

# Внедрение рекомендательных систем

# План

---

1

Практические  
аспекты внедрения  
рекомендательных  
систем

2

АБ-тесты

3

Многорукие  
бандиты

4

Совмещение "тяжелых"  
алгоритмов с онлайн  
алгоритмами

5

Эвристики

# Практические аспекты внедрения рекомендательных систем

# План внедрения рекоменд. систем



1. Изучение продукта
2. Метрики
3. Нагрузки
4. Данные
5. Архитектура сервиса
6. Валидация
7. Алгоритмы

# План внедрения рекоменд. систем



1. Изучение продукта
2. Метрики
3. Нагрузки
4. Данные
5. Архитектура сервиса
6. Валидация
7. Алгоритмы

# План внедрения рекоменд. систем



1. Изучение продукта
  - Где работает
  - Сколько элементов в блоке
  - Как отображается
  - Учесть разные версии продукта
  - ...
2. Метрики
3. Нагрузки
4. Данные
5. Архитектура сервиса
6. Валидация
7. Алгоритмы

# План внедрения рекоменд. систем



1. Изучение продукта
2. Метрики
  1. Бизнес-метрики
  2. Как бизнес-метрики коррелируют с ML
  3. Как вычисляются метрики
3. Нагрузки
4. Данные
5. Архитектура сервиса
6. Валидация
7. Алгоритмы

# План внедрения рекоменд. систем



1. Изучение продукта
2. Метрики
3. Нагрузки
  - Сколько пользователей в реальном времени
  - Сколько всего пользователей
  - Как объектов
  - Как часто прирастают
  - ...
4. Данные
5. Архитектура сервиса
6. Валидация
7. Алгоритмы



# План внедрения рекоменд. систем



1. Изучение продукта
2. Метрики
3. Нагрузки
4. Данные
  - Что хранится
  - Что нужно/можно собирать
  - Как собирается
5. Архитектура сервиса
6. Валидация
7. Алгоритмы

# План внедрения рекоменд. систем



1. Изучение продукта
2. Метрики
3. Нагрузки
4. Данные
5. Архитектура сервиса
  - Микросервис или часть монолита
  - Риал-тайм или нет
  - Как выкатывать новый алгоритм
6. Валидация
7. Алгоритмы

# План внедрения рекоменд. систем



1. Изучение продукта
2. Метрики
3. Нагрузки
4. Данные
5. Архитектура сервиса
6. Валидация
  - Онлайн или оффлайн
  - Проблема пузыря
  - ...
7. Алгоритмы

# План внедрения рекоменд. систем



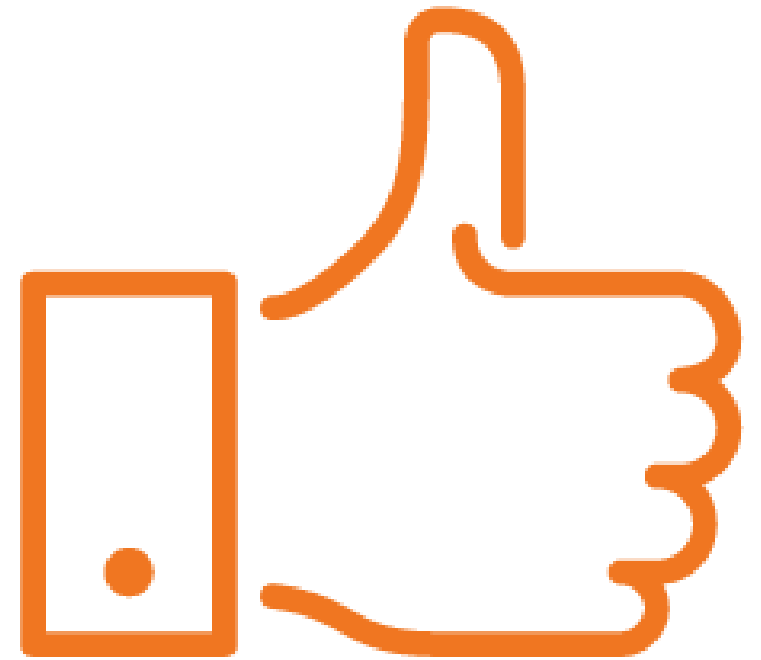
1. Изучение продукта
2. Метрики
3. Нагрузки
4. Данные
5. Архитектура сервиса
6. Валидация
7. Алгоритмы
  - Начать с тривиальных алгоритмов
  - Классических алгоритмов
  - Нейронные сети
  - Active Learning

Рекомендательные системы – это  
всегда часть какого продукта.

# Рекомендательные системы

---

**Рекомендательные системы** – это программы, которые пытаются предсказать, какие объекты (фильмы, музыка, книги, новости, товары) будут интересны пользователю, на основе определенной информации о его профиле.



Валидация / АВ-тесты

# Виды экспериментов



- Оффлайн эксперимент

Выполняется с использованием предварительно собранного набора данных о том, как пользователи выбирают или оценивают элементы. Используя этот набор данных, мы можем попытаться смоделировать поведение пользователей, которые взаимодействуют с системой рекомендаций.



# Виды экспериментов



- Исследования пользователя

Чтобы правильно оценить такие системы, необходимо собрать данные о реальных взаимодействиях пользователей с системой. Пользовательское исследование проводится путем набора испытуемых и их просят выполнить несколько задач, требующих взаимодействия с системой рекомендаций. Пока субъекты выполняют задания, мы наблюдаем и записываем их поведение, собирая любое количество количественных измерений, например, какая часть задания была выполнена, точность результатов задания или время, затраченное на выполнение задания.

# Виды экспериментов



- Онлайн-оценка

Во многих реальных системах используется система онлайн-тестирования, в которой можно сравнивать несколько алгоритмов. Обычно такие системы перенаправляют небольшой процент трафика в другую альтернативную систему рекомендаций и записывают взаимодействия пользователей с различными системами.

# Онлайн и оффлайн тестирование



Качество можно тестировать как на отложенной выборке так и в бою.

Только есть одна проблема: если у нас метрика CTR, как нам тестировать на отложенной выборке?

# АВ-тест (почти формально)



АВ-тестирование – это проверка гипотезы, о том, что реализации случайной величины в одной версии (А) *отличается* от реализации случайной величины в другой версии (В).

# АВ-тест



АВ-тестирование - это эксперимент, целью которого является сравнение двух вариантов одного и того же объекта. Версии А и версии В.

# АВ-тест



АВ-тестирование - это эксперимент, целью которого является сравнение двух вариантов одного и того же объекта. Версии А и версии В.

Версия А лучше версии В – допустимый тест, но НЕ самый лучший.

# Гипотезы



Как правило:

$H_0$  - между A и B нет разницы.

$H_1$  - ( $\overline{H_0}$ ) B лучше A.

# Матрица ошибок

Результат  
применения  
критерия

Таблица бок	Верная гипотеза	
	$H_0$	$H_1$
$H_0$	$H_0$ Верно принята	$H_0$ Неверно принята (ошибка II рода)
$H_1$	$H_0$ Неверно отвергнута (ошибка I рода)	$H_0$ Верно отвергнута




# Ошибки I и II рода



Ошибки не симметричны!

Ошибка I рода – отвергнуть  $H_0$ , когда она верна.

Ошибка I рода, как правило, критичнее и она жестко ограничена сверху.

Из всех критериев, которые удовлетворяют ограничениям выбирают, тот у которого Ошибка II рода минимальна.

Мы не делаем никаких выводов  
о гипотезе  $H_1$ !

И мы также не делаем вывода, что  $H_0$   
верна!

Отсутствие доказательств чего-то,  
**не** является доказательством обратного!

# Размер эффекта




Как правило в тестах оценивают или фиксируют минимально допустимый «размер эффекта».

Пример.

Оборачиваемость капитала выросла на 3%.

Абоненты стали больше пользоваться услугой на 2% процента.

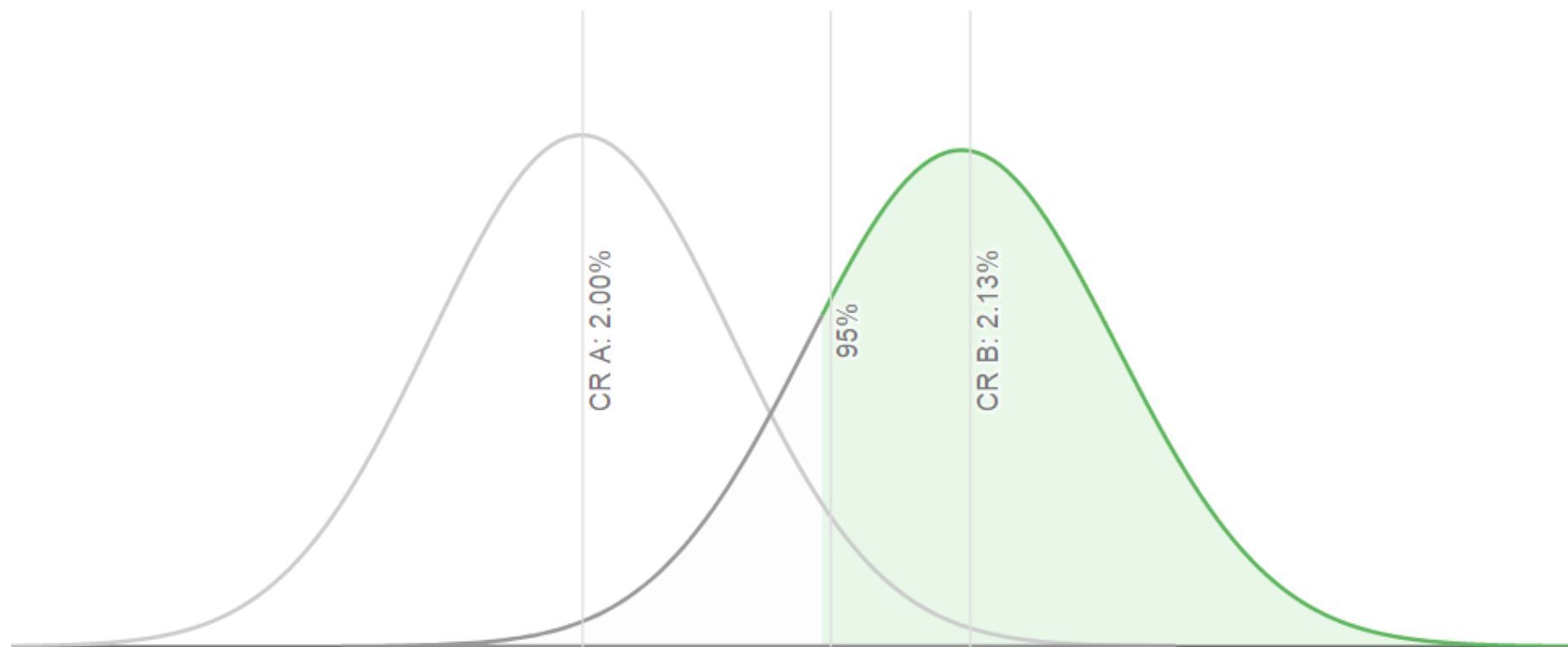
# Размер эффекта



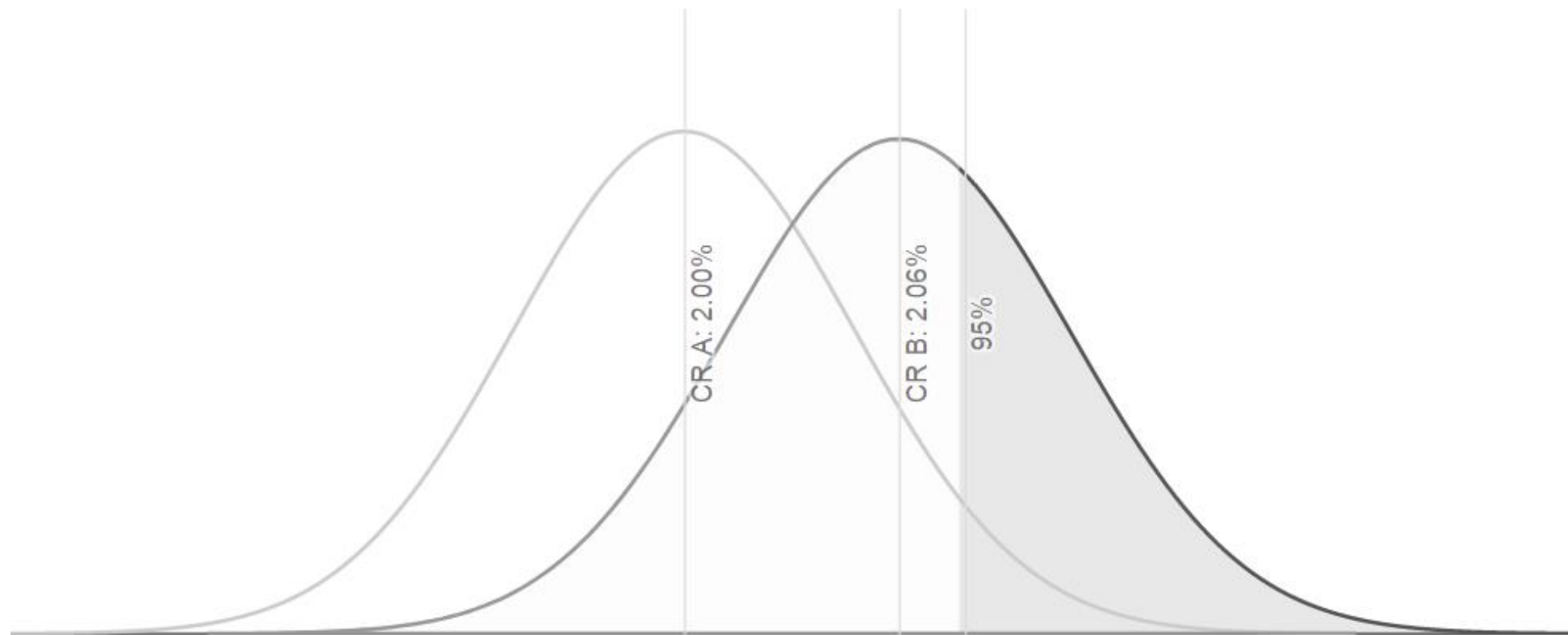
Нетривиально.

Чем **больший** размер эффекта вы хотите отловить, тем **меньше** нужно данных.

# Статистическая значимость



# Статистическая значимость



# Статистическая значимость

Вариант А		Вариант В	
Кол-во конверсий	100	Кол-во конверсий	115
Размер выборки	10000	Размер выборки	10000

Вердикт:

**между вариантами нет разницы**

результат не является статистически значимым



# Как проводить АВ-тест

# Разработайте гипотезу



Нужно определить две гипотезы:

- **Нулевая гипотеза** предполагает, что результаты, А и В на самом деле не отличаются и что наблюдаемые различия случайны.
- **Альтернативная гипотеза** — это гипотеза о том, что В отличается от А, и вы хотите сделать вывод об её истинности.

# Подготовьте эксперимент



Что нужно сделать, чтобы тест выдал корректные результаты:

- Создайте новую версию (B), отражающую изменения, которые вы хотите протестировать;
- Определите контрольную и экспериментальную группы;

# Подготовьте эксперимент



Что нужно сделать, чтобы тест выдал корректные результаты:

- Убедитесь, что пользователи будут видеть версии А и В в случайном порядке;
- Определите допустимые ошибки I (и II рода)
- Определите минимальный размер выборки;

# Тонкий момент



1. Ошибка I (и II) рода
2. Минимальный размер выборки
3. Размер эффекта

Связаны между собой.

Для эксперимента нужно зафиксировать 2 из этих параметров.

# Проведите эксперимент



- Обсудите параметры эксперимента с исполнителями;
- Выполните запрос на тестовой закрытой площадке, если она у вас есть;
- В самом начале проведения тестирования проверьте, действительно ли оно работает;
- И, наконец, **не смотрите** на результаты!

# Анализируйте результаты



Проверьте статистическую значимость!

- Если  $p$ -значение меньше уровня значимости, то можно отвергнуть нулевую гипотезу, имея доказательства для альтернативы.
- Если  $p$ -значение больше или равно уровню значимости, мы не можем отвергнуть нулевую гипотезу о том, что  $A$  и  $B$  не отличаются друг от друга.

# Могут быть следующие результаты

- Контрольная версия,  
А выигрывает или  
между версиями нет  
разницы.

- Версия В выигрывает.





# Как поставить АВ-тест



1. Что мы измеряем
2. Вероятность ошибки I (и II) рода
3. Размер эффекта
4. Размер выборки (рассчитывается на основе 2. и 3.)

Пример:

1. Средний чек
2. Ошибка I рода: 0.05
3. Увеличится на 50 р.
4. 100 000 человек

# НЕ АВ-тест



«Запустим АВ-тест и будем ждать пока не наберётся статистическая значимость.»

# НЕ АВ-тест



«Запустим АВ-тест и будем ждать пока не наберётся статистическая значимость.»

Почему так нельзя делать?

# НЕ АВ-тест



«Запустим АВ-тест и будем ждать пока не наберётся статистическая значимость.»

Почему так нельзя делать?

Тогда мы измеряем событие «Нулевая гипотеза» будет отвергнута хоть раз.

Если вы проверяете много гипотез,  
то пороги статистической  
значимости меняются!

# Проверка расчетов



Для проверки результатов тестирования используйте аналитика, (если его нет, то калькуляторы)! 😊  
Они помогут подсчитать достоверность теста.

- [Driveback](#) на русском;
- [Hungrysites](#) на русском;
- [Mindbox](#) на русском;
- [ABTestGuide](#) на английском, но с визуализацией графиками;
- <https://yandex.ru/adv/statvalue> на русском.

# Проверка расчетов

## Калькулятор достоверности A/B тестирования

Калькулятор позволяет быстро сравнить результаты A/B-тестов нескольких рекламных стратегий, а значит — определить наиболее эффективный набор инструментов и форматов.

Целевое p-value

80

%

Максимально допустимый коэффициент жесткости дополнительных вложений

2


Скачать Excel шаблон таблицы

Сегмент	Размер сегмента	Расходы в сегменте	Конверсии	Клики	CPA сегмента	Расходы относительно А	Конверсии относительно А	К о	Внедряем?
A	50	%			—	—	—	—	
B	50	%			—	—	—	—	—

# Особенности АБ-тестов для рекомендательных систем



# Особенности АБ для рекоменд. сист.



1. Удобно делать АВС-тесты.
  1. АБ – версии рек.сис
  2. С – сильный бейзлайн (2-5% пользователей)
2. Изменение визуал/функционал продукта влияет на качество рекомендаций. Если меняются рекомендательные блоки нужно повторять АБ-тесты.
3. Скорость ответа сервиса часто критичнее качества рекомендаций.

Вопросы