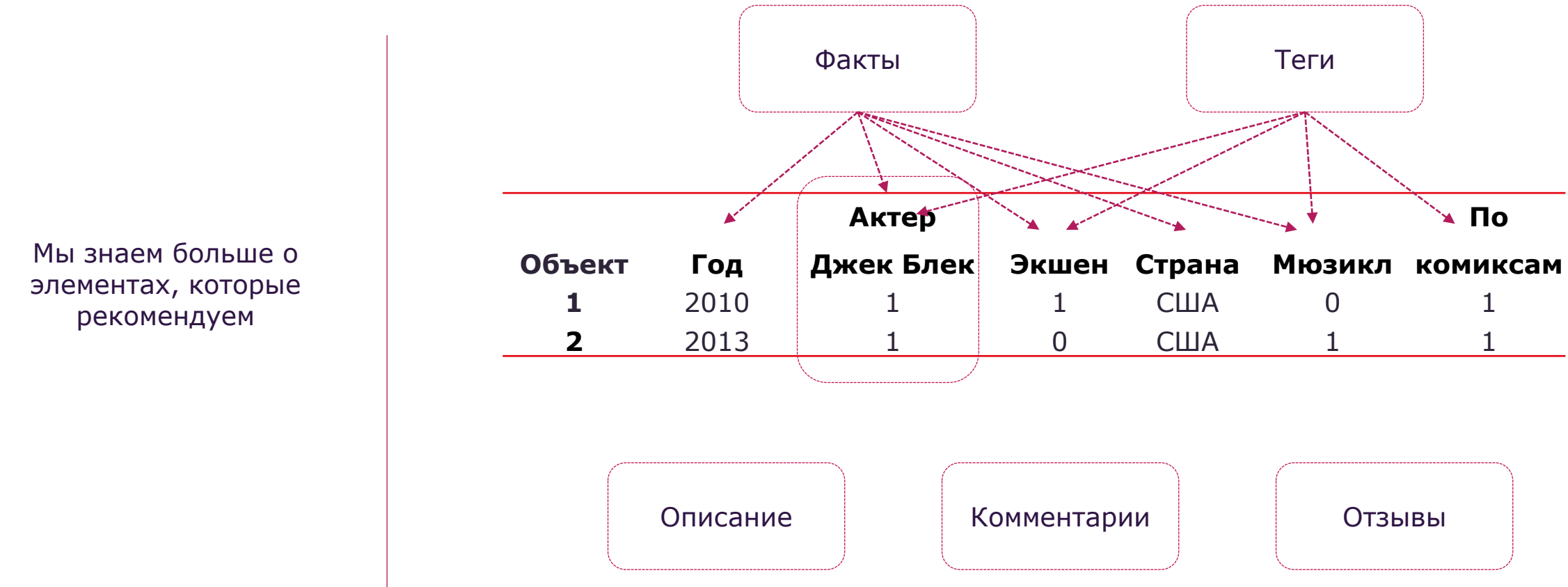


АНАЛИЗ КОНТЕНТА

Мы знаем больше о
элементах, которые
рекомендуем



АНАЛИЗ КОНТЕНТА. TF-IDF



АНАЛИЗ КОНТЕНТА. TF-IDF

$$\text{TF-IDF(слово, документ)} = \frac{(1 + \log(\text{частота слова}))}{(\log(\text{всего документов} / \text{кол-во документов со словом}))}$$

Мы знаем больше о
элементах, которые
рекомендуем

Объект	Описание	TF «педагог»	IDF «педагог»	TF * IDF
1	Фильм про педагога, который собрал свою музыкальную группу	1	1	1
2	Отличное роуд-муви по ностальгическим местам всех педагогов США	1	2	2
3	Студент-подросток должен победить бывших парней своей девушки. Обычная жизнь канадского студента.	1	3	3

АНАЛИЗ КОНТЕНТА. TF-IDF

$$\text{TF-IDF(слово, документ)} = \frac{(1 + \log(\text{частота слова}))}{(\log(\text{всего документов} / \text{кол-во документов со словом}))}$$

TF-IDF имеет дальнейшее развитие, найдите OKAPI-BM25

Объект	Описание	TF «педагог»	IDF «педагог»	TF * IDF
1	Фильм про педагога, который собрал свою музыкальную группу	1	0.405	0.057
2	Отличное роуд-муви по ностальгическим местам всех педагогов США	1	0.405	0.057
3	Студент-подросток должен победить бывших парней своей девушки. Обычная жизнь канадского студента	0	0	0

АНАЛИЗ КОНТЕНТА. LDA

LDA?
Генеративная модель?

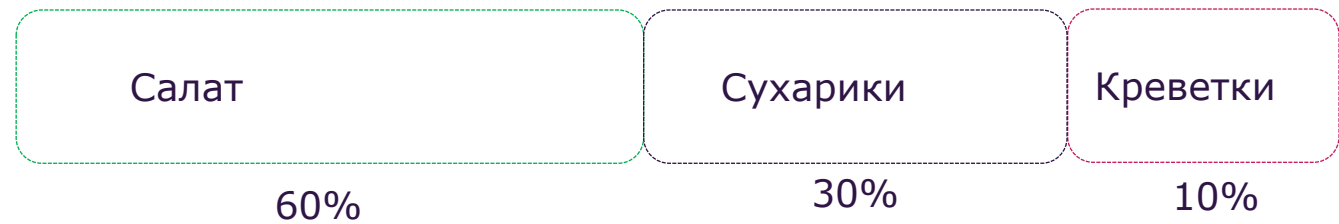
Креветки

Салат

Сухарики

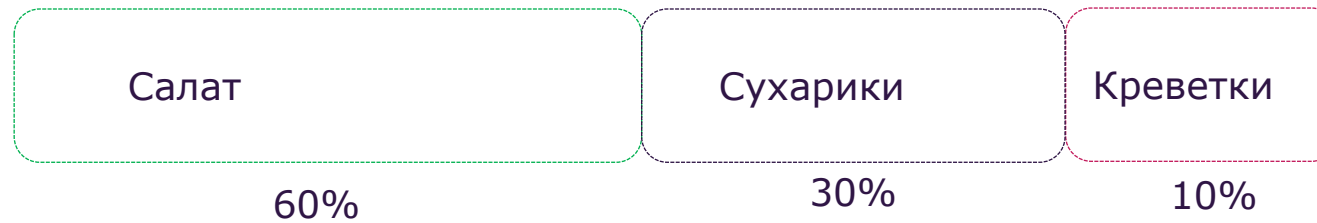
АНАЛИЗ КОНТЕНТА. LDA

LDA?
Генеративная модель?



АНАЛИЗ КОНТЕНТА. LDA

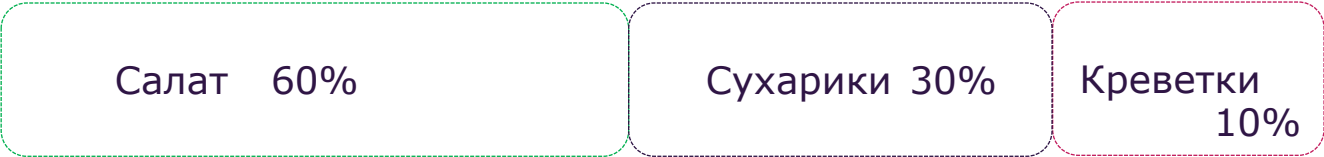
LDA?
Генеративная модель?



$$x = 0.6 * \text{Салат} + 0.3 * \text{Сухарики} + 0.1 * \text{Креветки}$$

АНАЛИЗ КОНТЕНТА. LDA

$$x = 0.6 \cdot \text{Салат} + 0.3 \cdot \text{Сухарики} + 0.1 \cdot \text{Креветки}$$



LDA?
Генеративная модель?

АНАЛИЗ КОНТЕНТА. LDA

LDA

$$x = 0.6 \cdot \text{Морская тема} + 0.3 \cdot \text{Капитан корабля} + 0.1 \cdot \text{Герой}$$

А на море не было шторма

Смелый капитан

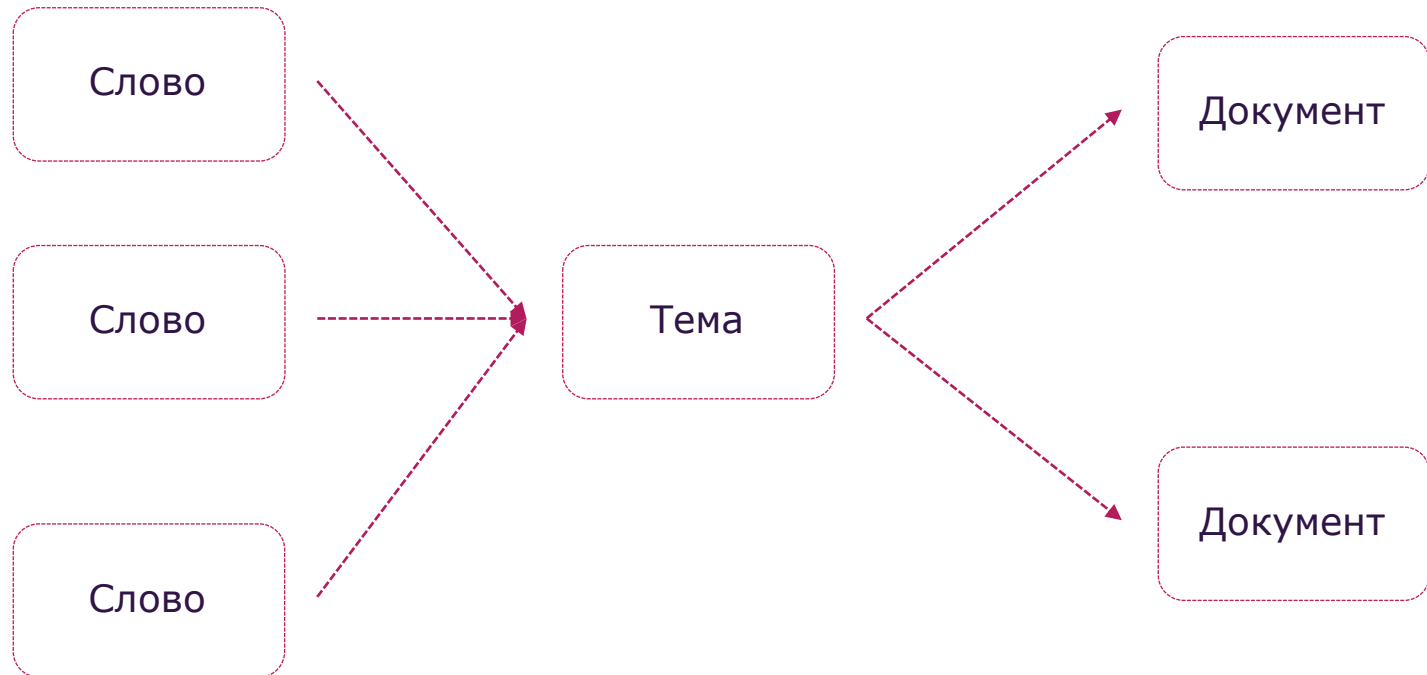
Он герой

АНАЛИЗ КОНТЕНТА. LDA

LDA

По схеме Гиббса –
начинается формирование
со случайного добавления
слова к теме

$$x = \text{вес} * \text{тема} + \dots \text{вес} * \text{тема}$$



Рассмотрим на примере



А есть что-нибудь не стандартное?



W2V в качестве рекомендаций?



Рассмотрим на примере





ОЦЕНКА И ТЕСТИРОВАНИЕ

ОЦЕНКА И ТЕСТИРОВАНИЕ

Netflix Prize свели задачу рекомендации
к задаче точного прогнозирования оценок



ОЦЕНКА И ТЕСТИРОВАНИЕ

Но! Это только один из многих компонентов эффективной рекомендательной системы.

Нужно также учитывать такие факторы, как разнообразие, контекст, правдивость, свежесть и новизна.

ОЦЕНКА И ТЕСТИРОВАНИЕ

Основные шаги
тестирования

Проверка
метрикой

Регрессионный
анализ

Автономное
тестирование

AB Test

ОЦЕНКА И ТЕСТИРОВАНИЕ



ОЦЕНКА И ТЕСТИРОВАНИЕ



ОЦЕНКА И ТЕСТИРОВАНИЕ

Проверка метрикой

ОЦЕНКА И ТЕСТИРОВАНИЕ

Проверка метрикой

Ошибка = факт - предикт

ОЦЕНКА И ТЕСТИРОВАНИЕ. MAE

Проверка метрикой

MAE -> 0

Пользователь 1

Пользователь 2

$$\frac{\text{СУММА Всех (ABS(Факт – Предикт))}}{\text{Кол-во элементов}}$$

Объект	Факт	Предикт	Ошибка
1	3	4	1
1	3	2	1
1	4	3	1
1	2	3	1
2	2	5	3

ОЦЕНКА И ТЕСТИРОВАНИЕ. MAE

Проверка метрикой

MAE -> 0

Пользователь 1

Пользователь 2

$$\frac{\text{СУММА Всех (ABS(Факт – Предикт))}}{\text{Кол-во элементов}}$$

Объект	Факт	Предикт	Ошибка
1	3	4	1
1	3	2	1
1	4	3	1
1	2	3	1
2	2	5	3
			1.4

Т.е. здесь средняя, абсолютная ошибка между фактом и предиктом, в этом наборе данных.

ОЦЕНКА И ТЕСТИРОВАНИЕ. MAPE (%)

Проверка метрикой

MAPE -> 0

Факт != 0

$$\frac{1}{\text{Кол-во элементов}} * \text{СУММА Всех} \left(\frac{\text{ABS(Факт - Предикт)}}{\text{Факт}} \right)$$

Объект	Факт	Предикт	Ошибка
1	3	4	0.3
1	3	2	0.3
1	4	3	0.25
1	2	3	0.5
2	2	5	1.5

0.57

ОЦЕНКА И ТЕСТИРОВАНИЕ. MeDianAPE

Проверка метрикой

MDAPE -> 0

MDAPE – не чувствительна к выбросам

Медиана ($\frac{ABS(Факт - Предикт)}{Факт}$)

Объект	Факт	Предикт	Ошибка
1	3	4	0.3
1	3	2	0.3
1	4	3	0.25
1	2	3	0.5
2	2	5	1.5
			0.3

ОЦЕНКА И ТЕСТИРОВАНИЕ. RMSE (L2 LOSS)

Проверка метрикой

RMSE -> 0

RMSE -> RMSPE

RMSE -> RMSLE

Квадратный корень ($\frac{\text{СУММА Всех (Факт - Предикт)}^2}{\text{Кол-во элементов}}$)

Объект	Факт	Предикт	Ошибка	²
1	3	4	1	1
1	3	2	1	1
1	4	3	1	1
1	2	3	1	1
2	2	5	3	9

Кв.корень(13 / 5) = 1,61

ОЦЕНКА И ТЕСТИРОВАНИЕ. RMSE (L2 LOSS)

Проверка метрикой

RMSE -> 0

RMSE -> RMSPE

RMSE -> RMSLE

Квадратный корень ($\frac{\text{СУММА Всех (Факт - Предикт)}^2}{\text{Кол-во элементов}}$)

Объект	Факт	Предикт	Ошибка	^2
1	3	4	1	1
1	3	2	1	1
1	4	3	1	1
1	2	3	1	1
2	2	5	3	9

Кв.корень(13 / 5) = 1,61

Т.е. здесь взвешенная средняя ошибка между фактом и предиктом, в этом наборе данных.

ОЦЕНКА И ТЕСТИРОВАНИЕ. ОЦЕНКА ТОЧНОСТИ

Проверка метрикой

Total population = P + N	Positive (PP)	Negative (PN)	Informedness, bookmaker informedness (BM) = TPR + TNR - 1	Prevalence threshold (PT) $= \frac{\sqrt{TPR \times FPR} - FPR}{TPR - FPR}$
Positive (P)	True positive (TP), hit	False negative (FN), type II error, miss, underestimation	True positive rate (TPR), recall, sensitivity (SEN), probability of detection, hit rate, power $= \frac{TP}{P} = 1 - FNR$	False negative rate (FNR), miss rate $= \frac{FN}{P} = 1 - TPR$
Negative (N)	False positive (FP), type I error, false alarm, overestimation	True negative (TN), correct rejection	False positive rate (FPR), probability of false alarm, fall-out $= \frac{FP}{N} = 1 - TNR$	True negative rate (TNR), specificity (SPC), selectivity $= \frac{TN}{N} = 1 - FPR$
Prevalence $= \frac{P}{P + N}$	Positive predictive value (PPV), precision $= \frac{TP}{PP} = 1 - FDR$	False omission rate (FOR) $= \frac{FN}{PN} = 1 - NPV$	Positive likelihood ratio (LR+) $= \frac{TPR}{FPR}$	Negative likelihood ratio (LR-) $= \frac{FNR}{TNR}$
Accuracy (ACC) $= \frac{TP + TN}{P + N}$	False discovery rate (FDR) $= \frac{FP}{PP} = 1 - PPV$	Negative predictive value (NPV) = $\frac{TN}{PN}$ $= 1 - FOR$	Markedness (MK), deltaP (Δp) $= PPV + NPV - 1$	Diagnostic odds ratio (DOR) $= \frac{LR+}{LR-}$
Balanced accuracy (BA) $= \frac{TPR + TNR}{2}$	$F_1 \text{ score}$ $= \frac{2PPV \times TPR}{PPV + TPR} = \frac{2TP}{2TP + FP + FN}$	Fowlkes–Mallows index (FM) $= \sqrt{PPV \times TPR}$	Matthews correlation coefficient (MCC) $= \frac{\sqrt{TPR \times TNR \times PPV \times NPV}}{\sqrt{FNR \times FPR \times FOR \times FDR}}$	Threat score (TS), critical success index (CSI), Jaccard index = $\frac{TP}{TP + FN + FP}$

ОЦЕНКА И ТЕСТИРОВАНИЕ. @K

Проверка метрикой

Релевантный контент в K позициях

K

ОЦЕНКА И ТЕСТИРОВАНИЕ. P@K

Проверка метрикой

Объект	Предикт	Отклик
1	4	1
1	2	0
1	3	1
1	3	0

ОЦЕНКА И ТЕСТИРОВАНИЕ. P@K

Проверка метрикой

Объект	Предикт	Отклик	P@K
1	4	1	1/1
1	2	0	1/2
1	3	1	2/3
1	3	0	2/4

0.5

ОЦЕНКА И ТЕСТИРОВАНИЕ. AP@K

Проверка метрикой

Ещё и ранжирование влияет на оценку?

Объект	Предикт	Отклик	P@K	Relevant	P@K * Rel
1	4	1	1/1	1	1/1 * 1
1	2	0	1/2	0	1/2 * 0
1	3	1	2/3	1	2/3 * 1
1	3	0	2/4	0	2/4 * 0

1

Сумма Rel

* 1.66

0.83

ОЦЕНКА И ТЕСТИРОВАНИЕ. AP@K

Проверка метрикой

Объект	Предикт	Отклик	P@K	Relevant	P@K * Rel
1	4	1	1/1	1	1/1 * 1
1	2	1	2/2	1	2/2 * 1
1	3	0	2/3	0	2/3 * 0
1	3	0	2/4	0	2/4 * 0

1

Сумма Rel

* 2

1

ОЦЕНКА И ТЕСТИРОВАНИЕ. AP@K

Проверка метрикой

Объект	Предикт	Отклик	P@K	Relevant	P@K * Rel
1	4	0	0/1	0	0/1 * 0
1	2	0	0/2	0	0/2 * 0
1	3	1	1/3	1	1/3 * 1
1	3	1	2/4	1	1/4 * 1

1

Сумма Rel

* 0.58

0.29

ОЦЕНКА И ТЕСТИРОВАНИЕ. MAP@K

Проверка метрикой

Объект	Предикт	Отклик	P@K	Relevant	P@K * Rel
1	4	1	1/1	1	1/1 * 1
1	2	0	1/2	0	1/2 * 0
1	3	1	2/3	1	2/3 * 1
1	3	0	2/4	0	2/4 * 0

1

Всего элементов

*

1

Сумма Rel

*

1.66

0.20

А есть что-нибудь не стандартное?



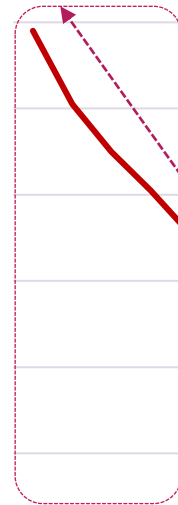
Метрики, которые действительно раскрывают
рекомендатор



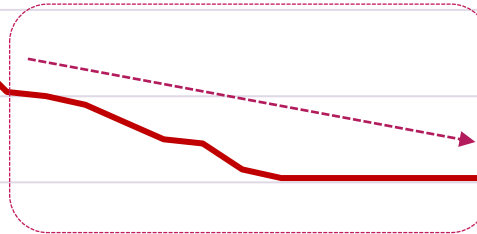
ОЦЕНКА И ТЕСТИРОВАНИЕ. РАЗНООБРАЗИЕ

Автономное
тестирование

Популярные



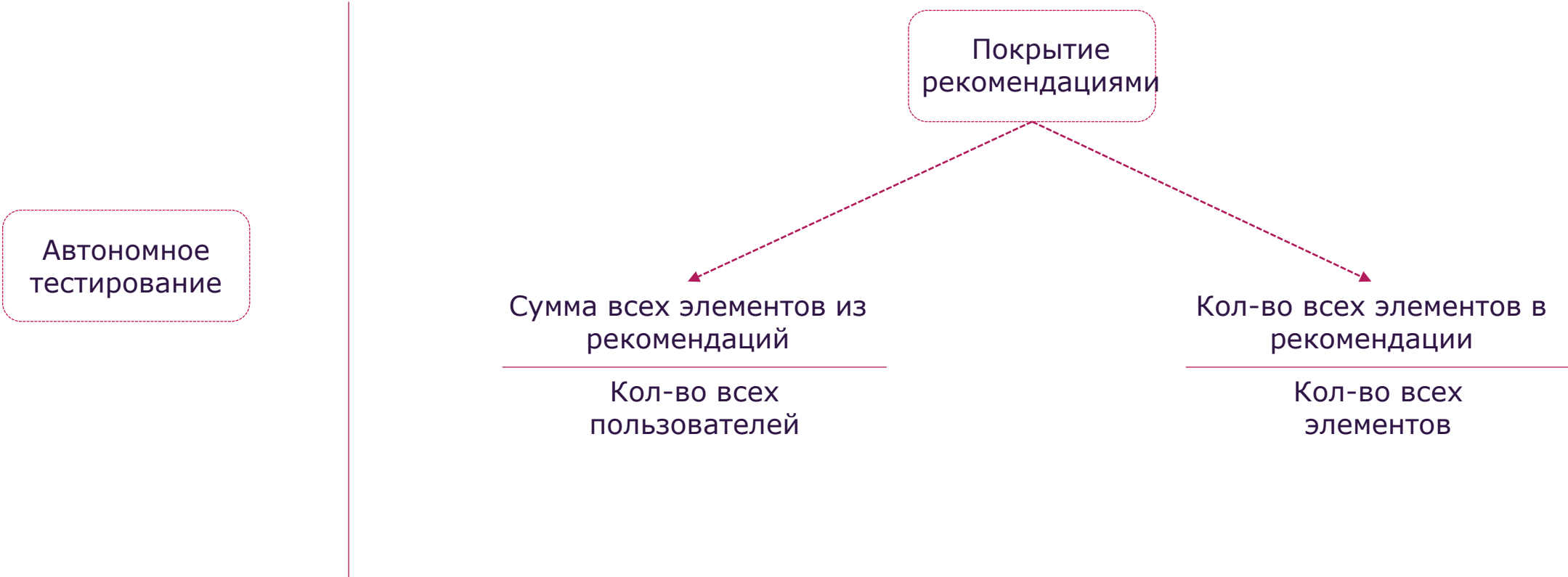
Не популярные



Эффект «Матфея».
Популярные становятся популярнее...

1 – список рекомендаций

ОЦЕНКА И ТЕСТИРОВАНИЕ. ОХВАТ / ПОКРЫТИЕ



ОЦЕНКА И ТЕСТИРОВАНИЕ. ИНТУИТИВНОСТЬ

Автономное
тестирование



Поисковые элементы = кол-во элементов из поиска

Неожиданность = $1 - \frac{\text{поисковые элементы в рекомендациях пользователя}}{\text{количество элементов в поиске}}$

Неожиданность
по пользователю

Кол-во всех
пользователей

?

Неожиданность
по пользователю

Кол-во всех
элементов

Рассмотрим на примере





ССЫЛКИ С ДОП.МАТЕРИАЛОМ

ДОП.МАТЕРИАЛ