
Оптимизация в поиске

— Почему цель важнее
скорости её достижения? —

Фёдор Романенко

Авито, Search Quality Lead

- Поисковая система Артефакт с 1999
- Яндекс.Картинки и Видео с 2005
- Свежесть и социальный поиск в Яндексе с 2009
- Инфраструктура качества поиска с 2013
- Качество поиска Авито с 2018



Оптимизация поиска — кому это нужно?

- огромная компания
- десятки миллионов пользователей
- поиск играет ключевую роль

Топ Рунета (2021)

- Яндекс ✓
- Wildberries ✓
- OZON ✓
- Mail.Ru / VK ✓
- Авито ✓
- 1С ✗
- Lamoda ✓
- HeadHunter ✓
- Delivery Club ✗
- AliExpress ✓

Релевантность

Решим когнитивную задачу

Релевантность в классифайде

«айфон 13 128 розовый» в Москве 232



Смартфон Apple iPhone 13 128Gb (m1ny3ru/A) Pink

65 990 ₽

6.1 дюйм, 128 Гб, 4 Гб

Смартфон Apple iPhone 13 128Gb (M1ny3ru/A) Pink.

Телефоны

● Шоссе Энтузиастов ⚡ 6–10 мин.
42 минуты назад



iPhone XR 128gb

20 590 ₽

Рыночная цена

Показать телефон

Написать

6.1 дюйм, 128 Гб, 3 Гб

iPhone XR 128gb. Оптом И В Розницу! Оригинальный Магазин. ...

Внимание Распродажа! · Любый Цвет, Любой Объем Памяти.

Запечатан в заводскую пленку. Полный комплект. Гарантия.

Документы. Доставка и самовывоз. iPhone весь модельный ряд. Цена

Телефоны

● Савёловская ⚡ до 5 мин.
8 часов назад ↑



iPhone 13 128Gb Все цвета в наличии

56 990 ₽

Рыночная цена

Показать телефон

Написать

128 Гб, 4 Гб

У нас вы можете купить новый оригинальный смартфон Apple iPhone 13 128 Gb. Указана минимальная цена за Японскую версию.

Стоимость других версии уточняйте в сообщениях и по телефону.

Запакованный, не обменный и не восстановленный. Мы успешно

Телефоны

● Багратионовская ⚡ 6–10 мин.
13 часов назад ↑



iPhone 13 (128), розовый

56 000 ₽

Рыночная цена

6.1 дюйм, 128 Гб, 4 Гб

iPhone 13 128gb новый, забираете прямо из салона, сначала проверка по уникальному номеру, потом оплата. Есть почти все модели. Звонить и писать с 8:00 до 22:00.

Телефоны

● Преображенская площадь ⚡ до 5 мин.

Доставка от 1 дня

2 недели назад



Как сделать ранжирование релевантным?

Бейзлайн

$$\text{score} = - \text{age}$$

Формула

$$\text{score} = (\text{title} == \text{query}) + \text{bm25f}(\text{query}, \text{desc})$$

ML-модель

$$\text{score} = \text{relev.predict}(f1(\text{query}, \text{title}), f2(\text{query}, \text{desc}))$$

Зачем тут ML?

- мы знаем, что должно быть наверху
- но у нас мало времени
- машина либо быстрая, либо умная
- ML прокачивает её когнитивные способности

Как научить ML релевантности?

Пусть прочтёт интернет и сама догадается!





Какие объявления релевантны запросу "айфон 13 128 розовый"?

1. Смартфон Apple iPhone 13 128Gb (mlny3ru/A) Pink
2. iPhone XR 128gb
3. iPhone 13 128Gb Все цвета в наличии



Самым релевантным объявлением для запроса "айфон 13 128 розовый" будет первое объявление, которое содержит точный запрос "iPhone 13 128Gb Pink".

Второе объявление не соответствует запросу, потому что оно не относится к модели, которую ищет пользователь (XR вместо 13) и не соответствует цветовой гамме (розовый).

Третье объявление может быть релевантным, но оно не указывает точно, есть ли модель в розовом цвете, что делает первое объявление более точным и соответствующим запросу.

Supervised learning

- взять выборку запросов
- скачать топы поиска
- разметить результаты на релевантность
- обучить pointwise классификатор

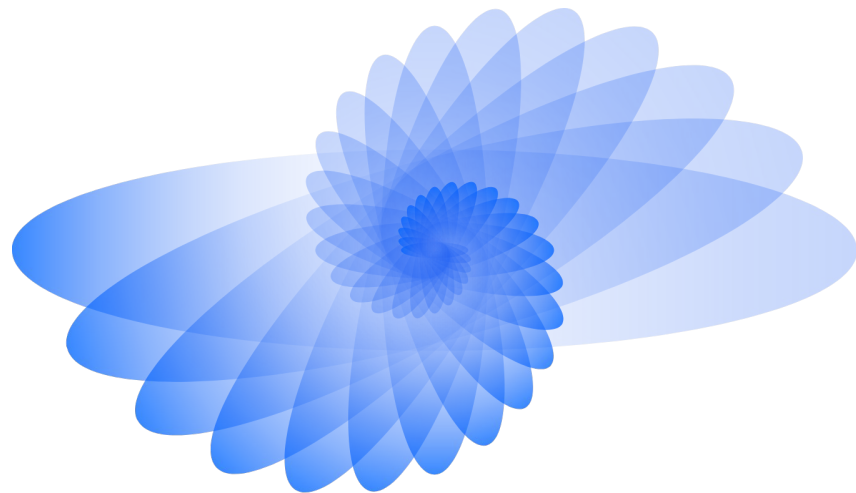
`relev.predict(query, item) → [0..1]`

... 100 тыс. оценок вполне хватит ✓

Кто соберет вам оценки?

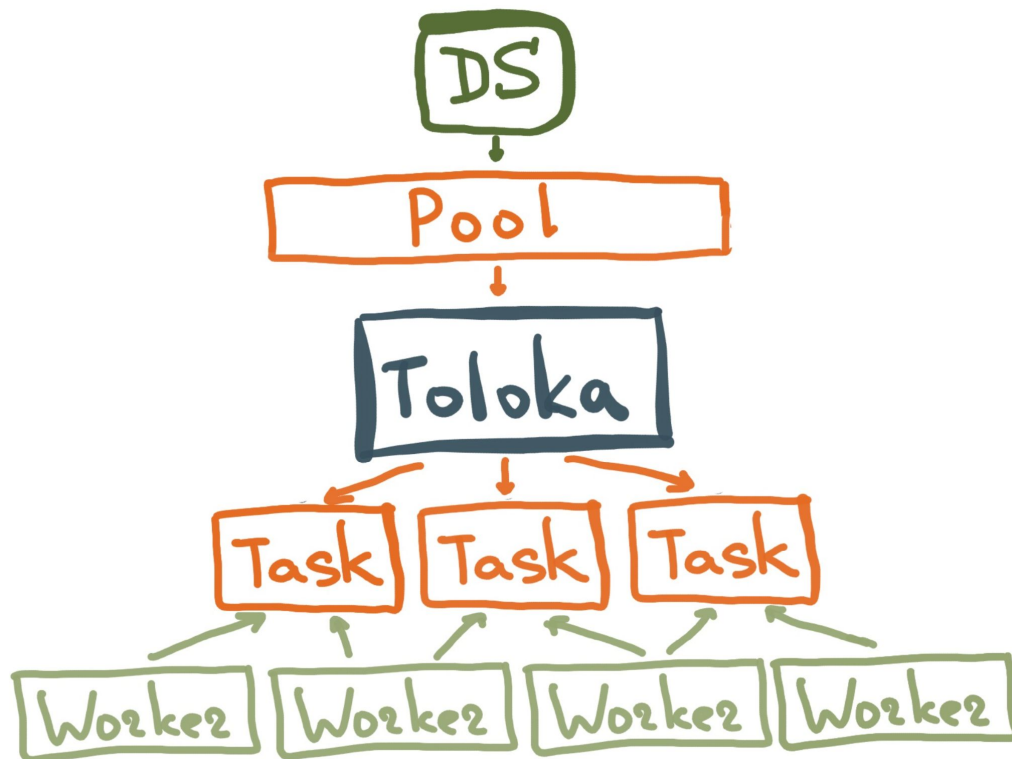


Toloka



toloka.ai

Сбор оценок через crowdsourcing

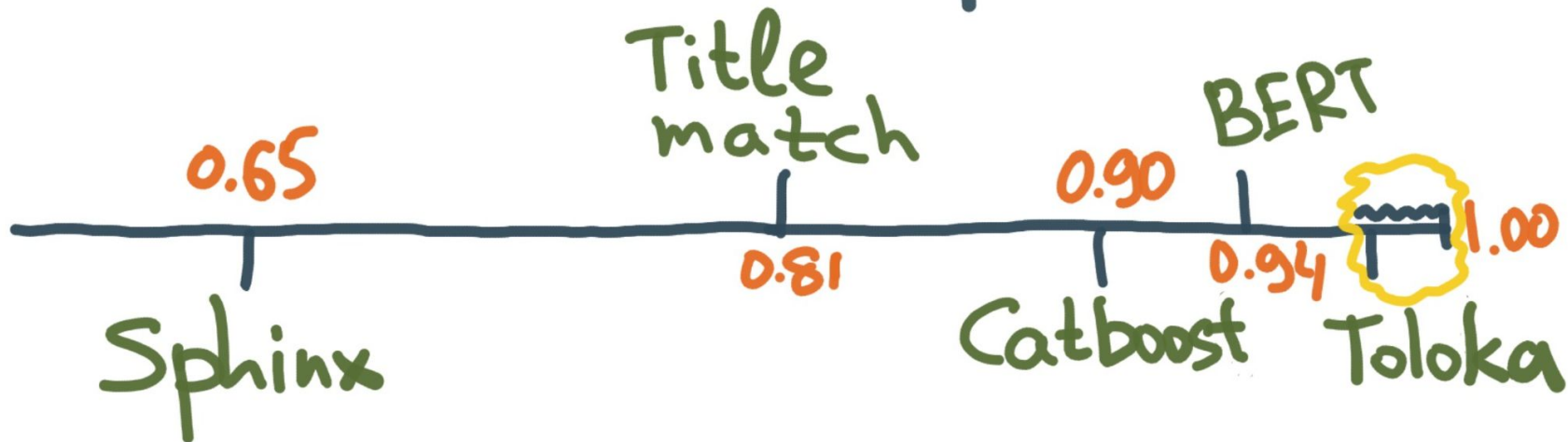


Качество таргета




Вы будете этим заниматься

По мере приближения к цели...

Relevance predict



Агрегация оценок в Толоке

		Василий	Дмитрий	Елена	Пётр	Константин	Итог
	кошка	x	✓	✓	x	x	x
	мышка	✓	✓	✓	x	✓	✓
	собака	x	✓	x	x	x	x
Σ	интеграл	✓	✓	x	✓	x	x

Maximum Likelihood Estimation of Observer Error-rates using the *EM* Algorithm

By A. P. DAWID† and A. M. SKENE‡

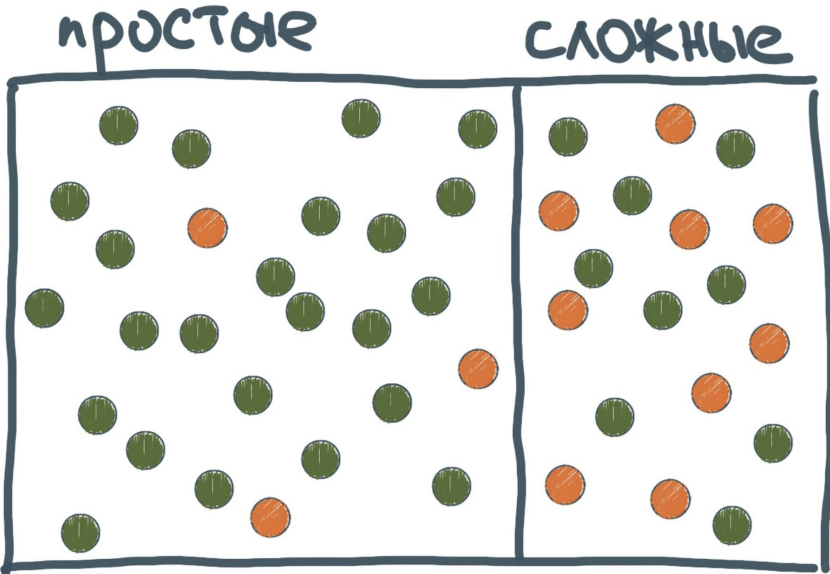
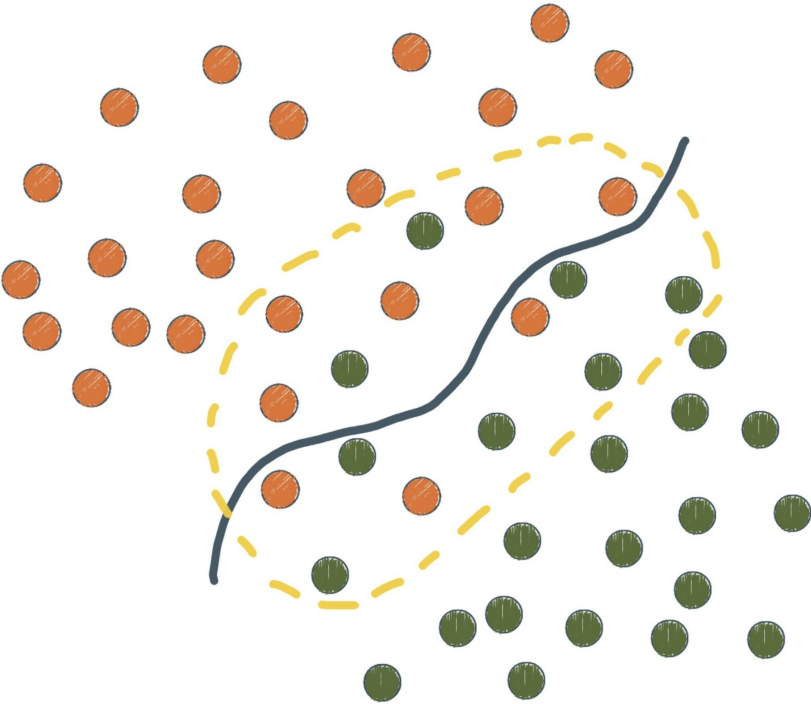
University College London

[Received October 1977. Revised May 1978]

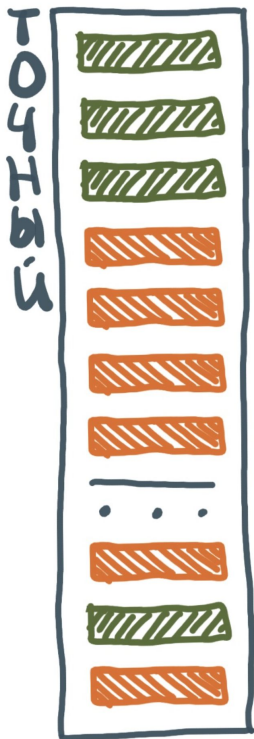
SUMMARY

In compiling a patient record many facets are subject to errors of measurement. A model is presented which allows individual error-rates to be estimated for polytomous facets even when the patient's "true" response is not available. The *EM* algorithm is shown to provide a slow but sure way of obtaining maximum likelihood estimates of the parameters of interest. Some preliminary experience is reported and the limitations of the method are described.

Нужна сложная задача



Релевантность: точность и полнота



iPhone 13 128gb (идеал)

51 000 ₺

Показать телефон

Написать

128 ГБ, 4 ГБ

iPhone 13 (128gb) без сколов, царапин и потёртостей. За качество телефонов отвечаю репутацией и даю гарантию 14 дней. Не вскрывался и не ремонтировался, читайте отзывы! - Телефон не восстановленный, не реф - Я всегда прийду вам на помощь в случае

Телефоны

● Водный стадион 🚶 до 5 мин.

Доставка от 1 дня

33 минуты назад ↑



Ура! Поехали?

Говорили нам, работает — не трогай!

Детские велосипеды в Москве 10 038

Базовый классифайд

Хобби и отдых

Велосипеды

Горные

Дорожные

BMX

Детские

Запчасти и аксессуары

Новые Б/у

Цена

Цена от до, руб.

Все

Частные

Компании

С Авито Доставкой ?

Рейтинг продавца

4 звезды и выше

Показать 10 тыс. объявлений



По date

Только в названиях

Сначала из Москвы

Сохранить поиск



Детский велосипед Stark tanuki 16 girl

4 000 ₽

● Братиславская ⚡ 16–20...
4 минуты назад



Велосипед детский алюминиевый 20"

20 000 ₽

● Дмитровская ⚡ 16–20 м...
4 минуты назад



Детский велосипед бу

1 000 ₽

● Щёлковская ⚡ от 31 мин.
10 минут назад



Велосипед детский для девочки

1 500 ₽

● Коптево ⚡ 21–30 мин.
16 минут назад



Детский велосипед «Wheeler»

5 500 ₽

● Улица Скобелевская ⚡...
18 минут назад



Трёхколесный велосипед


3 000 ₽

● Домодедовская ⚡ 16–2...
19 минут назад


Мы просто хотели добавить текстовый поиск...

🔍 Все категории


По дате Только в названиях Сначала из Москвы




Cybex zeno Bike (все расцветки)
Цена не указана
● Славянский бульвар ⚡ 6–10 мин.
2 минуты назад




Велосипед детский, игрушки, парта
Цена не указана
● Юго-Западная ⚡ 11–15 мин.
14 минут назад




Запчасти для велосипеда Lexus trike
Цена не указана
● Речной вокзал ⚡ 11–15 мин.
27 минут назад



Велосипед на литых дисках
Цена не указана
● Ломоносовский проспект ⚡ до 5 мин.
44 минуты назад



Запчасти на велосипед Lexus Trike
Цена не указана
● Улица Скобелевская ⚡ до 5 мин.



Stels Pilot 16 обмен
Цена не указана
Зеленоград
1 час назад

И сделать порелевантнее?



По умолчанию ▾

Только в названиях

Сначала из Москвы

Сохранить поиск

Показать на карте



Детский велосипед

8 000 ₽

● Бибирево 🚶 21–30 мин.
4 часа назад



Детский велосипед

5 000 ₽

● Отрадное 🚶 6–10 мин.
Доставка от 1 дня
11 часов назад



Детский велосипед

3 700 ₽

● Строгино 🚶 11–15 мин.
4 дня назад



Детский велосипед

2 500 ₽

● Румянцево 🚶 21–30 мин.
2 недели назад



Детский велосипед

5 000 ₽ 🚲

● Багратионовская 🚶 от 31 мин.
Доставка от 1 дня
10 часов назад



Детский велосипед

Бесплатно

● Улица академика Янгеля 🚶 до 5 мин.
2 недели назад

И вот, что вышло...

- 😍 поиск по запросу работает
- 😊 выдача релевантна
- 😐 все заголовки одинаковые
- 😞 в топе никому не нужное старье
- 😐 топ застыл
- 😱 продавцы не получают трафика
- 😬 монетизация сломалась

Забыли про свежесть, а зачем она?

Для покупателей

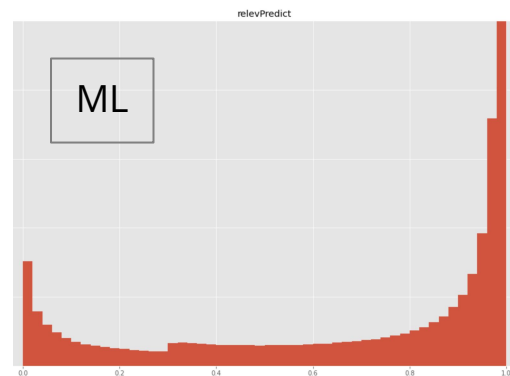
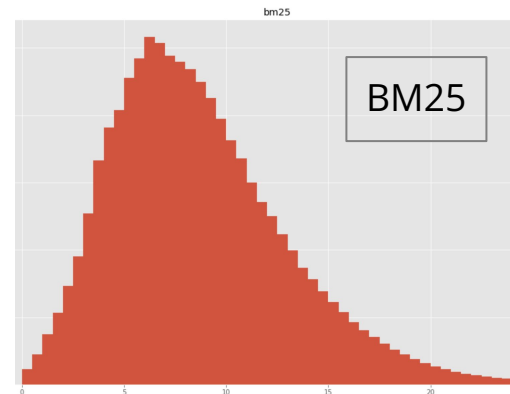
- найти посвежее
- помониторить выдачу

Для продавцов

- быстрее продать
- дать им всем трафик

Формулы релевантности не работают

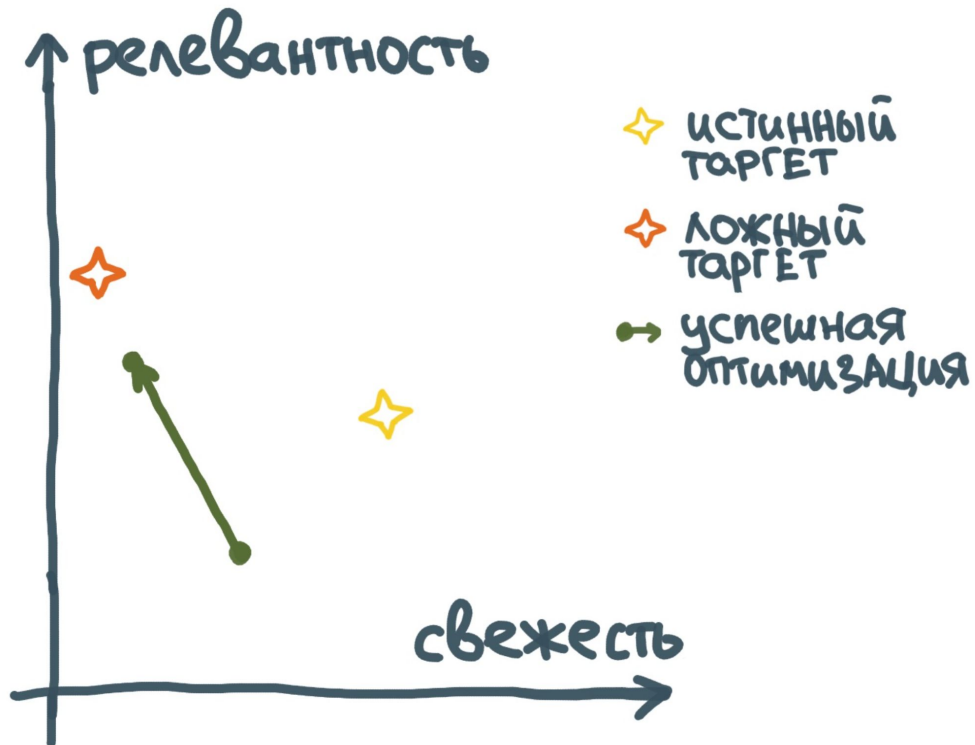
- они перекашивают выдачу,
 - убивают разнообразие
 - и полноту ответа
-
- только ML дает ~ 1.0 для всех типов релевантных
 - это нужно для учета нескольких критериев



Первое работающее решение

```
score = relev.predict(query, item)  
- 0.5 * item.age / (item.age + 3)
```

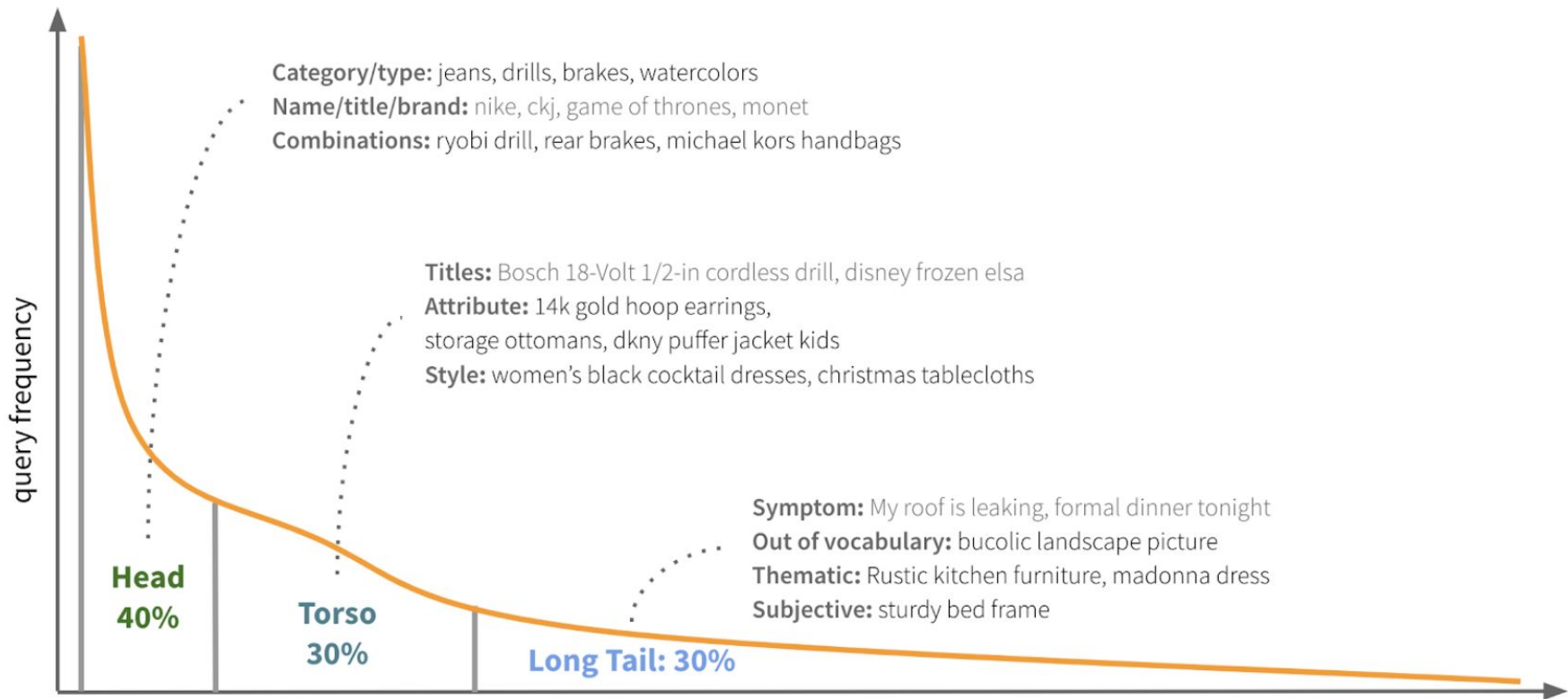
Урок: цель важнее оптимизации



Обучение по кликам

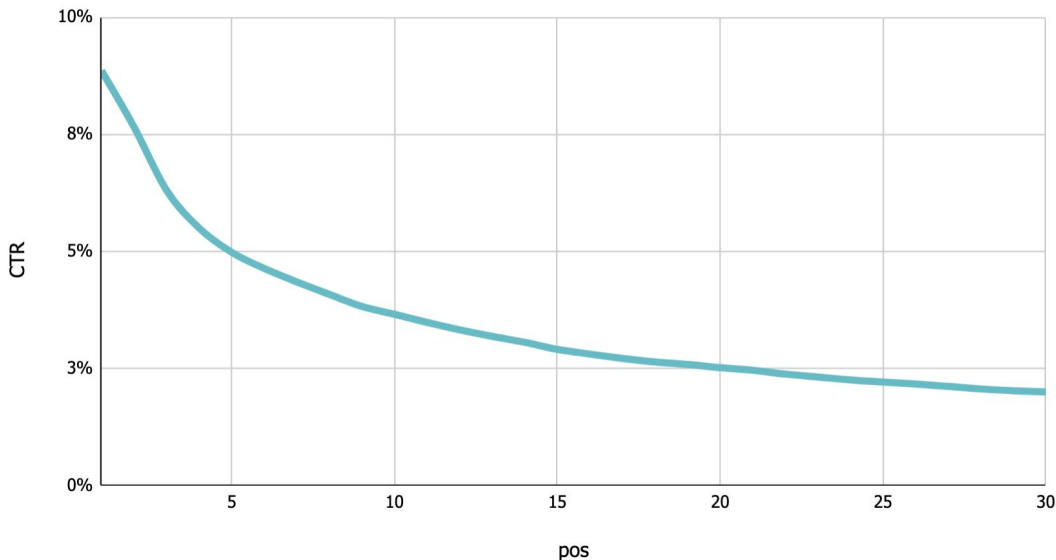
Много и бесплатно

Предсказать CTR(query, item)?

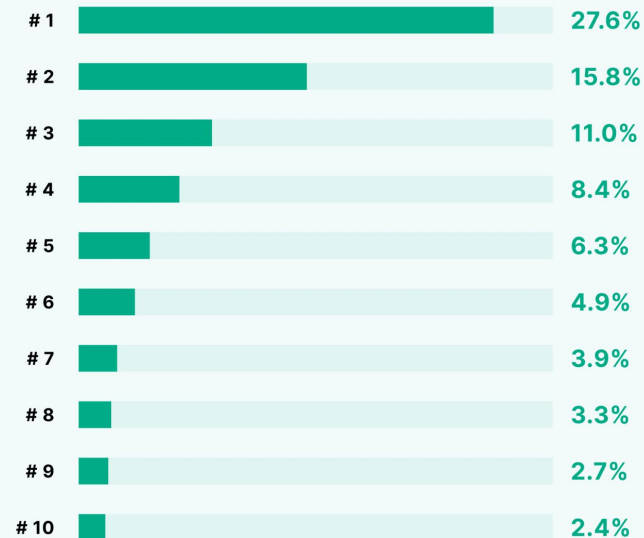


Предсказать clicked(serp, item)?

Avito sample CTR by position

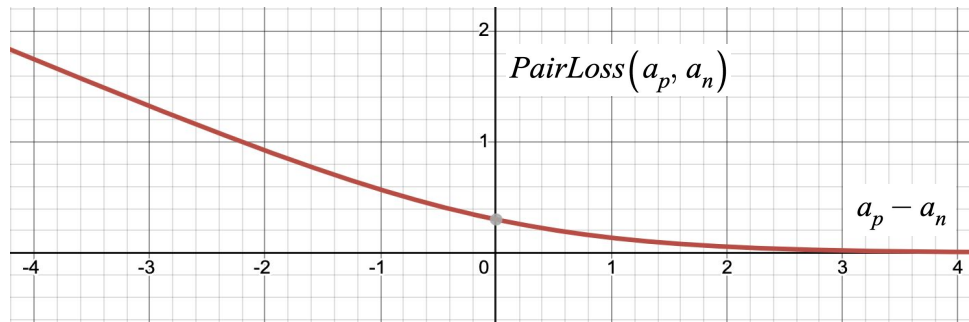


Google Organic CTR Breakdown By Position



CLICK THROUGH RATE

Предскажем кликнутый в паре



$$PairLoss(a_p, a_n) = -\log\left(\frac{1}{1 + e^{-(a_p - a_n)}}\right)$$

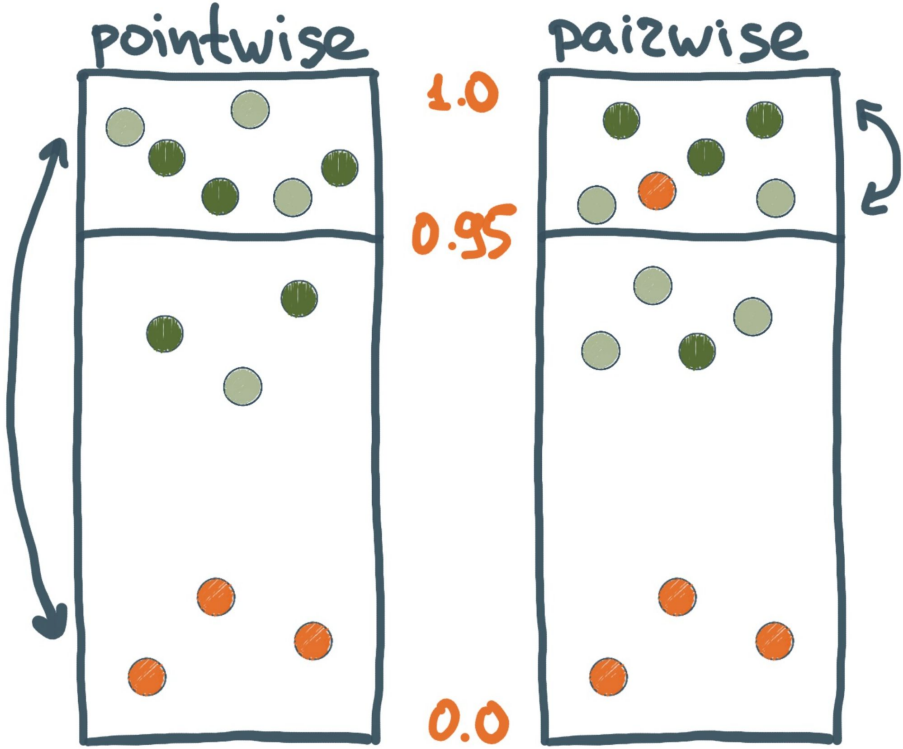
Составной таргет

Уточним нашу цель

Ранжируем комбинацией моделей

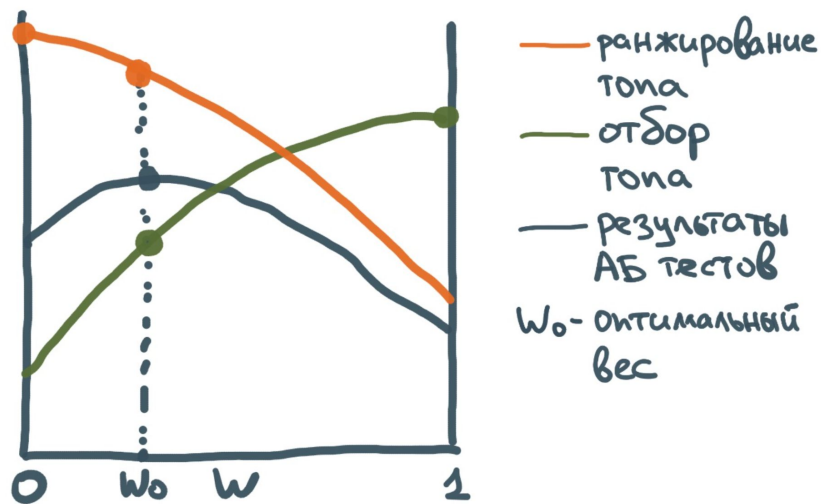
$$\begin{aligned} \text{score} = & 0.5 * \text{relev.predict}(\text{query}, \text{item}) \\ & + 1.0 * \text{click.predict}(\text{query}, \text{item}) \\ & - 0.4 * \text{item.age} / (\text{item.age} + 3) \end{aligned}$$

Pointwise vs Pairwise target rankings



Комбинация моделей — прагматичный выбор

- хуже отбирает топ, чем pointwise
- хуже ранжирует топ, чем pairwise
- но выигрывает АБ-тесты



$$\text{score} = w \cdot \text{pointwise} + (1-w) \cdot \text{pairwise}$$

А что ещё можно учесть?

- релевантность
- предсказание кликов / конверсий
- свежесть
- авторитетность / рейтинг продавца
- популярность
- близость / доставку
- цену / монетизацию ...



Видео



Картинки



Музыка



Кинопоиск



Маркет



Avito

North Star ✨ и качество поиска

½ времени: оптимизируем модель под таргет

½ времени: уточняем таргет






Много таргетов



в одной выдаче






Это ли они искали?

Я ✕ Найти   

Поиск [Картинки](#) [Видео](#) [Карты](#) [Товары](#) [Переводчик](#) [Все](#)

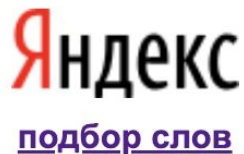
W **Гребенщиков, Борис Борисович — Википедия**
ru.wikipedia.org > [Гребенщиков, Борис Борисович](#)
Борис Борисович Гребенщиков — советский и российский поэт, музыкант, композитор, певец, гитарист. Основатель рок-группы «Аквариум».
[Основные даты биографии](#) · [Семья](#) · [Дискография](#) · [Проза](#) · [Религия](#)  

f **Борис Гребенщиков** 
[facebook.com](https://facebook.com/grebenshikov/) > [grebenshikov/](#)
The official page of **Boris** Grebenshikov Официальная страница **Бориса Гребенщикова**. ... **Борис Гребенщиков** и БГ+ в лондонском клубе Troubadour. Премьера на Дожде: фильм-концерт «Пятнадцать четвертей»... [Читать ещё](#) 



Советский и российский поэт, музыкант, композитор, певец, гитарист. Основатель рок-группы «Аквариум». Один из

Что им нужно?



борис гребенщиков

По словам

По регионам

История запросов

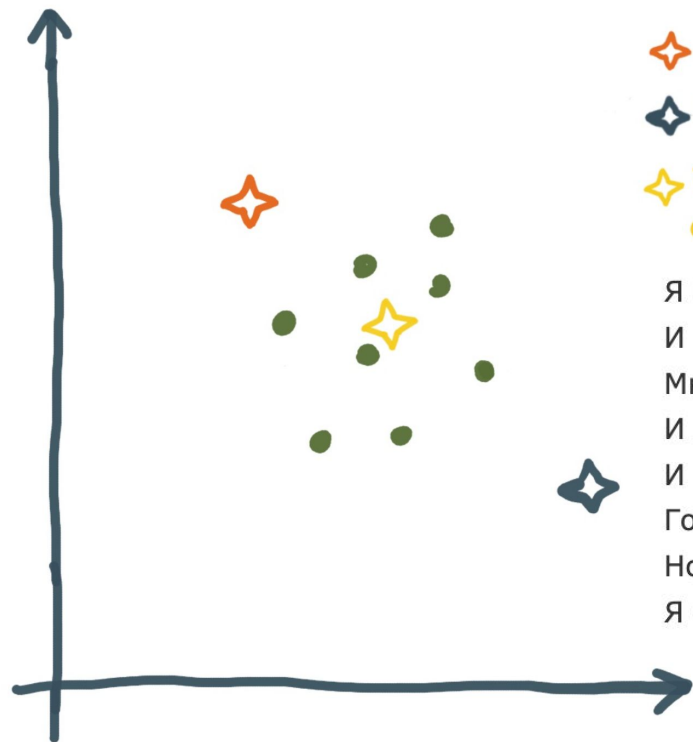
Что искали со словом «борис гребенщиков» — 55 664 показа в месяц

Статистика по словам

Показов в месяц [?]

борис гребенщиков	55 664
борис гребенщик	51 696
борис гребенщиков песни	3 958
борис гребенщиков слушать	3 527
где борис гребенщиков	1 214
борис гребенщиков скачать	1 196
борис гребенщиков фото	1 114
борис гребенщиков биография	1 107
борис гребенщиков сейчас	1 099
борис гребенщиков википедия	1 029

Усреднить таргеты: ни вашим, ни нашим



✧ песни

✧ биография

✧ песни -
автобиографии?

Я инженер на сотню рублей,
И больше я не получу.

Мне двадцать пять,
И я до сих пор не знаю, чего хочу.

✧ И мне кажется, нет никаких оснований
Гордиться своей судьбой,
Но если б я мог выбирать себя,
Я снова бы стал собой.

Неопределенность

- всем нужно одно и то же
- знаем точно, что
- но трудно определить, подходит ли результат?

Что делать?

- прокачивать ML

Неоднозначность

- разным нужно разное
- все знаем про результат
- но нельзя однозначно сказать, что он подходит всем

Что делать?

- блендинг



Поиск Картинки Видео Карты Товары Переводчик Все



Официальный сайт **Бориса Гребенщикова**

bg-aquarium.com

English. БГ | АКВАРИУМ. Читать ещё

О Группе · Книги · Концерты · Музыка · Картины · Контакт

~ официальный сайт



Гребенщиков, Борис Борисович — Википедия

ru.wikipedia.org > Гребенщиков, Борис Борисович

Борис Борисович Гребенщиков — советский и российский поэт, музыкант, композитор, певец, гитарист. Основатель рок-группы «Аквариум».

Основные даты биографии · Семья · Дискография · Проза · Религия

~ биография



Борис Гребенщиков — слушать онлайн бесплатно на...

music.yandex.ru > artist/453480

Борис Гребенщиков слушать лучшее онлайн бесплатно в хорошем качестве на Яндекс Музыка. Читать ещё

Альбомы · Инфо · Треки · Похожие

~ слушать онлайн бесплатно



Борис Гребенщиков - YouTube

youtube.com > Борис Гребенщиков

официальная страница **Бориса Гребенщикова**...

~ видео



Борис Гребенщиков

facebook.com > grebenshikov/

Борис Гребенщиков. 277,220 likes · 8,777 talking about this. The official page of **Boris Grebenshikov** ... **Борис Гребенщиков** и **БГ+** в лондонском клубе Troubadour. Премьера на Дожде: фильм-концерт «Пятнадцать четвертей»... Читать ещё

~



Борис Гребенщиков: беседы и песни на Би-би-си

bbc.com > russian/indepth/boris_grebenshchikov_60

Борису Гребенщикову 60. К юбилею поэта и музыканта Би-би-си открывает свои архивы и публикует фрагменты концертов, сыгранных в нашей студии, и интервью, прозвучавших в наших программах. Многое из его... Читать ещё

~ интервью



Boris Grebenshchikov (@bgrebenshikov) · Фото и видео...

instagram.com > bgrebenshikov/?hl=ru

Boris Grebenshchikov. Официальная страница **Бориса Гребенщикова** и группы Аквариум pomo.sh/bgrebenshikov. Читать ещё

~



Интервью БГ о войне, эмиграции и новом альбоме

holod.media > 2022/10/08/bg_interview/

Лидер «Аквариума» **Борис Гребенщиков** сразу после начала вторжения ... «Любовь во время войны» — песня, которую **Гребенщиков** написал восемь лет назад. Читать ещё

~



Гребенщиков Борис - аккорды(31), табы(4), видео(27)...

mychords.net > Наши > Гребенщиков Борис

Вм F# 2p Куплеты: **Борис Гребенщиков** Вм Песни любимых F# Песни выброшенных прочь Вм Похороненных без имени F# Замурованных в ночь Вм Песни перечеркнутых F# Песни сброшенных на. Читать ещё

~ песни



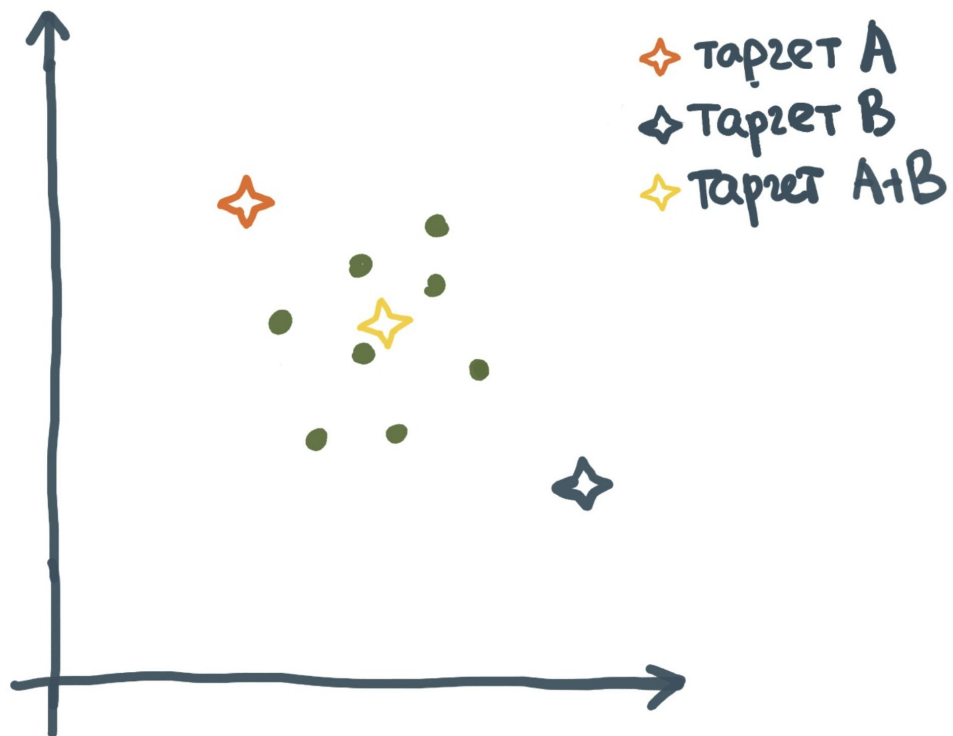
Борис Гребенщиков — фильмы — Кинопоиск

kinopoisk.ru > name/238568/

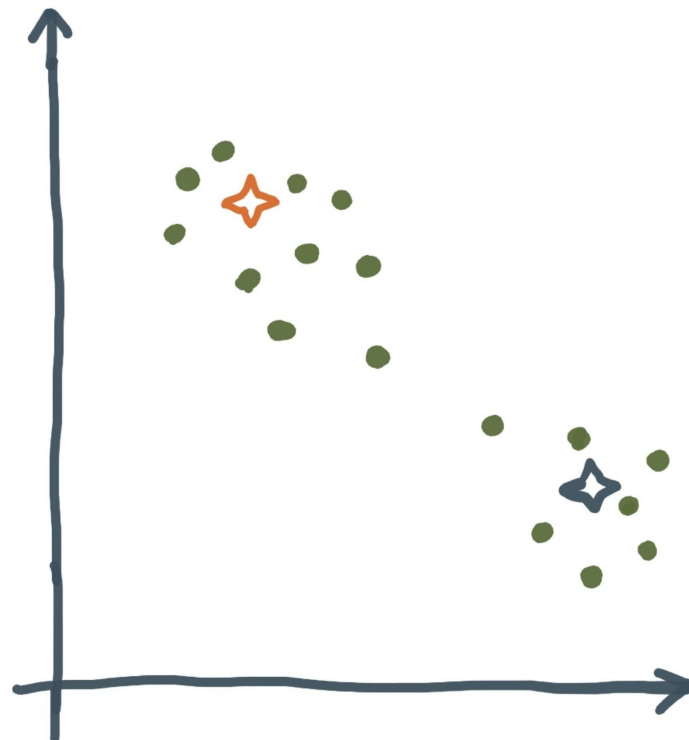
Борис Гребенщиков. О персоне. Карьера. ... «Кино» в кино: Цой, БГ и другие рокеры на экране. Читать ещё

~ фильмы

Ранкинг



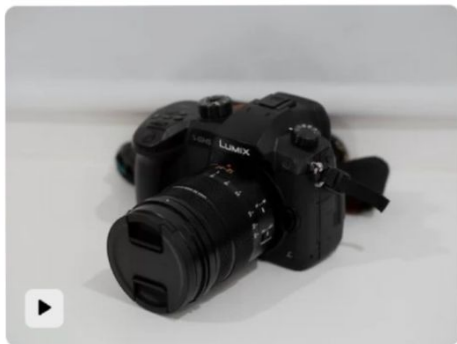
Блендинг



Много критериев

для оптимизации

Что важно в объявлении Авито?



Panasonic lumix GH5 + объективы

175 000 ₽

ВХ



Константин Трубин

4,9 ★★★★★ 46 отзывов

T&S

Продам лучшую камеру для съёмки видеоконтента — Panasonic lumix GH5. Идеальное состояние — чистая, красивая, ни царапинки. В комплекте два объектива: - Panasonic Leica DG Vario 12-60mm f/2.8-4.0. - Panasonic 45-200mm f/4.0-5.6. Также в комплекте: Зарядное

14 дней на возврат

T&S

Фототехника

● Сокол 🚶 до 5 мин.

3 часа назад ↑

ВХ

SX

×10 Продвинуто

MNЗ

ВХ - выбрать лучшее сейчас

SX - продать своё быстрее

T&S - на выдате безопасно

MNЗ - справедливое продвижение

Авито — требования к выдаче



Многокритериальный оптимизатор

- похож на линейное ранжирование
- критерий - фича
- веса фичей меняются по позициям

Используем матмодель блендинга!

Ранкинг

- оптимизируем одну позицию
- повторяем 10 раз независимо
- критерии одинаковы

Получаем:

- однообразную выдачу
- один критерий доминирует

Блендинг

- оптимизируем выдачу целиком
- выбор следующего зависит от предыдущих
- критерий меняется

Получаем:

- разнообразную выдачу
- пропорциональное влияние критериев

Процесс

Как с этим работать?

Дискавери ⇒ идея ⇒ реализация

- узнать все требования к выдаче
- оцифровать
 - важность критерия
 - качество результата
- положить в Блендер
- ...

Проверить на оффлайн-метриках

категория ↑	id ↑	# ↑	Δ dlvr ↑	Δ price ↑	Δ rating ↑	Δ relpred ↑	Δ title ↑	Δ █████ ↑	Δ █████ ↑	Δ age ↑	Δ █████ ↑
Не указано	-1	1149	-0.5%	-0.4%	-0.0%	-0.5%	-0.5%	1.4%	2.3%	-4.1%	-3.0%
Транспорт	1	1171	-0.6%	-0.7%	-0.2%	1.2%	0.6%	4.3%	3.0%	-4.0%	-4.6%
Работа	110	286	-	2.8%	-0.0%	-1.1%	-0.9%	-0.3%	-0.1%	-7.3%	1.3%
Услуги	113	194	-	-0.7%	-0.0%	-0.4%	-0.3%	1.1%	2.0%	-4.3%	-3.4%
Для дома и дачи	2	753	-0.4%	-1.0%	-0.1%	-0.8%	-0.5%	1.3%	1.8%	-2.4%	-2.2%
Животные	35	229	0.4%	-1.8%	0.0%	-0.0%	0.1%	0.7%	0.5%	-4.0%	0.1%
Недвижимость	4	612	-	-2.4%	-0.1%	0.4%	0.5%	1.9%	2.0%	1.2%	-4.4%
Личные вещи	5	750	-0.3%	-2.7%	-0.1%	-0.2%	0.3%	1.3%	3.3%	9.2%	-2.7%
Электроника	6	495	0.3%	-2.1%	-0.2%	-0.9%	-0.6%	2.7%	2.9%	1.6%	-7.3%
Хобби и отдых	7	295	-0.5%	-1.7%	-0.0%	-0.5%	-0.3%	0.8%	2.5%	3.0%	-3.5%
Готовый бизнес и оборудование	8	66	-5.4%	1.0%	-0.0%	-1.9%	-1.5%	1.4%	0.9%	3.9%	-2.3%

Проверить на АБ-тесте

Metric	MoE	t-stat	Lift	Mean	Num	Den
contact_████████████████████	2%	7.1σ	4.3%	0.085	441.7K	5.2M
create_order_████████████████████	2.6%	5.5σ	4.5%	0.009	48.9K	5.2M
████████████████████_widget	1.1%	3.5σ	1.2%	0.267	1.4M	5.2M
loc_████████████████████	1.7%	3.4σ	1.8%	0.065	338.2K	5.2M
████████████████████_grouped_ serp	1.6%	4.8σ	2.3%	0.161	831.1K	5.2M
████████████████████_clicks	1.1%	3.3σ	1.1%	0.509	2.6M	5.2M
total_████████████████████_confirm	1.3%	3.5σ	1.4%	0.645	3.3M	5.2M
total_████████████████████_confirm	1.3%	3.4σ	1.4%	0.660	3.4M	5.2M
user_loc_████████████████████	1.2%	4.0σ	1.5%	0.054	278.8K	5.2M
click_████████████████████_block	1.9%	7.1σ	4.2%	0.087	453K	5.2M
search_████████████████████	0.2%	3.7σ	-0.3%	0.519	111.3M	214.3M
total_████████████████████_widget	1.2%	4.2σ	1.5%	0.441	2.3M	5.2M
total_████████████████████_clicks	1.1%	3.4σ	1.2%	0.706	3.7M	5.2M
user_████████████████████_block	0.3%	4.9σ	-0.4%	1.902	9.8M	5.2M

профит!

ДОКЛАД

AI

11.04 / 16:00 – 16:45 (UTC+3)



Федор Романенко

Авито

**AI в вашем продукте больше не
следует правилам. Теперь
придется тестировать не
правильность, а эффективность**

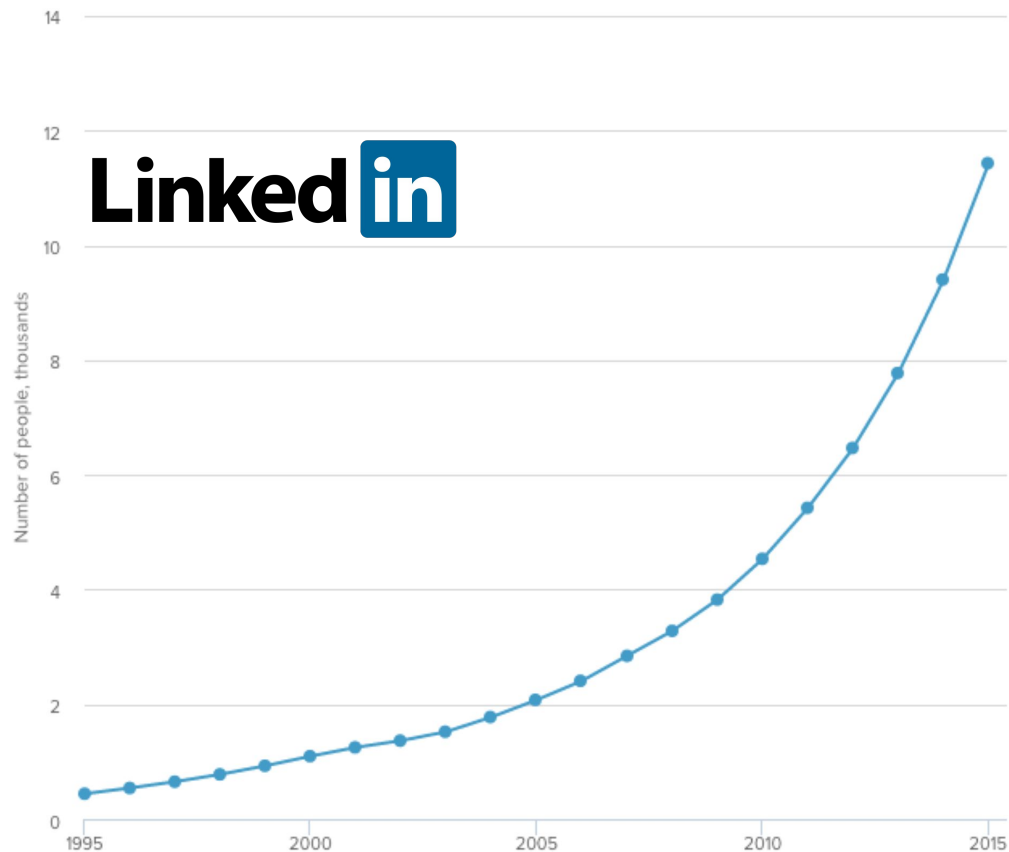
Послесловие

Какие выводы?

Развитие Data Science



CUMULATIVE NUMBER OF DATA SCIENTISTS OVER TIME



Чем заняться DS-инженеру? Мечты

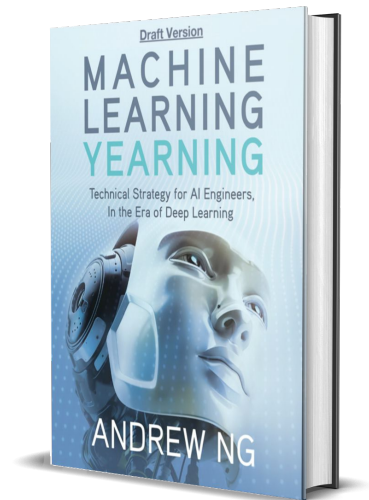


- но... архитектуру GPT-10 будете разрабатывать не вы,
- и у вас не будет \$1B на ее обучение,
- хотя использовать научитесь быстро

Чем заняться DS-инженеру?

Повседневность

- на работе будет много табличных данных
- мидлы будут оптимизировать чужие таргеты
- сеньоры — определять таргеты и ставить задачу ML!



avito.tech 2023

Фёдор Романенко

АВИТО

Search Quality Lead

fedor57@gmail.com

Telegram [@fedor57](https://t.me/fedor57)